BioSystems 155 (2017) 29–41

Contents lists available at ScienceDirect

BioSystems

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

Applying attractor dynamics to infer gene regulatory interactions involved in cellular differentiation

Ahmadreza Ghaffarizadeh[a,*], Gregory J. Podgorski[b,c], Nicholas S. Flann[a,d,e]

[a] Computer Science Department, Utah State University, 4205 Old Main Hill, Logan, UT 84322, United States
[b] Biology Department, Utah State University, 5305 Old Main Hill, Logan, UT 84322, United States
[c] Center for Integrated BioSystems, 4700 Old Main Hill, Logan, UT 84322, United States
[d] Institute for Systems Biology, 401 Terry Ave N, Seattle, WA 98109, United States
[e] Synthetic Biomanufacturing Institute, 1780 N. Research Park Way, Suite 108, North Logan, UT 84341, United States

ARTICLE INFO

Article history:
Received 6 August 2016
Received in revised form 6 December 2016
Accepted 22 December 2016
Available online 28 February 2017

Keywords:
Gene regulatory networks
Nested canalyzing functions
Attractor dynamics
Genetic algorithm

ABSTRACT

The dynamics of gene regulatory networks (GRNs) guide cellular differentiation. Determining the ways regulatory genes control expression of their targets is essential to understand and control cellular differentiation. The way a regulatory gene controls its target can be expressed as a gene regulatory function. Manual derivation of these regulatory functions is slow, error-prone and difficult to update as new information arises. Automating this process is a significant challenge and the subject of intensive effort. This work presents a novel approach to discovering biologically plausible gene regulatory interactions that control cellular differentiation. This method integrates known cell type expression data, genetic interactions, and knowledge of the effects of gene knockouts to determine likely GRN regulatory functions. We employ a genetic algorithm to search for candidate GRNs that use a set of transcription factors that control differentiation within a lineage. Nested canalyzing functions are used to constrain the search space to biologically plausible networks. The method identifies an ensemble of GRNs whose dynamics reproduce the gene expression pattern for each cell type within a particular lineage. The method's effectiveness was tested by inferring consensus GRNs for myeloid and pancreatic cell differentiation and comparing the predicted gene regulatory interactions to manually derived interactions. We identified many regulatory interactions reported in the literature and also found differences from published reports. These discrepancies suggest areas for biological studies of myeloid and pancreatic differentiation. We also performed a study that used defined synthetic networks to evaluate the accuracy of the automated search method and found that the search algorithm was able to discover the regulatory interactions in these defined networks with high accuracy. We suggest that the GRN functions derived from the methods described here can be used to fill gaps in knowledge about regulatory interactions and to offer hypotheses for experimental testing of GRNs that control differentiation and other biological processes.

1. Background

Two goals of systems biology are to obtain a blueprint of gene and protein interactions and to understand how these molecular interactions give rise to emergent cell and organismal level properties. The increasing availability of high-throughput gene expression data coupled with effective automated literature mining tools has advanced the ability to map molecular interactions. However, there is a crucial need to harness this trove of information to infer the structure of gene regulatory networks (GRNs) and to understand how they can lead to emergent biological properties.

There are many approaches to modeling GRNs that control differentiation and other cellular processes. Each of these approaches differs in its scalability, simplicity, parameterization and the amount of information required (Berestovsky and Nakhleh, 2013; Hecker et al., 2009). The four most common methods are: (1) information theory models, (2) Bayesian networks, (3) systems of equations, and (4) Boolean networks. Of these approaches, Boolean networks have the advantages of simplicity and the capability to model GRN dynamics without knowledge of any kinetic parameters (Krumsiek et al., 2011). The properties of Boolean networks and their utility for modeling have been studied extensively (Aldana et al., 2003; Drossel, 2009). Although Boolean networks

* Corresponding author.
E-mail addresses: ghaffarizadeh@aggiemail.usu.edu (A. Ghaffarizadeh), greg.podgorski@usu.edu (G.J. Podgorski), nick.flann@usu.edu (N.S. Flann).

http://dx.doi.org/10.1016/j.biosystems.2016.12.004
0303-2647/

have limitations in their ability to model some important features, such as signal amplification or responses to continuously varying environmental conditions like temperature or nutrient availability (D'haeseleer et al., 2000), these disadvantages are balanced by the strength of Boolean networks for providing qualitative interpretations of regulation, particularly state switching. These are among the reasons for choosing a Boolean network modeling framework for this current study.

Boolean networks are dynamical networks in which each gene is either expressed or not expressed. Every gene in the network is represented by a node, with each node associated with a regulatory function that specifies its output as either on (1) or off (0) based on the values of inputs to the node. First proposed for use in studies of gene regulation by Kauffman (1969), Boolean networks are a major contributor to our knowledge of gene regulatory networks and have been used to represent GRN structure and dynamics in many systems, including *Drosophila* development (Albert and Othmer, 2003; Bodnar, 1997), angiogenesis (Bauer et al., 2010), eukaryotic cell dynamics (Shmulevich et al., 2005), and yeast transcription networks (Kauffman et al., 2003). Boolean networks often generate outputs that are in excellent agreement with those of known biological networks.

