



Session 6.3

Computational toxicology

Comparative Analysis of Gene Networks at Multiple Doses and Time Points in Livers of Rats Exposed to Acetaminophen

Hiroyoshi Toyoshiba¹, Hideko Sone¹, Fred Parham², Richard D. Irwin², Gary A. Boorman² and Christopher J. Portier²

¹National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan; ²National Institute of Environmental Health Sciences, RTP, NC, USA

Summary

Gene interaction networks (GINs) have been developed as a potential tool to understand biological mechanisms or responses to chemical toxicity. Our recent studies developed two algorithms to identify and quantify gene interaction networks. One is to identify networks based on a statistical criteria and the other is to test hypothetical networks using microarray data. These methods were combined with clustering analysis, networks were clustered and father analysis was performed to quantify linkages in the networks and to follow microarray change over time and dose levels in the model of rat exposed to acetaminophen. These analysis methods could be used for risk assessment and could improve the mechanistic understanding of chemical toxicity.

Keywords: gene interaction networks, computational biology, risk assessment, microarray

Introduction and questions

Several methods of microarray data analysis have been developed in order to gain a mechanistic understanding of chemical toxicity. Clustering analysis (Bharadwaj et al., 2005; Hamadeh et al., 2002; Pan et al., 2003) has become a powerful tool to characterise chemical toxicity through the patterns of expression levels of thousands of genes. Recently, a number of approaches have tried to understand the mechanisms of chemical toxicity by means of pathway analysis or linkage analysis between genes. Gene interaction networks (GINs) have been developed to analyse links between genes to understand the biological mechanisms or responses to chemical toxicity. Several methods have been developed under different mathematical frameworks, i.e. ordinary differential equations (Chen et al., 1999; Moles et al., 2003) and Boolean networks etc. Among these methods, the Bayesian networks (Friedman et al., 2000) approach is a powerful method by

which acyclic directed networks are used to characterise networks. We have recently developed two algorithms based on statistical criteria and statistical quantification of linkages in GINs. Both methods were developed based on Bayesian networks: the TAO-Gen (Theoretical Algorithm for identifying Optimal Gene interaction networks) algorithm (Yamanaka et al., 2004) is used to identify the best GIN based on a statistical criteria, and a statistical algorithm using log-linear function form (Toyoshiba et al., 2004) is used to quantify linkages between genes. In this manuscript, father analysis of the network structure was introduced by applying the method to data measured from livers of rats 6, 24 and 48 hours after exposure to 50 mg/kg, 150 mg/kg and 1500 mg/kg of acetaminophen. These nine data sets were acquired with 10 replicates. For each of the data sets, the best network was estimated by TAO-Gen algorithm. The networks were clustered into two groups by k-mean clustering. The clustering analysis separated the networks into the high dose group (1500

mg/kg) and low dose group (50 mg/kg and 150 mg/kg). Using our method (Toyoshiba et al., 2004), each linkage in the networks was quantitatively analysed to identify whether the linkage was an activator or inhibitor. The results of the analysis showed that the network may change with dose level and that the relationship of the linkage could change over the different time points.

This method offers a new approach to analyse GINs quantitatively and is capable of improving dose-response analysis and risk assessments. Also, the method developed here could contribute to reduction of the use of animals in experiments in the future. For example, the method showed the clustering networks and combination of network analysis and clustering analysis could provide pattern recognition as well as linkage information. Chemicals with the same network structure could be predicted to have the same spectrum of effects on humans and animals. Hence, studies on a specific chemical could be used to predict the toxicity of other chemicals with the same network structures.

