Chapter 5

Integration of Microarray and Metabolic Data

Christina Chan, Zheng Li, and Shireesh Srivastava

Numerous mathematical methods have been developed to reconstruct biological networks, for example gene (1,2), metabolic (3), and protein-protein (4,5) networks, from experimental data. Reconstructing pathways and networks provides a framework for predictive modeling and hypotheses testing to gain more insight into living organisms, disease mechanisms, and targeted therapeutics. A popular method used to reconstruct pathways and networks is Bayesian Network (BN) analysis. We initially evaluated this methodology against the known biochemical networks to gain confidence in the networks that are uncovered from the experimental data using BN analysis. From the metabolic data we inferred the known sub-networks, such as the tricarboxylic acid (TCA) and urea cycles. Extending this methodology to gene networks presents unique challenges because of the size of the gene networks. One of the main shortcomings with BN learning is the computational inefficiency when applied to large number of nodes (genes), such as microarray data. Therefore, we developed approaches that identify a smaller subset of relevant genes (active pathways) for network reconstruction to circumvent the computational inefficiency.

Introduction

Understanding cellular processes is integral to our ability to manipulate cells and identify key variables that may lead to proper function of the cellular or tissue system. The ability to predict cellular responses as a function of the genetic, metabolic, and environmental make-up of the system is important; not only in enabling us to direct proliferation and differentiation pathways *in vitro*, but also in understanding how the changes in these variables may influence their response *in vivo*. An important step in predicting cellular responses is to identify the underlying complex genetic and metabolic networks. However, the determination of such networks experimentally is

a tedious and daunting task, and most of the current experimental analyses have only identified a small subset of possible cellular networks. Another important consideration is that cellular networks change and evolve with time. For example, the regulatory networks of an undifferentiated cell may be different from a differentiated cell, or the network may change during stress, such as during diseased condition or change in environment. Although the experimental identification of cellular networks in such a multitude of conditions, for a variety of cells may be extremely difficult, it is now possible to obtain high throughput information on the metabolic and genomic responses of the cells. Mathematical analyses can be applied to identify or infer complex networks and analyze their systematic properties and behavior (6,7,8). This approach has the potential to facilitate a more complete determination of biochemical and gene regulatory networks as well as their evolution. Reconstructing networks from high throughput data may help in understanding the underlying mechanism(s) behind cellular processes. The biomedical applications of network reconstruction are numerous, ranging from improving understanding of disease mechanisms to identifying effective drug targets.

Recently, a number of computational methods (1-5, 9-12) have been developed to reconstruct networks from experimental data. Among them, the most popular approaches are based on Bayesian Network (BN) analysis (3,5,9,10). The first attempts at network reconstruction using BN analysis endeavored to infer gene regulatory pathways and networks from microarray and simulated data of prokaryotes and yeasts (9,10). More recently, BN analysis has been applied to reconstruct protein signaling network(s) of primary human immune cells (5). Our group used a Bayesian based framework to reconstruct metabolic sub-network structures, e.g., TCA and urea cycles, from hepatocellular metabolic data (3), to confirm the ability of the proposed approach to reconstruct *known* biological networks. This provided some degree of confidence in the *novel* networks that may be inferred from experimental data, such as gene regulatory networks from microarray data. The advantage of the Bayesian framework over other data-driven methods is the ability of the Bayesian approach to perform cause and effect analyses, thus identifying causal relationships. This is accomplished without *a priori* detailed knowledge or assumptions of the biological system and the governing equations, but rather is based upon the concept of conditional probability (9). In addition to expressing causal relationships, the graphical representation and use of probability theory in BN make it amenable to learning incomplete as well as unmeasured data. Furthermore, similar to model-driven methods, the BN approach permits the evaluation of several hypothetical networks using quantitative measures, such as a Bayesian metric score, to access the likelihood of a proposed network structure.