Kauffman (1969) originally suggested that cell types are attractors in the dynamics of Boolean networks that execute GRNs. Subsequent studies have supported this view (Davila-Velderrain et al., 2015; Zhou et al., 2011; Huang et al., 2005; Enver et al., 2015). In the state space of a dynamical system, an attractor is a closed subset of the states toward which the network tends to converge, regardless of the starting condition. Kauffman proposed that Boolean networks need to be robust against the intrinsic noise of the system in order to model GRNs appropriately and he introduced canalyzing functions (CFs) as regulatory functions that provide robustness to noise in GRNs. In biological networks, eukaryotic gene regulation appears to be strongly biased towards canalyzing functions (Harris et al., 2002). With CFs, the value of one input is dominant and determines (canalyzes) the output of the function irrespective of the value of the other inputs. CFs shift the dynamics of the system from the chaotic domain to the critical domain where GRNs are believed to operate (Nykter et al., 2008; Kauffman, 1993). Nested canalyzing functions (NCFs), which are used in this work, are a special case of canalyzing functions. When regulatory functions are represented with NCFs, all inputs have the potential to canalyze the output, but there is a dominance hierarchy, or ordering, of the inputs that is defined for the function (Kauffman et al., 2003).

Previous approaches have been taken to identify Boolean networks from gene expression data. Early work demonstrated that GRNs could be inferred given sufficient time series data (Akutsu et al., 1999; Lä et al., 2003). These studies were extended to deal with noise (Shmulevich et al., 2002; Maki et al., 2001) and experimental perturbations of the network (Ideker et al., 2000). Methods specifically designed to identify GRNs for cellular differentiation have been developed that seek to match network attractors to different cell types that are characterized by distinct gene expression signatures (Pal et al., 2005; Layek et al., 2011).

In this work, we use the concept of attractors as cell types in the development of a search mechanism to identify Boolean networks that can control cellular differentiation. The method employs a genetic algorithm (Holland, 1992) that searches the space of possible GRNs to identify networks that produce a set of attractors with gene expression profiles that match those of the target cell types. Previous work (Pal et al., 2005; Layek et al., 2011) has demonstrated that this problem is inherently under-constrained because there are many networks capable of producing the target attractors, with only a few of these being biologically realistic. To mitigate this problem and focus the search only on biologically plausible networks, regulatory functions are limited to NCF and data from biological studies is applied. Often this data is fragmentary, but it is sufficient to narrow the set of potential targets networks.

The computationally derived regulatory functions form dynamic models useful for prediction, for GRN engineering (Zhou et al., 2011), and to inform experimental studies to reveal regulatory network architecture. We demonstrate the utility of the method for two test cases: pancreatic cell differentiation starting from a pancreatic endoderm precursor cell, and myeloid cell differentiation starting from a common myeloid progenitor cell. A study using defined synthetic networks is included to evaluate the accuracy of the method in discovering completely known networks. The modeling framework described here can potentially be extended to other network-driven biological processes in which there is only partial knowledge of gene or protein interactions.

2. Results

2.1. Overview of the approach

Our initial studies demonstrated that many networks were able to produce attractors that matched the gene expression patterns of target pancreatic and myeloid cell types. However, the majority of these GRNs contain interactions between network proteins that are not supported by the biological literature. Combining information on the types of regulatory interactions between proteins, the effects of knockout mutations, and knowledge of biological network topology, and using this information to filter the initially discovered GRNs, sharply restricted the number of candidate networks. These constraints, including the restriction of regulatory functions to nested canalyzing functions (Kauffman et al., 2004), yielded a set of similar, but not identical, high-scoring GRNs. Within this ensemble of GRNs, particular pairs of interacting proteins were scored for whether the regulatory protein influenced its target protein through activation or inhibition, and these predictions were compared to reports in the literature. Most often, when a set of similar candidate networks was discovered, the predicted form of interaction between the regulatory protein and its target were in agreement with published reports.

2.2. Searching for simple GRNs: pancreatic cell differentiation

The pancreas is an exocrine and endocrine organ that secretes digestive enzymes and hormones, including insulin. The development of the pancreas and the differentiation of its cell types has been intensively studied (Zhou et al., 2011, 2014; Habener et al., 2005; Oliver-Krasinski and Stoffers, 2008; Zaret and Grompe, 2008; Jensen, 2004). A simplified lineage tree that shows the differentiation of the pancreatic exocrine and endocrine cells is given in Fig. 1. This figure also shows the discretized expression levels of 5 key transcription factors (TFs) in exocrine, β/δ progenitor, and α/PP progenitor cells as reported in Zhou et al. (2014). Fig. 2 shows both experimentally validated and proposed interactions between these genes in pancreatic cell differentiation (Zhou et al., 2014).

The discretized transcription factor levels were used to infer a gene regulatory network by the approach described in Section 4. The search algorithm was run initially 100 times over the unconstrained GRN space in which no restrictions were placed on either regulatory functions or network topology. In this initial run, the objective function was calculated based only on consistency between the known gene expression levels for each target cell type and the network attractors. For each of the 100 unconstrained runs of the algorithm, we saved the best solution as a candidate GRN. All of these solutions produced an attractor set that exactly matched the target cell type expression signatures. This run