Materials and methods

Estimation procedure

Let us define the GINs. $G = [G_1, G_2, \dots, G_P]$ represents P random variables associated with P genes with g_i ($i=1, 2, \dots, P$) as the actual observation of G_i . Let g_i be $n \times 1$ matrix denoting actual n samples from gene G_i . Log-linear functional form is defined as the following,

$$\log(g_j) = \sum_{i=1}^P I_{ij} \cdot \beta_{ij} \cdot \log(g_i) + \varepsilon_j \quad (1)$$

where I_{ij} represents an indicator function having 1 if there exists a linkage from gene G_i to gene G_j and 0 if there is no linkage. β_{ij} is

a coefficient in the log-linear regression and finally, ε_j is assumed to follow the normal distribution $N(0, \sigma_j^2)$. Hence, the model has I_{ij} , β_{ij} and σ_j^2 as the parameters and the first two parameters are particularly interesting. The indicator functions can determine the network structure and the coefficient β_{ij} is obtained by the mean of the posterior distribution which tells whether the linkage is statistically significant (Toyoshiba et al., 2004). A matrix consisting of I_{ij} called transition matrix determines the structure of the network and a matrix consisting of β_{ij} is called parameter matrix. The transition matrix is estimated by TAO-Gen algorithm (Yamanaka et al., 2004). Once the transition matrix is given, the parameter matrix can be estimated with the posterior distribution of β_{ij} (Toyoshiba et al., 2004). The method tells whether the linkage is an activator or inhibitor and also the level of significance. If a gene has a regulatory effect on G_i , then that gene is referred to as a ‘‘Parent of G_i ’’ and we refer to it as belonging to the set $\text{Pa}(G_i)$.

Clustering networks

The transition matrix can be considered to be the single sequence consisting of two elements 0 and 1. The last element of the i^{th} row precedes the first element of the $i+1^{\text{th}}$ row. Hence, the network structure can be summarised as a single sequence. Clustering analysis can be performed with these sequences to elucidate similarities or differences between networks. The distance between two sequences was defined as the sum of the absolute difference of each element of the two sequences. K-means clustering was performed and the networks were clustered into two groups and two centred networks were obtained.

Quantification of the centred networks

After estimating two centred networks, each path in the centred networks was quantified by Markov Chain Monte Carlo sam-

Tab. 1: Seventeen (17) genes used to develop GINs to understand how the apoptosis and oxidative stress genes are linked after exposed to Acetaminophen (APAP).

	Abbreviation	Clone ID	Description
Apoptosis	MADH2	AA818109	<i>Rattus norvegicus</i> Smad2 protein (Smad2) mRNA, complete cds
	NFKB1	AA858801	<i>Rattus norvegicus</i> nuclear factor kappa B p105 subunit mRNA, 3' end
	AKT1	AA858883	Rat mRNA for RAC protein kinase alpha, complete cds
	TP53	AA875052	Rat mRNA for nuclear oncoprotein p53
	PCNA	AA924358	Rat proliferating cell nuclear antigen (PCNA/cyclin) mRNA, complete cds
	SRC	AA926101	ESTs, Highly similar to NEURONAL PROTO-ONCOGENE TYROSINE-PROTEIN KINASE SRC (<i>M.musculus</i>)
	PPARG	AI111890	<i>Rattus norvegicus</i> mRNA for PPAR-gamma protein, complete cds
Oxidative stress	MAPK3	AA875555	<i>Rattus norvegicus</i> microtubule-associated protein-2 kinase mRNA, partial cds
	GSK3B	AA964758	<i>R. norvegicus</i> mRNA for tau protein kinase I
	MAPK1	AI058662	ESTs, Highly similar to Phosphorylated Map Kinase Erk2 (<i>R.norvegicus</i>)
	CDKN1B	AI071529	<i>Rattus norvegicus</i> mRNA for p27, complete cds
	GRB2	AI071530	Rat mRNA for Ash-m, complete cds
	NFE2L2	AI071579	<i>Rattus norvegicus</i> NF-E2-related factor 2 mRNA, complete cds
	ITGB1	AA819911	<i>Rattus norvegicus</i> integrin beta-1 subunit (integrin beta-1) mRNA, complete cds
Significant changed genes	CASP2	AA956222	<i>Rattus norvegicus</i> caspase 6 (Mch2) mRNA, complete cds
	MT1a	AA900218	Rat metallothionein-i (mt-1) mrna
	HMOX1	AA874884	Rat heme oxygenase gene, complete cds