Despite substantial advances and progress in our understanding of biological processes, there are limitations with the existing approaches for reconstructing networks. A major shortcoming of BN learning is the computational inefficiency when applied to

large number of nodes, as is the case with microarray data. A number of approaches have been proposed to alleviate this problem. They either take advantage of the genome-wide interaction data, such as protein-protein and protein-DNA interaction data, or require large amounts of perturbation data to reconstruct an overall molecular network (1,4). However, the availability of such data is limited for mammalian systems. Therefore, we developed approaches to identify a smaller subset of relevant genes (active pathways) for network reconstruction, by capitalizing upon the integration of multi-source and multi-level data to identify the active pathways (13). In this paper, we demonstrate that mathematical analyses can aid in identifying known and potentially *novel* pathways in a cellular system under different experimental conditions. Nonetheless, experimental verification of the *novel* networks would still be required.

Materials and Methods

Materials, Cell Culture, and Assays

The details of the experimental system and measurements are described in (3,13). The modeling approaches discussed in this paper are briefly described below.

Bayesian Networks (BN)

BN are directed acyclic graphs (DAG) whose nodes correspond to variables and whose arcs represent the dependencies between variables. The dependencies are determined by the conditional probabilities of each node x_i, given its parent node p_a, $\Pr(x_i \mid p_a(x_i))$. A BN i) assumes conditional independence, such that each node is independent to its non-descendants, given its parents, in other words, x_i and x_j are conditionally independent to each other given p_a, then

$$Pr(x_i \mid x_j, p_a(x_i)) = Pr(x_i \mid p_a(x_i)) \qquad [1]$$

and ii) consists of the joint distribution defined by a set of variables $\{x_i\}$ as:

$$\Pr(x_1, ..., x_n) = \prod_{i=1}^{N} \Pr(x_i \mid p_a(x_i)) \qquad [2]$$

Inferring the BN from information theory

Metabolic flux data is much smaller in scale than microarray data; nevertheless, applying BN analysis to flux data is still computationally prohibitive. Many algorithms exist that can infer the BN structure; only a few are computationally efficient enough to deal with large datasets. Information theory-based learning algorithm is one such method, which we have applied to the flux data to infer the underlying metabolic regulatory network.

This algorithm has been applied to real-world data with hundreds of variables and records (14), and is described in (3).

Fisher's Discriminant Analysis
Fisher's Discriminant Analysis (FDA) was applied to identify the measured metabolic fluxes that contributed to the separation of different phenotypes (cytotoxic versus nontoxic). FDA identifies the projection axes that maximize the ratio of the between-group and the within-group variations. Details on the FDA algorithm can be found in (15).

Gene Set Enrichment analysis (GSEA) of the gene data
The genomic responses of the cells to the treatments were evaluated with cDNA microarray analyses. The expression levels of 19,522 genes were measured, and the data was analyzed using GSEA. GSEA aims to identify the gene sets whose coordinated change differentiates two phenotypes. The software GSEA-P, from http://www.broad.harvard.edu/gsea/, was used for the GSEA analysis. The gene sets with a high significance of enrichment are considered important in separating the distinct phenotypes.

Integrating the gene expression and metabolic flux profiles
Multi-block partial least squares (MBPLS) is a hierarchical multivariate analysis method (16,17), where the variables are divided into different blocks based upon *a priori* knowledge, for example, according to different stages of an industrial process (16) or different metabolic pathways in a cell (18). Here, the genes are separated into different blocks based upon their functional roles in different pathways. This facilitated the identification of an important block (e.g., a gene set) to a desired dependent variable (e.g., metabolic flux), and then further identified the important genes within the block. Important genes sets were identified by evaluating the weights of each block and the importance of individual genes was identified by evaluating the regression coefficients of the genes within the block. For more details of the MBPLS algorithm, refer to Hwang et al., (2004).

Results and Discussion

Reconstructing sub-networks from metabolic data
Using BN analysis we reconstructed the metabolic sub-networks, namely, the TCA and urea cycles, from the metabolic flux data (3), which we reproduce here in brief. The analysis was able to infer the relationships shown in Figure 5.1B, which compared well with the TCA cycle (Figure 5.1A). The direct connections between the metabolites citrate, α-ketoglutarate, succinyl-CoA and malate, in the TCA cycle were not learned by BN analysis, but rather were identified as being linked to oxidative phosphorylation

