Module network rewiring in response to therapy

Tao Zeng, Ru Zhao Wang, Luonan Chen
Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
* to whom correspondence should be addressed

Abstract—Response to stress is an important biological mechanism to react to environment variations. Different from distinguishing stresses like heat shock, ER stress, and oxidative stress, the study of response to an artificial signal like drug in therapy would be an alternative and also attractive way to understand the cellular response mechanism, which also benefits clinical application. Although differentially expressed genes are usually thought to be therapy responsive genes in many previous researches, more and more attention is diverted from single genes to functions or pathways, in particular for cancer therapy analysis. Thus, comparing with purely molecule (e.g., gene) rewiring, understanding functional reorganization or module rewiring would be more important for systematically studying therapy response or other dynamic biological processes. Therefore, in this paper we propose a model of module network rewiring to characterize functional reorganization, in contrast to gene network rewiring. Specifically, we develop a new framework named as module network rewiring analysis (MNRA) to investigate relevant network modules and their re-connections during an antiviral therapy. In MNRA, we aim to study module dynamics from the network viewpoint, by defining a module network with a module as a node and a path connecting two modules as an edge, which is a network for the molecular interaction system on a higher level. By MNRA experiments on expression data of patients with Hepatitis C virus infection (HCV) receiving Interferon therapy, we found that (1) the consistent module (a set of genes) separates two new subtypes of patients which were not discovered by differentially expressed genes; (2) the patient-group specific module network rewiring reveals necessary functional connections bridged by biological paths; (3) the hierarchical structures of temporal module network rewiring show that they can be taken as spatial-temporal markers to diagnose whether a patient has therapy response or not. Thus, MNRA indeed can provide biologically systematic clues for potential pharmacogenomic applications and has ability to characterize complex dynamic processes for many biological systems.

I. INTRODUCTION

Response to stress is an important biological mechanism to react to environment variations. Although the cellular response to distinguish stresses like heat shock, ER stress, and oxidative stress are natural cell decision [1], studying response to artificial signal like drug in therapy is an alternative and also attractive way to understand its complex mechanism. In particular, on- or off-status of drug response is determinate and manipulatable, and thus can be taken as control or case condition for comparison study. Generally, differentially expressed genes are thought to be therapy responsive genes, and such genes have been intensively investigated in many previous researches [2]. However, it is well recognized that individual genes or proteins can not fully facilitate biological functions, and it is their interactions or networks (or modules) that ultimately hold responsible for dynamical behaviors and functions of living organisms. Therefore, the study on response to therapy based on a molecular network or its local structures, e.g., network motifs and network modules, can not only elucidate its essential principle to react to drug at the network level, but also reveal fundamental mechanisms on general stresses, thereby benefitting efficient identification of biomarkers in pharmacogenomic [3].

In contrast to conventional studies on therapy response that mainly focused on the responsive genes [4]–[6], recently more and more attention is diverted from single genes to particular functions or pathways like DNA-damage response in cancer therapy [7], [8]. But how genes’ cooperation or even biological functions’ collaboration to achieve an effective or non-effective response to therapy is still far from clear. Actually, the mechanism for dynamically biological processes in therapy response is rooted from the functional reorganization. Therefore instead of rewiring of gene regulations (or gene network rewiring), we develop a novel model for module network rewiring to describe such a functional reorganization and further reveal its dynamic features of a therapy response on the functional network level.

Specifically, the functional reorganization model (or module network rewiring model) contains two aspects. One is to determine the groups or subnetworks of cooperative genes across different times or conditions like days before and after drug injection, i.e., detecting consistent modules among the changing networks. The other is to identify the varying collaborations or interactions between different gene groups or modules during the dynamic biological process just like therapy response, i.e., connecting the consistent modules by the changing interactions to construct the module network. Actually, many previous studies have discussed the cooperative genes (as modules) or varying gene cooperation (as gene network rewiring), where a gene module is generally considered as a basic unit to perform biological functions at the network level. Different from mining the conserved gene expression (expression module) [9], discovering network modules is a difficult computation task, and many methods have been applied to this problem [10], [11], but there is few work to detect the consistent modules across different conditions, and also few researchers have exploited dynamic interactions among modules to analyze network functions although there are some simple schemes using the overlapped
genes to link different modules [12]–[14]. In addition to detecting network modules, study on gene network rewiring behind cellular response has also attracted much attention and becomes a key biological model in recent studies [15], however, these methods mainly paid attention on characterizing the changes between molecules, rather than dynamical regulations during spatio-temporal rewiring processes. In order to reveal the essential mechanism of functional reorganization at the network level, it is necessary to develop a theoretical model to combine the network module and network rewiring together in a dynamic and unified manner.