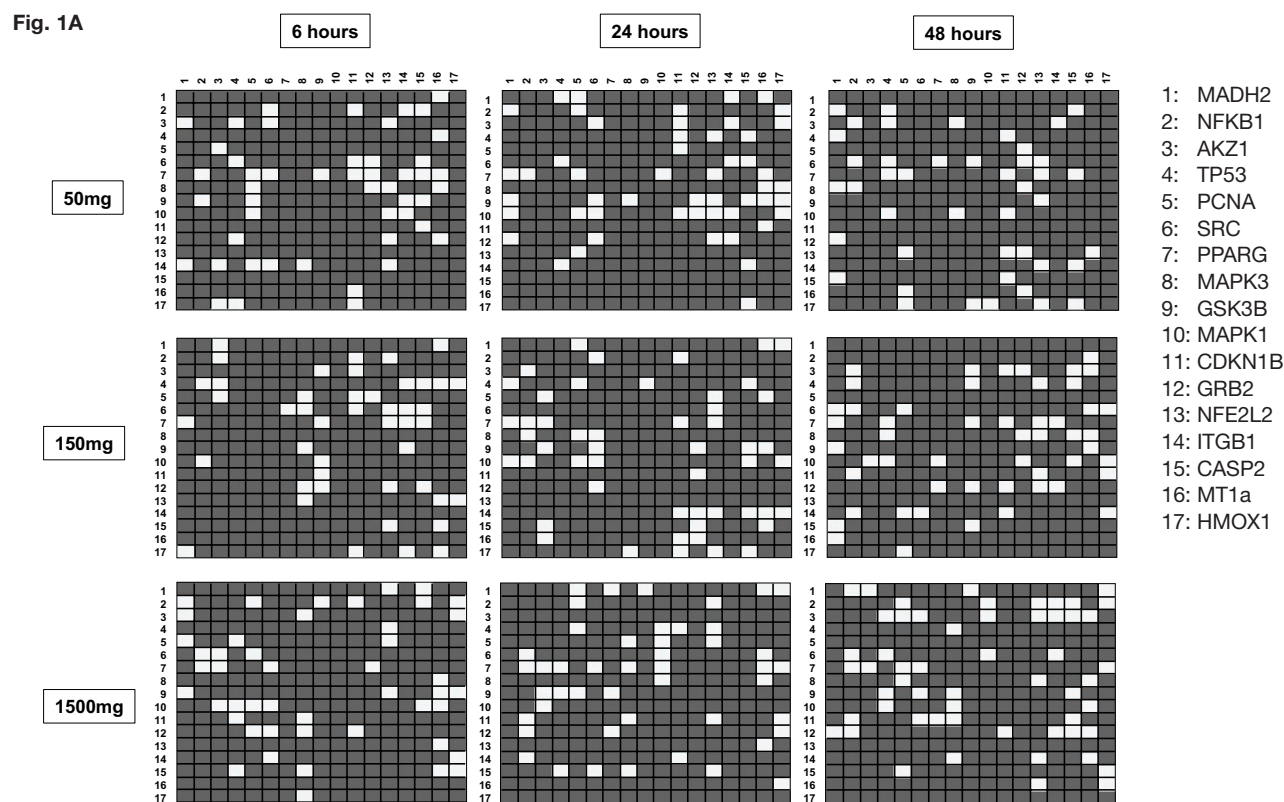
pling methods as described (Toyoshiba et al., 2004). The quantification defines the statistical significance of the linkages between the different genes and the quality of each linkage (activator or inhibitor).

Results

Seventeen genes (MADH2, NFKB1, AKT1, TP53, PCNA, SRC, PPARG, MAPK3, GSK3B, MAPK1, CDKN1B, GRB2, NFE2L2, ITGB1, CASP2, MT1a, HMOX1) from a single gavage dose study in Fischer rats at multiple doses and multiple times after exposure to acetaminophen were identified as potentially belonging to a network governing the liver response (tab. 1).