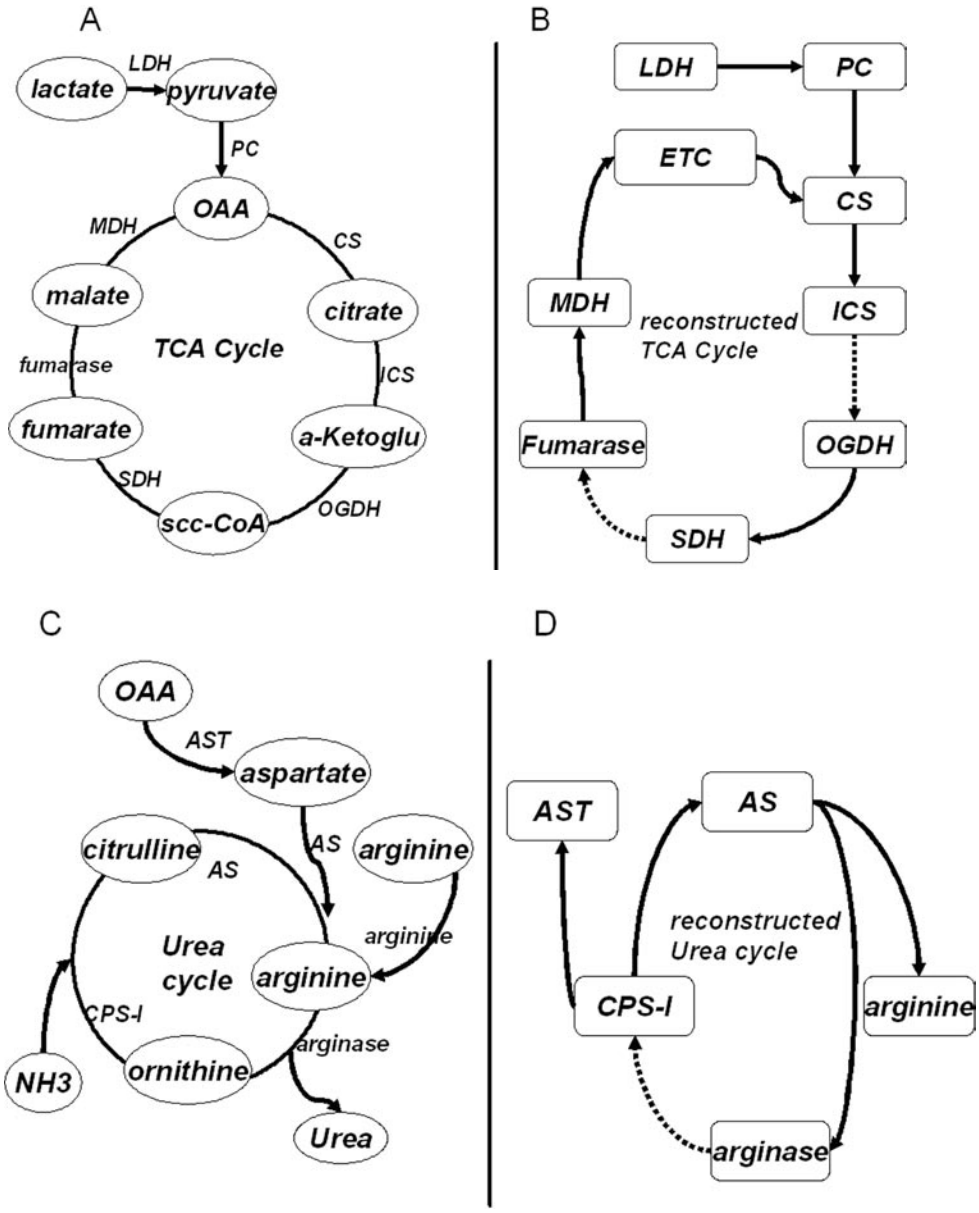

Figure 5.1. Reverse Engineered TCA and urea cycle learned by mutual information based algorithm. A) Actual metabolic network of TCA cycle, each node represents a metabolite and connections (arcs) between nodes represent the metabolic fluxes. B) TCA cycle inferred by Bayesian network analysis, each node represents a flux in Figure 5.1A, each arc between the nodes represent a causal relation between the fluxes, the solid connections were learned and the dashed connections were not learned by Bayesian network analysis. C) Actual metabolic network of urea cycle, each node represents a metabolite and connections between nodes represent the metabolic fluxes. D) Urea cycle inferred by Bayesian network analysis, each arc between the nodes represent a causal relation between the fluxes, the solid connections were learned and the dashed connections were not learned by Bayesian network analysis. OAA: oxaloacetate, AST: aspartate aminotransferase, AS: arginosuccinate synthetase, CPS: Carbarmoyl-p synthetase, PC: pyruvate carboxylase , CS: citrate synthase, ICS: iso-citrate synthase, OGDH: α-ketoglutarate dehydrogenase, SDH: succinate dehydrogenase, MDH: malate dehydrogenase.