Considering the specificity of drug response compared to other types of stress responses, we develop a new computational framework, named as module network rewiring analysis (MNRA), to investigate the relevant network modules and their re-rewiring dynamics during antiviral therapy, based on spatio-temporal gene expression data and molecular interaction data. In MNRA, we aim to study module dynamics from the network viewpoint, by defining a module network with each module as a node and each path connecting two modules as an edge, which is a macro-network and is also a graph model for the molecular interaction system on a higher level. Note that a module in this paper is actually a consistent module whose edges within the module have no rewiring during the biological process. Clearly, rewiring in a module network can reflect the global variation and local stability of the network in a unified way.

By experiments on the expression data of patients with Hepatitis C virus infection (HCV) receiving Interferon therapy [16], we show that MNRA reveals distinct changes of functional relationships for drug treatment response. In particular, our results provide the biological insights which imply potential clinical applications: Different from the differentially expressed genes, module genes are able to distinguish two new subtypes of therapy responders and non-responders; Based on the phenotypes of therapy response, the rewiring of the patient-group specific module network reveals many important functional connections; The numerical encoding of module network reorganizations can be taken as spatio-temporal markers which can diagnose whether or not a patient has therapy response. In summary, MNRA method and its preliminary experimental results strongly demonstrate that our module network model has a strong analysis power comparing with traditional gene network based model, and therefore has great potential to be used in various stress response researches as well as studies on functional reorganization and biological system transitions.

II. METHOD

A. Model of module network rewiring

There are two kinds of datasets used in modelling module network rewiring. One dataset is organized as a tensor data and denoted as $D(G, P, T)$, whose each element $d_{g,p,t}$ represents the gene expression value of one gene $g \in G$ for some sample $p \in P$ measured at time point $t \in T$. The gene expression profile of a gene $g$ at a time point $t$ with all samples is $D(g, P, t)$. And the gene expression profile of a gene $g$ of one sample $p$ during a time period $T_{s,c}$ (it starts at time point $s$ and ends at time point $e$) is $D(g, p, T_{s,e})$. The other dataset is the physical interaction network (PIN) [17] composed of the known protein-protein interactions and TF-target interactions collected in different databases (KEGG [18], ITTF [19], Tred [20], and iRefIndex [21]), which is organized as an indirected network $PIN = \{G, E\}$ whose node (gene/protein) set is $G$ and edge (interaction) set is $E$. To measure the rewiring of a network is to evaluate the existence of the interactions (edges) in the network. In computational analysis, this is achieved by measuring the weight of each edge or the association between a pair of genes involved in an interaction. Commonly, the genes’ pairwise association is calculated as Pearson Correlation Coefficient (PCC). For convenience, the correlation between two genes for a group of samples related to a time point is noted as $PCC_{g_1, g_2, t}$, while the correlation between genes for a series of time points related to a sample is noted as $PCC_{g_1, g_2, p}$. Formally, these two correlation values are calculated as follows:

$$PCC_{g_1, g_2, t} = \frac{\sum_{p \in P} (d_{g_1, p, t} - \bar{d}_{g_1, t}) (d_{g_2, p, t} - \bar{d}_{g_2, t})}{\sqrt{\sum_{p \in P} (d_{g_1, p, t} - \bar{d}_{g_1, t})^2} \sqrt{\sum_{p \in P} (d_{g_2, p, t} - \bar{d}_{g_2, t})^2}}$$

(1)

$$PCC_{g_1, g_2, p} = \frac{\sum_{t \in T} (d_{g_1, p, t} - \bar{d}_{g_1, p}) (d_{g_2, p, t} - \bar{d}_{g_2, p})}{\sqrt{\sum_{t \in T} (d_{g_1, p, t} - \bar{d}_{g_1, p})^2} \sqrt{\sum_{t \in T} (d_{g_2, p, t} - \bar{d}_{g_2, p})^2}}$$

(2)

where, $\bar{d}_{g_1, t} = \frac{\sum_{p \in P} d_{g_1, p, t}}{|P|}$ and $\bar{d}_{g_1, p} = \frac{\sum_{t \in T} d_{g_1, p, t}}{|T|}$ are means of expressions of gene $g$ among a group of samples $\{p \in P\}$ or at several time points $\{t \in T\}$ respectively.

We develop the module network rewiring model with several steps whose details are given in following subsections.

