Gene 518 (2013) 17–25

Contents lists available at SciVerse ScienceDirect

Gene

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

An estimation method for a cellular-state-specific gene regulatory network along tree-structured gene expression profiles

Ryo Araki, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda *

Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1-5, Yamadaoka, Suita, Osaka 565-0871, Japan

ARTICLE INFO

Available online 21 December 2012

Keywords:
Cellular states
Tree-structured gene expression profiles
Gene regulatory network

ABSTRACT

Background: Identifying the differences between gene regulatory networks under varying biological conditions or external stimuli is an important challenge in systems biology. Several methods have been developed to reverse-engineer a cellular system, called a gene regulatory network, from gene expression profiles in order to understand transcriptomic behavior under various conditions of interest. Conventional methods infer the gene regulatory network independently from each of the multiple gene expression profiles under varying conditions to find the important regulatory relations for understanding cellular behavior. However, the inferred networks with conventional methods include a large number of misleading relations, and the accuracy of the inference is low. This is because conventional methods do not consider other related conditions, and the results of conventional methods include considerable noise due to the limited number of observation points in each expression profile of interest.

Results: We propose a more accurate method for estimating key gene regulatory networks for understanding cellular behavior under various conditions. Our method utilizes multiple gene expression profiles that compose a tree structure under varying conditions. The root represents the original cellular state, and the leaves represent the changed cellular states under various conditions. By using this tree-structured gene expression profiles, our method more powerfully estimates the networks that are key to understanding the cellular behavior of interest under varying conditions.

Conclusion: We confirmed that the proposed method in cell differentiation was more rigorous than the conventional method. The results show that our assumptions as to which relations are unimportant for understanding the differences of cellular states in cell differentiation are appropriate, and that our method can infer more accurately the core networks of the cell types.

1. Introduction

Systems biology aims to understand cellular processes by using mathematical models (Cantone et al., 2009). One of the important themes in systems biology is to reverse-engineer the dynamics of gene regulatory relations, called gene regulatory networks (GRNs), from gene expression profiles. Several methods have been developed for the mathematical modeling of these dynamics from gene expression profiles. Boolean network (Xiao, 2009), graphical Gaussian model (Grzegorczyk, 2007; Toh and Horimoto, 2002), mutual information model (Margolin et al., 2006), Bayesian network (Heckerman, 1996), and relevance network (Butte and Kohane, 2000; Butte et al., 2003) are widely used.

GRNs have been inferred under various biological conditions, such as stimulated cellular response (Nagashima et al., 2007; Shinozaki et al., 2003) and cell differentiation (Carter et al., 2004; Siersbak and Mandrup, 2011; Siersbak et al., 2012; Tokuzawa et al., 2010; Zhang et al., 2011). The aim of this research is to identify the key gene regulatory relations in a given cellular state; this is because the key relations in a state help us to understand the differences between various cellular states. Cellular states are described by GRNs in this area of research; different states have different GRNs, and the states are changed by the conditions. To find the key relations in cellular states, conventionally, the GRN of the cellular state has been inferred from a single gene expression profile under the condition of interest.

It is not an easy task for conventional methods to identify the key gene regulatory relations in a cellular state. This is because conventional methods run into two problems. The first is that conventional methods infer the GRN of a cellular state from a single gene expression profile under the condition of interest; these are called *exact samples* (Hu, 1994; Hu et al., 2000). With the conventional methods, the inferred GRN includes relations in common with other states

Abbreviations: GRN, gene regulatory network; GEO, Gene Expression Omnibus; GEP, gene expression profile; ROC, receiver operating characteristic; MSC, mesenchymal stem cell; RN, relevance network; GGM, graphical Gaussian modeling.

* Corresponding author.
E-mail address: matsuda@ist.osaka-u.ac.jp (H. Matsuda).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.11.090

and relations specific to the given state. Another problem is that for the GRNs inferred from the expression profiles that are derived from public databases such as Gene Expression Omnibus (GEO), more than 90% of profiles have fewer than 10 observation points. The result is that the inferences produced by the conventional methods include a large number of candidate relations that are common to other states or unique to the given state. Due to above two problems, the inferred GRNs of conventional methods have low accuracy.

In this paper, we propose a more accurate method of inferring the key GRN in a cellular state from multiple gene expression profiles under varying biological conditions. It should be noted that conventional methods aim to infer a GRN under a condition of interest, and they are not intended to compare multiple gene networks, that is, to infer each GRN independently from *exact samples*. Our method is different from these conventional methods in that it utilizes the exact profile along with profiles from other conditions, called *relevant samples* (Hu, 1994; Hu et al., 2000), for inferring the key GRN in a cellular state. We assume that the key GRN in a cellular state is a relation specific to that state, and that removing common relations from the GRN of the state enables a more accurate inference of the key GRN. We also use multiple gene expression profiles that compose tree structure, and in this tree, both *exact sample* and *relevant samples* are included. The root of the tree corresponds to an original state of a target state and the comparison states, and the leaves represent the target state and the comparison states. The profiles from the root to the target leaf are merged and used as a profile of the target state, and the profiles from the root to the comparison leaves are merged and used as profiles of the comparison state. These merged profiles decrease the candidate relations in the inferred GRNs, and the specific relations are extracted from the GRN of the target by comparing them with the GRNs of the comparisons. Therefore, our method can estimate the key GRN of the target state with higher accuracy.