and the electron transport chain. This inference is encouraging, since oxidative phosphorylation in which ATP is formed by electron transfer from NADH and $FADH_2$ to O_2, is indeed linked to the TCA cycle. Another important hepatic function is the urea cycle. BN analysis was able to reconstruct many of the reactions in the urea cycle (Figure 5.1D). A comparison of the actual pathways (Figure 5.1C) and the inferred network is shown in Figure 5.1. The connection between arginine and citrulline was not learned by BN analysis. This could be due to the noise in the data or missing data (3).

BN analysis offers the advantage of characterizing the underlying causal structure within the data, unlike correlation-based approaches which cannot identify the causal/ parent variable that is responsible for the observed correlation between two variables (3). Modulating the common cause or parent variable as opposed to the variable(s) that are simply highly correlated will more likely elicit a response in the target variable(s). For example, in our previous paper (15), the aspartate aminotransferase pathway was selected as a potential pathway to optimize or restore urea production. However, in the sub-network inferred by the BN analysis, shown in Figure 5.1D, the argininosuccinate synthetase and argininosuccinase reactions have common causes, the carbamoyl-P-synthetase and ornithine transcarbamylase reaction, suggesting that the argininosuccinate synthetase pathway is relevant to, but not the cause of, urea synthesis. The network reconstructed by the BN analysis suggested that to optimize urea production, the variable(s) which should be modified are carbamoyl-P-synthetase and ornithine transcarbamylase rather than argininosuccinate synthetase. This suggests that ammonia, regardless of its source, combines with HCO_3- to form carbamoyl phosphate, which is the driving force for urea synthesis and the source of argininosuccinate synthetase and aspartate aminotransferase activation. Thus, BN analysis identified novel/parent targets to optimize cellular functioning under altered environmental conditions. These novel targets need to be validated with further experiments.

Identifying the active pathways from gene expression data

As mentioned above, BN analysis is computationally inefficient when used to infer large networks from gene expression data. Therefore, this section describes a hierarchical framework we developed to identify a subset of relevant genes (active pathways) for network reconstruction (15). In brief, the hierarchical framework consisted of three stages. First, discriminant analysis was used to identify the metabolites that were most relevant in differentiating the phenotype of interest. Second, GSEA was applied to the gene expression data to identify the sets of genes that were transcriptionally altered and correlated statistically significantly to the desired phenotype. Third, a multi-block partial least squares analysis (MBPLS) regression model was used to integrate the expression of the enriched gene sets with the metabolic profiles to identify the genes that regulate the metabolic pathways found to be important in separating the phenotype.

Metabolites involved in differentiating the phenotype

We have found that mono- and poly- unsaturated fatty acids, tumor necrosis factor (TNF)-α and their combinations induced a non-toxic phenotype, whereas, the saturated fatty acid palmitate induced a toxic phenotype in HepG2 cells (19) and TNF-α increased this toxicity. For the various conditions, the rates of uptake/release of 27 metabolites, of glucose, fatty acid, and amino acid metabolism were also measured (13). To identify the metabolites responsible for separating the phenotypes (cytotoxic versus non-toxic as defined by the level of lactate dehydrogenase [LDH] release) and highly correlated with the cytotoxic phenotype, FDA and (Pearson's) correlation analysis were applied. FDA found that beta-hydroxybutyrate (BOH), acetoacetate (AcAc), and intracellular triglyceride (TG) accumulation were responsible for separating the cytotoxic palmitate phenotype from the rest of the non-toxic FFA phenotype (13). Correlation analysis identified that BOH and acetoacetate were strongly positively related to the cytotoxicity, while TG accumulation had strong negative correlation (Table 5.2). These results are in agreement with other studies which have identified that increased beta-oxidation is associated with increased reactive oxygen species (ROS) generation (20), while channeling of palmitate to TG reduces its toxicity (21). Therefore, the effect of a gene on these processes would indicate how that gene would affect the cytotoxicity.