1) Construction of PIN co-expression network: We first construct molecular networks based on available data, e.g., expression data and the reference network. There are two kinds of context-based networks in our study. One is the patient-specific network (across whole time periods) for one patient or a time window, and the other one is the time-specific network (across patients or samples) at one time point or a time window, respectively.

2) Construction of module network: Based on the correlation matrix, we need to group genes with similar behavior into modules. We use the following steps whose details are given in following subsections.
the time-specific network is denoted as PIN(P, t), where t is
time point and P means a group of patients.

2) Decomposition of PIN co-expression network by consistent modules: We decompose molecular networks based
on edge consistency across samples and times. In order to
find consistent sub-networks across multiple context-based
networks (focused on time-specific networks here), a con-
secutive version of frequent graph mining method [23] is
adopted. The overlapped edges presenting in the networks
{PIN(P, t)}_{t \in T} consist of the so-called consistent network S.
Similarly, the overlapped edges only appearing in the networks
after particular time points consist of the appearance-consistent
network \(S^k\), while the overlapped edges only existing in the networks
before particular time points consist of the absence-
consistent network \(S_k\). Formally, the three kinds of consistent
networks can be obtained according to following definitions.

\[
S = \cap_{t \in T} PIN(P, t) \quad (3)
\]
\[
S^k = \cap_{t \in T, t \geq k} PIN(P, t) - \cup_{t \in T, t < k} PIN(P, t) \quad (4)
\]
\[
S_k = \cap_{t \in T, t \leq k} PIN(P, t) - \cup_{t \in T, t > k} PIN(P, t) \quad (5)
\]

As a natural partition of a consistent network, the maximal
connected components of this network are consistent modules,
or appearance-consistent modules or absence-consistent
modules, respectively. Furthermore, two commonly adopted
criteria were used to filter the possible trivial modules, i.e.,
remove the modules with genes less than 3, and remove the
modules without enough biological significance according to
the functional enrichment analysis [24].

3) Reconstruction of PIN co-expression network: module network:
Then, we reconstruct molecular networks by using
all consistent modules, i.e., construct module networks
from any original patient-specific or time-specific molecular
networks. Given there is a list of modules \(\{M_{\lambda \in \Omega}\}\) and
a molecular network PIN, the target module network \(\Psi^P\)
consist of four feature sets. The node set \(V(\Psi^P) = \{M_{\lambda \in \Omega}\}\)
consists of all given modules and each module is a meta-
node of network \(\Psi^P\). The edge set \(E(\Psi^P)\) contains pairs of
modules and each edge means that there is at least one path
or route of genes in the molecular network PIN to directly link
the two respective modules. The node weight set \(W_V(\Psi^P)\)
and edge weight set \(W_E(\Psi^P)\) contain the so-called activity
scores of modules and their connections. The weight of one
node module or a pair of connected modules is calculated as
follows:

\[
W_V(\Psi^P) = \sum_{x \in M_{\lambda \in \Omega}, y \in M_{\lambda \in \Omega}} (e_x + e_y)PCC_{x,y} \mid M_{\lambda} \in V(\Psi^P) \quad (6)
\]
\[
W_E(\Psi^P) = \sum_{x \in M_{\lambda \in \Omega}, y \in M_{\lambda \in \Omega}, sp(x, y) > 0} sp(x, y) \mid (M_{\lambda}, M_{\mu}) \in E(\Psi^P) \quad (7)
\]

where \(e_x\) is the differential expression of gene \(x\), and \(sp(x, y)\)
is the length of shortest route between genes \(x\) and \(y\) on the
context-based molecular network, and \(sp(x, y)\) is equal to the
given maximum value when \(x\) is not connected to \(y\). Note
that \(W_V\) is related to the correlation but \(W_E\) is related to the
distance between the two components. A module is considered
to be more active or more functional in the biological system
when its members have close correlation and high differential
expression simultaneously. Otherwise, the module is inactive.
On the other hand, the weight of modules’ association reflects
the closeness among all genes in the two modules. A small
weight \(W_E\) (distance) between two modules means that the
two modules may be the two components of another function
module on a high level of the biological network because of
their strong correlation. Otherwise, the two modules will be
independent when the weight of their connection is large.

4) Functional annotation of consistent modules: We aim to
study module network rewiring that represents the functional
reorganization, rather than molecule network rewiring. Based
on network modules, we analyze the functions of each module
by network ontology analysis [25].