The genes can be divided into three basic groupings; apoptosis (Apt) related genes (MADH2, NFKB1, AKT1, TP53, PCNA, SRC, PPARG), oxidative stress (Ost) related genes (MAPK3, GSK3B, MAPK1, CDKN1B, GRB2, NFE2L2, ITGB1) and other (Oth) significantly elevated genes (CASP2, MT1a, HMOX1). CASP2 was selected because of its known role in

stimulating apoptosis (Harvey et al., 1997), activating HMOX1 (Panahian et al., 1999) and protecting against oxidative stress (Beattie et al., 2005). The TAO-Gen algorithm was applied to the data for each dose (50, 150 and 1500 mg/kg) at each follow-up time (6, 24 and 48 hours) resulting in nine separate models (fig. 1A). Significance for each linkage in the TAO-Gen algorithm was assessed using Markov Chain Monte Carlo and only the significant linkages were retained. Figure 1A is a graphical representation of these linkages. In this representation using the matrix form, paths lead from genes in columns to genes in rows. For example, in the graph of 50 mg/kg at 6 hours, AKT1 in the third row has paths to MADH2, TP53, SRC and NFE2L2. These linkages are highlighted in yellow in the 1st, 4th, 6th and 13th columns of the third row. To compare these networks, k-means clustering (k=2) was applied to nine transition matrices to determine which are closer structures. The resulting two clusters split the networks by dose and not by time (fig. 2). The first cluster contained all times for the low (50 mg/kg) and middle (150 mg/kg) dose groups, whereas the other cluster contained all times for the high dose (1500 mg/kg) group. The two centred



$$\log(g_j) = \sum_{i=1}^{17} t_{ij} \cdot p_{ij} \cdot \log(g_i) + \epsilon_j$$

Fig. 1B Linkages from AKT1 to other genes in the low dose groups at 50, 150 and 1500 mg/kg/day exposure.

Fig. 1A: Nine estimated GINs by dose and exposure time in male rats exposed to acetaminophen where yellow boxes indicate linkages between genes and green indicates no linkage with the parent gene on the vertical axis and the child on the horizontal axis.

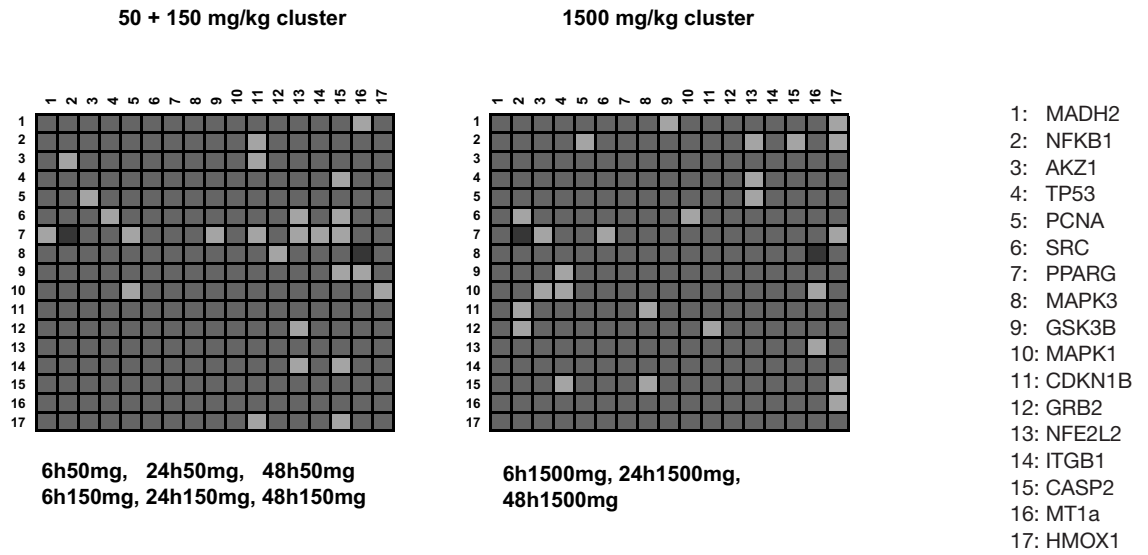


Fig. 2: Two networks resulting from k-means clustering of the estimated networks shown in fig. 1. The left cluster consists of all data for the 50 and 150 mg/kg dose groups and the right cluster consists of all data for the 1500 mg/kg dose group.