Gene sets found enriched by GSEA

With single-gene analysis, one often encounters either a long list of statistically significant genes without any unifying biological theme or the important genes may not meet statistical significance and are thus not selected (22). Interpretation of the former can be overwhelming and ad hoc, and dependent on one's area of expertise. With the latter, since relevant biological differences may oftentimes be modest relative to the noise in the microarray, important genes may be missed by the analysis. To overcome these limitations, GSEA, unlike the single-gene analysis, aims to identify the gene sets whose coordinated changes differentiate phenotypes (22). The gene sets with a high significance of enrichment are considered important in separating the distinct phenotypes. GSEA was applied to identify the gene sets/processes most associated with the palmitate-induced cytotoxicity (13). Of the gene sets evaluated, those related to oxidative stress, such as ROS, glutathione, oxidative phosphorylation and electron transport chain (ETC) were significantly enriched. We have previously shown the importance of ROS generation by palmitate in the toxicity (19). Thus, the GSEA identified a potential mechanism for the observed toxicity (13).

Pathways involved in inducing the phenotype

MBPLS analysis was applied to identify the gene sets and the underlying genes that were strongly related to the cytotoxicity, by evaluating the regression coefficients

Table 5.1 Correlation between the metabolite production/ uptake and lactate dehygrogenase (LDH) release.

Metabolite	Correlation Coefficent
Triacylglycerol Synthesis	-0.52324
Glycerol	-0.48656
Glucose	-0.47572
Lactate	-0.38699
Glutamate Uptake	-0.19394
Cysteine Uptake	-0.18678
Ornithine Uptake	-0.16018
NH_3 Uptake	-0.13579
Glycine Uptake	-0.11056
Aspartate Uptake	-0.10666
Arginine Uptake	-0.09454
lsoleucine Uptake	-0.07904
Histidine Uptake	-0.07651
Alanine Uptake	-0.07491
Tyrosine Uptake	-0.06592
Lyssine Uptake	-0.05902
Valine Uptake	-0.05244
Glutamine Uptake	-0.00651
Phenylalanine Uptake	-0.00258
Leucine Uptake	0.017462
Threonine Uptake	0.043442
Proline Uptake	0.130492
Serine Uptake	0.157344
Fatty Acid Uptake	0.173699
O_2 In	0.574845
Acetoacetate	0.853407
Betahydroxybutyrate	0.935881