5) Spatial and temporal module network rewiring: At one
time window, different patients have respective module
networks with the same nodes (modules) but different edges
(module connections), which are resulted from the spatial
module network rewiring and can reflect the patient specificity.
Meanwhile, for one patient \(p\), there is also a sequence of
module networks with the same nodes but different edges
among them at different time windows. These changes of
the network structures in a temporal order is defined as the
module network rewiring chain \(C_T^p = \{\Psi^P_t\}_{t \in T}\) for the
corresponding patient, which can describe the dynamics of a
module network (or temporal module network rewiring) during
a biological process like therapy response and potentially
reflect the patient’s response genotype. In fact, the spatial
module network rewiring can be considered as the trivial
condition of the difference among module network rewiring
chains \(C_T^p = \{\Psi^P_t\}_{t \in k}\) when the chains’ length is equal to
one. Therefore, the module network rewiring chain can be used
to unify the spatial and temporal module network rewiring.

B. Framework of module network rewiring analysis

In the experimental study of therapy response, we used
the module network rewiring analysis (MNRA) based on the
above model and method. In this application of MNRA, there
are several patients’ gene expressions at some check points,
such as the time point before drug injection, and a few
following days. We conducted the analysis mainly in three
steps to extract relevant characteristics of module network
rewiring, which can be taken as the potential markers of
therapy response. At first, the temporal consistent network
without the patient-specificity are identified and the modules
are naturally detected from maximal network components,
under the assumption that a therapy is to affect the association
among functions represented by modules. Then, for the two
patient groups with therapy response or not, we rebuild again
their networks in different time points in order to trace the
rewiring with the patient-response specificity. This is

2012 IEEE 6th International Conference on Systems Biology (ISB)
978-1-4673-4398-5/12/$31.00 ©2012 IEEE
Xi’an, China, August 18–20, 2012
a static strategy by using the snapshot of the module network to distinguish patients with different responses. For a dynamical strategy, every patient can also have several structures of module networks corresponding to different time windows during the therapy period. The module network rewiring chain for any patient is expected to reflect patients’ responses to the therapy in a dynamic manner, so that each chain is encoded to a 0-1 sequence (feature vector). Finally, traditional hierarchical clustering is used to show the categories of the module network rewiring chains corresponding to patients, and decision-tree is used to build a prediction model for judging therapy effect on patients at early time point.

III. RESULT AND DISCUSSION

The time course microarray data was obtained from NCBI GEO with ID GSE11342 [16]. In the original study, patients received PegIntron at 1.5 g/kg (based upon weight at initial visit) plus ribavirin administered subcutaneously once a week for 10 weeks. The data in this study consists of the blood gene expression profiles of twenty genotype 1 hepatitis C patients who were previously untreated, and had no other cause of chronic liver disease. These patients’ blood drawn for analysis on day 1 prior to first injection of Interferon (base line) and at days 3, 6, 10, 13, 27, 42 and 70. According to the HCV-RNA determined by qRT-PCR with a lower limit of detection of 29 IU/ml, there are 11 responders and 9 non-responders until day 70. More details of these biological experiments can be found in reference [16]. Totally, after general pre-processes (gene symbol conversion, missing value filled with zero, duplicate genes’ expression average, interpolation of temporal data), our analyzed gene expression data contains 10342 genes of 20 patients on 8 time points.

A. Comparison of module genes and differentially expressed genes

An natural question is what are differences between the consistent module genes and significantly differentially expression genes reported in previous study [16]. In fact, except a few overlapped genes between them (seeing the genes with red nodes in Fig. 1 (E) and (F)), there are mainly two different parts. One is the connectivity on the known biological network: the consistent module genes are grouped on the network (Fig. 1 (F)); and previously reported DEGs have less direct interactions (Fig. 1 (E)). The other is that the consistent module genes (Fig. 1 (C)-(D)) but not DEGs (Fig. 1 (A)-(B)), can discover differential expression patterns among patients with or without drug responses. The responders and non-responders both can be divided into two sub-groups which have negatively correlated gene expression profiles on those consistent module genes, so that two possible new subtypes of patients may be found which were not shown by previously reported DEGs. These facts support that it is biological meaningful to find the consistent functional group (genes) as the network module in the therapy process, which can lead to discover more response-specific details on the biological network level.

B. The consistent modules representing responsive functions

For all 48 consistent modules and some appearance-consistent modules across all therapy time points, their biological meanings are measured by the P-value of functional (pathway) enrichment analysis in g-Profiler [26]. The modules with P-value no more than 0.05 will be used to build static module network and investigate its dynamic characteristics in the following analysis.