2. Method

In this study, we defined cellular-states-specific GRNs as a set of gene regulatory relations that determine the differences of cellular states. It is difficult for conventional methods to extract the relations specific to a cellular state because, as mentioned, the small number of observation points in the profiles leads the inferred GRN to include a large number of candidates, and not to consider other cellular states lead to include relations common and specific to the other states. Thus, our method considered the other states related to target state and decreased the candidate relations by applying tree-structured gene expression profiles. Below is the notation we will use for the explanation of our methods.

2.1. Notation

$S=\{S_1,\ldots,S_m\}$: A set of cellular states.
S_i: An original cellular state of S_i.
$G=\{g_1,\ldots,g_n\}$: A set of genes g.
$GEPs=\{GEP_1,\ldots,GEP_m\}$: A set of gene expression profiles of S.
$GEP_i[G,T_i]$: A gene expression profile including observation points T_i and a set of gene G.
$T_i=\{t_1^i,\ldots,t_j^i\}$:A set of observation points t of GEP_i.

We explain our tree-structured profiles with this notation. To simplify the explanation we will describe our method with three gene expression profiles. This example is easily applied to more gene expression profiles. One profile is observed from S_0 to S_1, a second is observed from S_1 to S_2 and another is from S_1 to S_3 (see Fig. 1). Let S_2 be a target cellular state, then the conventional methods infer a target GRN from the profile that is observed from S_1 to S_2. The results include key interactions for understanding the behavior of a target state; however, the inferred GRN has too many candidate relations to determine the key relations of this behavior because the profile has only a few observation points and the conventional inference of the GRNs does not consider other cells S_3. For understanding the differences in the behaviors, it is necessary to utilize other profiles and to compare the GRNs in order to detect the differences between the target cellular state and the other cellular states. Thus, our methods use other profiles both for decreasing the candidate relations in the inferred GRNs and for removing the relations common to S_2 and S_3.

For extracting a GRN specific to a cell, we assumed that there are two kinds of common relations in the GRNs. One is the gene relations common to all cellular states, such as the relations maintaining cellular states. A conceptual view of this assumption is shown in Fig. 2. Another assumption is that there are relations common to a particular series of cellular states, such as relations inherited from an original state S_0. We assume that we cannot distinguish the behavior unique to a target state from these relations (see Fig. 3). And common relations to a particular series of states cannot be estimated from profiles with few observation points, because the relations are quiet.

We propose two methods following from these assumptions for a more accurate estimation of cellular-states-specific relations.

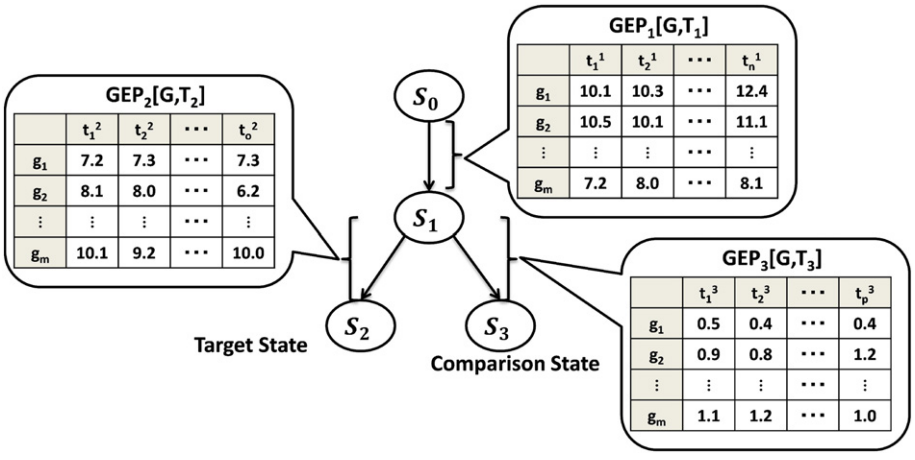

Fig. 1. Tree-structured gene expression profiles. This figure shows an example of tree-structured gene expression profile composed of 3 profiles. The circled S corresponds to cellular states. One profile observes the transitions from the root S_0 to S_1, another profile is from S_1 to S_2, and the third is from S_1 to S_3. Cellular states are changed from S_0 through S_1 to S_2 and S_3.