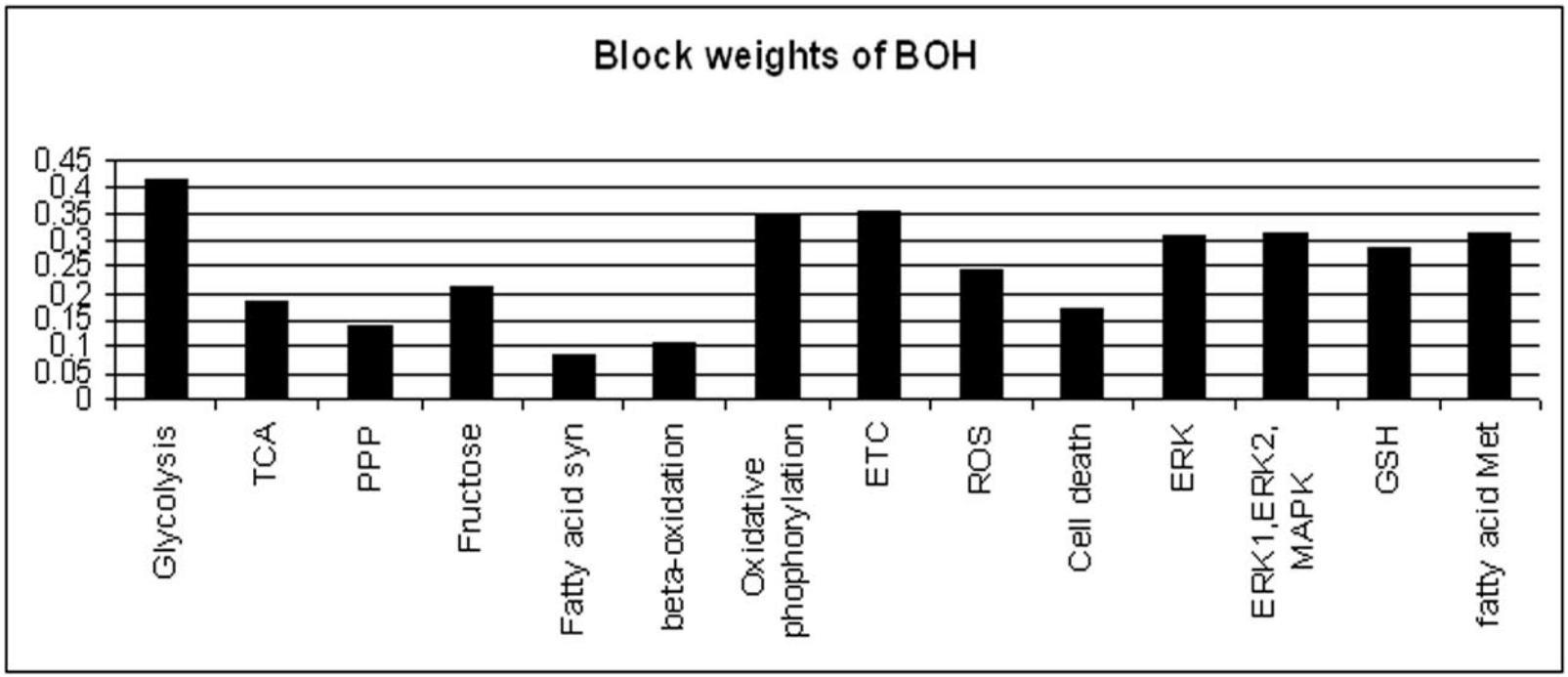

Figure 5.2. Multi-block partial least squares (MBPLS) regression coefficients of different gene sets for beta-hydroxybutyrate (BOH).

Table 5.2 Genes Selected by Hierarchical Approach. Access#: Accession Number, RC(BOH): regression coefficient for genes to predict beta-hycroxybutyrate (BOH), abs(RC): absolute value of the regression coefficients.

Access #	RC (BOH)	abs (RC)	Gene Name
AA425826	-4.24E-02	4.24E-02	mitogen-activated protein kinase kinase 2 (MAP2K2)
N56898	3.89E-02	3.89E-02	glutathione S-transferase M5 (GSTM5)
AA664101	3.64E-02	3.64E-02	aldehyde dehydrogenase 1 family, member A1 (ALDH1A1)
AA489666	-3.53E-02	3.53E-02	neutrophil cytosolic factor 1 (47kDa, chronic granulomatous disease, autosomal 1) (NCF1)
AA406536	3.33E-02	3.33E-02	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase) (NDUFS1)
AA680322	3.15E-02	3.15E-02	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa (NDUFA4)
AA455235	-2.93E-02	2.93E-02	aldehyde dehydrogenase 1 family, member A3 (ALDH1A3)
AA43630	2.93E-02	2.93E-02	aldehyde dehydrogenase 3 family, member B2 (ALDH3B2)
AA136566	-2.77E-02	2.77E-02	forkhead box M1 (FOXM1)
AA664007	-2.75E-02	2.75E-02	serine/threonine kinase 25 (STE20 homolog, yeast) (STK25)
T72259	2.72E-02	2.72E-02	cytochrome P450, subfamily IIA (phenobarbital-inducible), polypeptide 7 (CYP2A7)
W88587	-2.41E-02	2.41E-02	GNAS complex locus
AA256532	-2.41E-02	2.41E-02	insulin-like growth factor 1 receptor
T58873	-2.36E-02	2.36E-02	FOS-like antigen 2
AA486570	-2.32E-02	2.32E-02	glutathione S-transferase M4 (GSTM4), transcript variant 3
H59758	2.31E-02	2.31E-02	v-raf murine sarcoma 3611 viral oncogene homolog 1 (ARAF1)
N30404	-2.26E-02	2.26E-02	copper chaperone for superoxide dismutase (CCS)
AA035384	2.21E-02	2.21E-02	succinate dehydrogenase complex, subunit D, integral membrane protein (SDHD)
AA055585	2.20E-02	2.20E-02	core promoter element binding protein (COPEB)
T52484	2.20E-02	2.20E-02	nerve growth factor, beta polypeptide (NGFB)