The pathways enriched in modules suggest several pathways would play consistent roles in a therapy (the complete table of all modules and their functional enrichment is not shown here due to page limit). It is not surprising that the pathway of Drug metabolism - other enzymes (KEGG:00983) and Drug metabolism - cytochrome P450 (KEGG:00982) are observed in module \{CYP2C9, CYP3A43, UGT1A10, GSTM5, UGT1A9, GPX3, UGT1A8, CYP2A6, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A1, UGT1A3\}, and pathways of Viral mRNA Translation (REAC:192823) and Viral Protein Synthesis (REAC:192841) are enriched in module \{RPS13, RPS12, RPS11, RPS10, RPS17, RPS16, RPL23A, RPS14, RPL7, RPL27A, RPS18, RPL3, RPL37A, RPS3, RPS2, RPL23, CFL1, TPT1, EIF3H, RPS15, RPL24, RPL5, RPL22, H3F3A, RPL10A, RPL4, RPS24, RPLP0, RPS23, RPS20, RPLP1, RPS9, RPL14, RPL17, RPL10, RPL12, EEF1B2, RPL19, RPL28, RPL38, RPL37, RPL34, RPL35, RPL32, RPL30, RPL31, UBC, EEF1G, PAK2\}. This confirms that our analyzed HCV therapy response has two global characteristics as the input and output of the drug therapy, i.e. viral genetic events are influenced by the drug therapy in a drug-specific metabolic way.

Fig. 1. Differences among module genes and differentially expressed genes
C. The module network rewiring reveals functional relationship in effective therapy response

As discussed in the above subsections, the module genes keep consistent connections during the therapy, and they are involved in several disease-relevant or therapy-relevant pathways. Thus, it is worth to investigate the so-called activity variance of those modules and their rewiring connections along with the drug therapy, which can disclose the biological differences between the therapy responders and non-responders, even among different patients.

Average differential expression values E-score (fold-change here) [16] (Fig. 2 (A)) and newly defined expression-correlation scores EC-score (as (7)) (Fig. 2 (B)) both indicate that a few modules (marked with star in Fig. 2 (A)-(B)) have significant statistic difference between groups of responders and non-responders (t-test with 0.05 as the significance threshold).

More obviously, the interaction weights between module pairs are more powerful to distinguish two groups of patients with a different therapy responses (Fig. 2 (C)). Both responders and non-responders tended to construct close module interactions among the therapy. But clearly shown in Fig. 2 (C), responders have the stronger module connection during the therapy. This means, for response to therapy, functional modules indeed need to communicate at certain time periods of therapy. Otherwise, without necessary genes to shorten/bridge the connections of these modules on biological network, the drug therapy maybe non-effective for patients just like non-responders. Therefore, the interactions between modules and their temporal rewiring actually reflect the responsive relationships among disease-specific and drug-specific functions enriched in network modules.

D. The module network rewiring chain on patient distinguishes responder and non-responder

As the module network rewiring chain can start and end at any pair of check time points, we have tested which time period is the best to classify patients. In the traditional hierarchical analysis in PermutMatrix [27], most responders and non-responders indeed have distinct modules’ connection weights at certain time periods, especially the earlier (first and second) time windows (Fig. 3). Then, we built the decision tree classifier in Weka [28] based on the chain encoding vectors and used leave-one-out validation to evaluate this classification model’s performance. The accuracy and AUC of prediction model are both above 0.85, which demonstrate that module network rewiring chain can be used as spatio-temporal markers for biomedicine study and application. We will conduct further validation tests on such markers for other independent datasets in future. In all, different from previous methods, the proposed MNRA can distinguish the consistent and varying interactions (or relations), and supply the topological indicators of rewiring module network to be some novel and better biomarker candidates than conventional molecular biomarkers. Although MNRA requires more (temporal) samples for each patient to extract network-based diagnostic indicators, these data will be necessary and abundant in future clinical application like personalization of treatment.

IV. CONCLUSION

Different from previous study on the conserved function based on the consistent molecular interactions or dysfunctions with the rewired interactions of molecules behind, we...
studied the rewired associations between the consistent functions based on the module network rewiring. The proposed MNRA lets us analyze dynamic processes on a functional interaction level rather than the molecular interaction level. By using MNRA on the gene expression data from HCV patient receiving Interferon injection, we successfully found several relevant modules and their varying associations which clearly distinguish patients with good therapy response or not. Clearly, MNRA can be straightforwardly applied to other kinds of dynamic processes like cell cycle, cell differentiation and cellular stress response in a similar manner. In addition, MNRA can also be applied to identify the network biomarkers [29], [30] for distinguishing the normals and diseases, and the dynamical network biomarkers for distinguishing the normals and pre-diseases [31].

REFERENCES