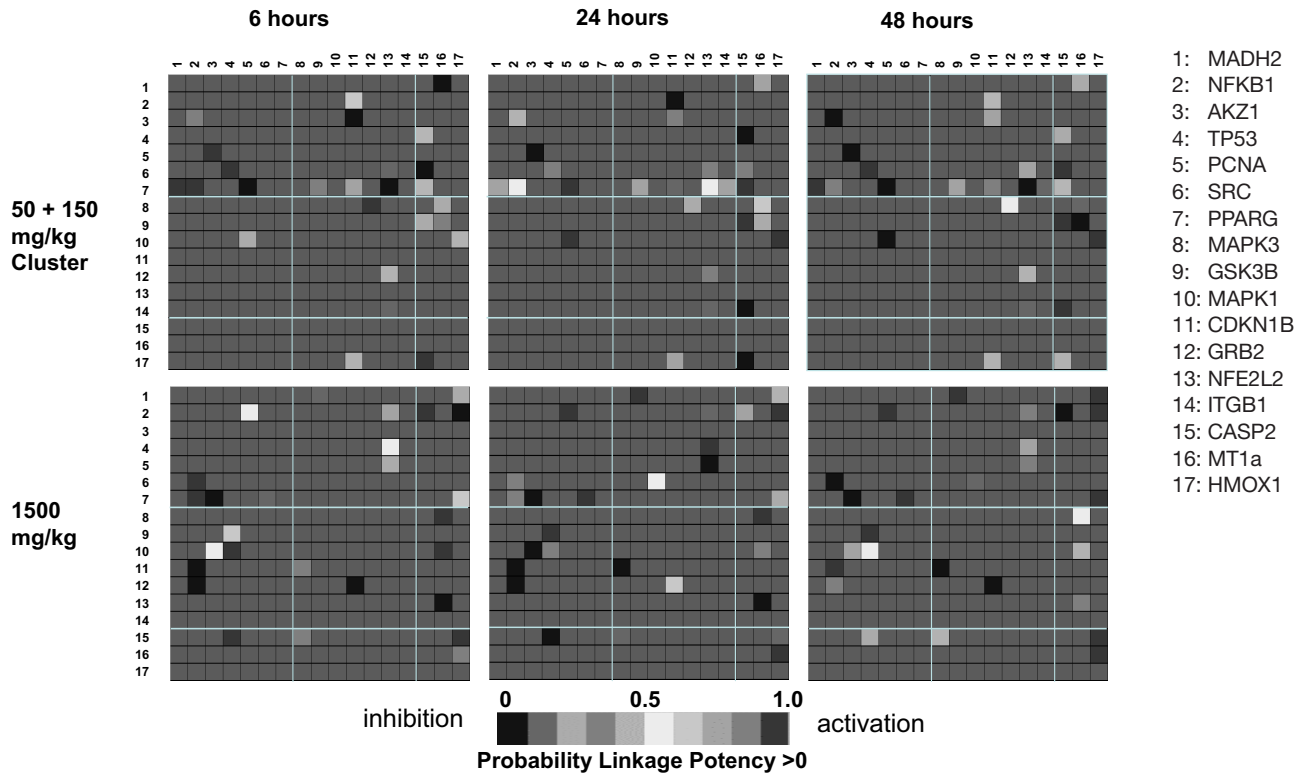


Fig. 3: Significance of the quantified paths in the two centred networks obtained from clustering (fig. 2) after network clustering at each time point. Red indicates significant activation and blue significant inhibition for the path.



networks illustrate significant differences with only two common linkages (shown in red in fig. 2).

To determine the role and the significant level of the linkages between genes in figure 2, a statistical algorithm (Toyoshiba et al., 2004) was applied. To investigate changes over time points, two centred networks were quantified at different time points. The lowest dose (50 mg/kg) and the middle dose (150 mg/kg) data were combined at each time point. Significant levels of each linkage are illustrated in fig. 3. Deeper red and blue represent more significant activator and inhibitor respectively. For example, the linkage from SRC to CASP2 is an inhibition at the earliest time point, changing to activation at later time points. By this quantification analysis, changes in the roles of the linkages in the networks could be illustrated over the time points of the experiment.