of the individual genes and gene sets (13). Glycolysis, oxidative phosphorylation, and ETC had very high block weights. The selection of these gene sets indicates that the alterations in the energy transduction pathways are strongly related to the toxicity caused by free fatty acids (FFA). In addition, extracellular related kinase (ERK), ERK1/2 MAPK, ROS, glutathione, and fatty acid metabolism gene sets also had high weights, indicating that alterations in fatty acid metabolism, ERK/MAPK signaling and redox state of the cells play important role in the FFA toxicity. Among the genes, Mitogen-activated protein kinase kinase 2 (MAP2K2) had the highest negative regression coefficient to BOH (Table 5.2), suggesting potentially protective role for this kinase. MAP2K2 is an upstream MAP kinase which phosphorylates and activates extracellular signaling kinases (ERKs) ERK2 and ERK3. ERK2 and 3 regulate diverse cellular processes such as growth and differentiation. Because multiple processes could be affected by altering the levels or activities of MAPKs, MAP2K2 was not tested further. Specifically, we looked into those genes that had very high positive regression coefficients as their roles could be easily evaluated using pharmacological inhibitors. Glutathione-S-transferase M5 (GSTM5) had the highest positive regression coefficient to BOH (Table 5.2). However, its basal levels in hepatocytes is very low (24) and its purported role is in the detoxification of harmful aldehydes. This suggested that inhibiting GSTM5 may not be very effective in reducing toxicity. Similarly, while the aldehyde dehydrogenase 1 family member A1 (ALDH1A1) had the second highest regression coefficient, and another aldehyde dehydrogenase (ALDH1A3) had very high negative regression coefficient (Table 5.2). This suggested that to test the roles of these genes, alterations in the specific isoforms would be required and the application of inhibitors of ALDH, which are not specific to any isoform, would lead to confounding results. On the other hand, two isoforms of NADH dehydrogenases had very high positive regression coefficients to cytotoxicity. While their individual coefficients were slightly smaller than glutathione S-transferase M5 (GSTM5) as well as aldehyde dehydrogenase 1A1 (ALDH1A1), their combined coefficients were greater than any other gene(s). Additionally, NADH dehydrogenases have also been suggested as a major source of ROS in the cells (23). This indicated that inhibition of NADH dehydrogenases could reduce the ROS levels, and possibly, the LDH release. Because both isoforms had high positive coefficients, the roles of these NADH dehydrogenases could be easily tested using pharmacological inhibitor. As shown in Table 5.3, the ROS level and LDH release were significantly reduced in the presence of rotenone, an inhibitor of NADH dehydrogenase. Thus the predicted role of NADH dehydrogenase in inducing the cytotoxic phenotype of palmitate was experimentally validated.

In summary, integrating the metabolic and genetic profiles can identify a smaller subset of genes relevant to the phenotype of interest, see Table 5.2. This smaller subset of genes can then be subsequently subjected to BN analysis for network reconstruction with reduced computational cost as described above for the metabolic network reconstruction.

Table 5.3. Relative reactive oxygen species (ROS) and absolute lactate dehydrogenase (LDH) release in the presence of palmitate and palmitate + rotenone

	ROS	LDH
Control	1 ± 0.11	1.07 ± 0,4
Palmitate	1.95 ± 0.1**	4.67 ± 0.79**
Palmitate/Rotenone	1.06 ± 0.14	2.81 ± 0.12**
** P<0.01 with T-test		

Acknowledgements

This work is supported in part by the National Science Foundation (BES 0222747, BES 0331297, and 0425821), the National Institute of Health (R01 GM079688-01), the Environmental Protection Agency (RD83184701-0) and the Whitaker Foundation.

Correspondence

All correspondence should be addressed to:

Christina Chan
Michigan State University
Department of Chemical Engineering and Materials Science
1257 EB
East Lansing, MI 48824
Phone: (517) 432-4530
Fax: (517) 432-1105
e-mail: krischan@egr.msu.edu

References

1. di Bernardo, D. Thompson MJ, Gardner TS, Chobot SE, Eastwood EL, Wojtovich AP, Elliott SJ, Schaus SE, Collins JJ. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. Nature Biotechnology 2005;23:377-383.