Discussion

This study illustrates a new approach with a potential for gaining a mechanistic understanding of chemical toxicity, which is a step further than mere network or clustering analysis. Network type analysis could provide insight into the mechanisms of toxicological responses, but it would be very difficult to interpret if the linkages are increased and complicated. On the other hand, clustering type analysis is useful to compare similarities or differences but does not provide insight into mechanistic changes.

The approach used in this study could provide not only insight into the mechanisms but could also be used to compare networks. In the model experiment with acetaminophen, the high dose group was clustered separately from the low dose group. This might suggest that the network structure at high dose is different from that at the low dose. Also, the quantification analysis could demonstrate how the qualities of the linkages change over time. These results suggest that the network structure may differ between different doses of a substance and that the quality of the linkages may change in the course of exposure to the substance.

To demonstrate another way to use this technique, the transition matrix and the parameter matrix were estimated under low dose acetaminophen, and it was evaluated how well the data from the high dose acetaminophen treatment fitted the model estimated at low dose. Genes that do not fit the model estimated with the low dose data might indicate that the substance has different effects at the higher dose or they might indicate that different linkages exist at the high dose. This information contributes to the understanding of the mechanism of toxicity.

Network structure could be used to characterise chemicals. Studies on one specific chemical of a group could be used to predict the toxicity of other chemicals in the same group. This is an obvious contribution to the 3Rs and may help us to understand the mechanisms of chemical toxicity at lower cost.

References

- Beattie, J. H., Owen, H. L., Wallace, S. M., et al. (2005). Metallothionein overexpression and resistance to toxic stress. *Toxicol. Lett.* 157(1), 69-78.
- Bharadwaj, L., Dharmi, K., Schneberger, D., et al. (2005). Altered gene expression in human hepatoma HepG2 cells exposed to low-level 2,4-dichlorophenoxyacetic acid and potassium nitrate. *Toxicol. In Vitro* 19(5), 603-619.
- Chen, T., He, H. and Church, G. (1999). Modeling gene expression with differential equations. *Proc. Pac. Symp. On Biocomputing* 4, 29-40.
- Friedman, N., Linial, M., Nachman, I. and Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 7(3-4), 601-620.
- Hamadeh, H. K., Knight, B. L., Haugen, A. C. et al. (2002). Methapyrilene toxicity: anchorage of pathologic observations to gene expression alterations. *Toxicol. Pathol.* 30(4), 470-482.
- Harvey, N. L., Butt, A. J. and Kumar, S. (1997). Functional activation of Nedd2/ICH-1 (caspase-2) is an early process in apoptosis. *J. Biol. Chem.* 272(20), 13134-13139.
- Moles, C. G., Mendes, P., Banga and J. R. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* 13(11), 2467-2474.
- Pan, H., Zhu, J. and Han, D. (2003). Genetic algorithms applied to multi-class clustering for gene expression data. *Genomics Proteomics Bioinformatics* 1(4), 279-287.
- Panahian, N., Yoshiura, M. and Maines, M. D. (1999). Overexpression of heme oxygenase-1 is neuroprotective in a model of permanent middle cerebral artery occlusion in transgenic mice. *J. Neurochem.* 72(3), 1187-1203.
- Toyoshiba, H., Yamanaka, T., Sone, H. et al. (2004). Gene interaction network suggests dioxin induces a significant linkage between aryl hydrocarbon receptor and retinoic acid receptor beta. *Environ. Health Perspect.* 112(12), 1217-1224.
- Yamanaka, T., Toyoshiba, H., Sone, H. et al. (2004). The TAO-Gen algorithm for identifying gene interaction networks with application to SOS repair in *E. coli*. *Environ. Health Perspect.* 112(16), 1614-1621.

Correspondence to

Christopher Portier
 Laboratory of Molecular Toxicology
 National Institute of Environmental Health Sciences
 PO Box 12233 (MD A3-02)
 Research Triangle Park
 NC, 27709, USA
 e-mail: portier@niehs.nih.gov