2. Basso K., Margolin A., Stolovitzky G., Klein U., Dalla Favera R., and Califano A., Reverse engineering of regulatory networks in human B cells.

Nature Genetics 2005;37(4): 382-90.

3. Li, Z., and Chan, C., Inferring pathways and networks with a Bayesian framework. FASEB J 2004;18(6): 746-8.

4. Ideker, T., Ozier, O., Schwikowski, B. & Siegel Andrew, F. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 2002;18 Suppl 1:S233-40.

5. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A. , Nolan, G.P. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Science 2005; 308:523-529.

6. Barabási A.-L. &. Oltvai. Z. N. Network biology: understanding the cell's functional organization. Nature Reviews Genetics 2004;5:101-113.

7. Jeong H., Mason S.P., Barabási A.-L., and Oltvai Z.N., Lethality and centrality in protein networks, Nature 2001;411: 41-42.

8. Newman M. E. J. The structure and function of complex networks. SIAM Review 2003;45(2):167–256.

9. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. Journal of Computational Biology 2000;7(3-4):601-20.

10. Pe'er, D., Regev, A., Elidan, G. and Friedman, N. Inferring subnetworks from perturbed expression profiles, Bioinformatics, 2001;17, Sup l.1: s215-s224.

11. Liang, S., Fuhrman, S. and Somogyi, R. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures, Pacific Symposium on Biocomputing, 1998;3:18 -29.

12. Chen, T., He, H. L., and Church, G.M. Modeling Gene Expression with Differential Equations, Pacific Symposium on Biocomputing, 1999; 4:29-40.

13. Li Z., Srivastava S., Mittal S., Norton P., Resau J., Haab B., Chan C. A hierarchical approach to identify pathways that confer cytotoxicity in HepG2 cells from metabolic and gene expression profiles, BMC Systems Biology, 2007; accepted.

14. Cheng, J., Kelly, J., Bell, D.A. and Liu, W. Learning belief networks from data: An information theory based approach, Artificial Intelligence Journal, 2002;137: 43-90.

15. Chan, C., Hwang, D. H., Stephanopoulos, G. N., Yarmush, M. L., and Stephanopoulos, G., Application of Multivariate Analysis to Optimize Function of Cultured Hepatocytes, Biotechnology Progress, 2003;19: 580-598.

16. MacGregor, J.F., Jaeckle, C., Kiparissides, C. and Koutoudi, M. Process Monitoring and Diagnosis by Multi-Block PLS Methods. AIChE Journal, 1994;40 (5): 826-838.

17. Lopes J.A., Menezes J.C., Westerhuis J.A., Smilde A.K.. Multiblock PLS analysis of an industrial pharmaceutical process. Biotechnololgy and Bioengineering, 2002;80(4):419-27.

18. Hwang D, Stephanopoulos G, Chan C. "Inverse modeling using multi-block PLS to determine the environmental conditions that provide optimal cellular function" Bioinformatics, 2004; 20(4):487-99.

19. Srivastava S., and Chan C. Hydrogen peroxide and hydroxyl radical mediate palmitate-induced cytotoxicity to hepatoma cells: relation to mitochondrial permeability transition. Free Radical Research 2007;41(1): 38-49.

20. Sanyal AJ, Campbell-Sargent C, Mirshahi F, Rizzo WB, Contos MJ, Sterling RK, Luketic VA, Shiffman ML, Clore JN. Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. Gastroenterology, 2001;120(5):1183-92.

21. Listenberger LL, Han X, Lewis SE, Cases S, Farese RV Jr, Ory DS, Schaffer JE. Triglyceride accumulation protects against fatty acid-induced lipotoxicity. Proc Natl Acad Sci U S A, 2003;100(6):3077-82.

22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome- wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

23. Moller IM. Plant mitochondria and oxidative stress: Electron Transport, NADPH Turnover, and Metabolism of Reactive Oxygen Species. Annual Review of Plant Physiology and Plant Molecular Biology, 2001;52:561-591.

24. Takahashi Y, Campbell EA, Hirata Y, Takayama T, Listowsky I. A basis for differentiating among the multiple human Mu-glutathione S-transferases and molecular cloning of brain GSTM5. Journal of Biological Chemistry, 1993; 268(12):8893-8.