Application of Structural Equation Modeling for Inferring Toxicity-Dependent Regulation in Human Embryonic Stem Cells

Sachiyo Aburatani,
Computational Biology Research Center
National Institute of AIST
Tokyo, Japan
s.aburatani@aist.go.jp

Wataru Fujibuchi
Center for iPS Research and Application
Kyoto University,
Kyoto, Japan
w.fujibuchi@cira.kyoto-u.ac.jp

Abstract—**Chemical toxicity threat our daily health, especially for embryos. Reveling toxicity-dependant regulation in human embryo is one of the effective approaches to prevent some chemical effects. In previous study, we developed a network inference approach, based on Structural Equation Modeling (SEM). In this study, we improved the SEM approach and applied this enhanced approach to expression profiles in human embryonic stem cells exposed to various chemicals. The inferred gene regulatory models among neurodevelopment related genes clarify the differences between chemicals, and the network shapes reflected the features of chemical toxicities. The effects of Acrylamide toxicity finally aggregated to a neuronal cell-related gene, even though Diethylnitrosamine disturbed normal cell differentiation-related genes. Furthermore, gene regulatory network with Thalidomide was complicated, but embryonic development-related genes were estimated as the finally effected genes by Thalidomide toxicity.**

Keywords-Structural Equation Modeling; Embryonic Stem Cell; Gene Regulatory Network; Chemical Toxicity.

I. INTRODUCTION

We are exposed to many chemicals in our daily life, and chemical toxicity is known to exert harmful effects on human health. Actually, some diseases are caused by exposure to environmental pollution [1][2], including chemicals such as methylmercury [3][4], and so on. Furthermore, some chemical toxins are threatening, since they can cause abnormal cell differentiation in embryos [5][6][7]. Clarifying the details of the toxic stress response in embryonic cells is crucial for the prevention of harmful chemical effects [8][9].

To gain a better understanding of the role of the toxic stress response, a gene regulatory network is useful. With the gene expression information, the regulatory networks among the genes can be inferred. Various algorithms, including Boolean and Bayesian networks, have been developed to infer complex functional gene networks [10][11]. In our former investigation, we developed an approach based on graphical Gaussian modeling (GGM). The GGM approach is combined with hierarchical clustering for calculations with massive amounts of gene expression data, and we can infer the huge network among all of the genes by this approach [12][13]. However, GGM infers only the undirected graph, whereas the Boolean and Bayesian models infer the directed graph, which shows causality.

Recently, we developed a new statistical approach, based on Structural Equation Modeling (SEM) in combination with factor analysis and a four-step procedure [14]. This approach allowed us to reconstruct a model of transcriptional regulation that involves protein-DNA interactions, from only the gene expression data. Furthermore, SEM approach allows us to strictly evaluate the inferred model by using fitting scores. The SEM approach is available for the detection of causality among selected genes, as the linear relationships between genes are assumed to minimize the difference between the fitted model covariance matrix and the calculated sample covariance matrix [15][16][17].

Here, we applied the SEM approach to the limited expression data of neurodevelopment related genes in human embryonic stem cells exposed to various chemicals. The chemicals were considered to be toxic and to adversely affect the neurodevelopment related genes. Thus, inferring the gene regulatory network among neurodevelopment related genes will help to elucidate the toxic stress response in the human embryo. Since the regulatory interactions among the genes were unclear, a new approach for assuming an initial model should be developed for the application of SEM. In this study, we used an improved SEM approach that includes a new method for constructing a preliminary initial model, in the absence of known regulatory interactions. The resulting gene expression data clarified the chemical-specific interactions among the neurodevelopment related genes.

II. MATERIAL AND METHODS

A. Expression data

We were provided the expression data which were measured in previous investigation [6], and the details of data are follows. The nine genes considered to be affected by chemicals were measured in the human embryonic stem cells: GATA2, Lmx1A, MAP2, Nanog, Nestin, Nodal, Oct3/4, Pax6 and Tuj1 [6][18]. As an internal control, the expression of beta-actin was also measured. The expression data were obtained from human embryonic stem cells exposed to 15 chemicals [6][18]. The toxicity of each chemical was classified into one of three types: Neurotoxic, Carcinogenic and others. The human embryonic cells were exposed to each chemical for several time periods: 24 hours,

  ISBN

48 hours, 72 hours and 96 hours. Each chemical was also tested at 5 concentrations: very low, low, middle, high and very high. The expression of the selected genes was measured twice under each condition by RT-PCR, and thus 300 expression patterns per gene were measured [18].

The measured expression level of each gene was normalized as follows:

$$E_g = \frac{1}{N}\sum_{i=1}^{N}\log_2\left(\frac{e_g^i}{e_{bActin}^i}\right) \quad (1)$$

Here, N is the number of repeated experiments, e_g^i is the measured expression level of gene g under one set of conditions, and e_{bActin}^i is the beta-actin expression level measured under the same conditions. The expression level of each gene was divided by that of beta-actin, for intracellular normalization. To minimize the experimental error, the logarithms of the normalized expression data were obtained and averaged.

B. Extraction of causalities from expression data

For the iteration of model fitting in SEM, an initial model should be assumed from known information. To construct the initial model among the 9 neurodevelopment genes from the time series expressions, we applied cross correlation to the expression profiles measured for each chemical and each concentration.

Cross correlation is utilized as a measure of similarity between two waves in signal processing by a time-lag application, and it is also applicable to pattern recognition [19]. The cross correlation values range between -1 and +1. In a time series analysis, the cross correlation between two time series describes the normalized cross covariance function. Let $X_t = \{x_1, \cdots, x_N\}$, $Y_t = \{y_1, \cdots, y_N\}$ represent two time series data including N time points, and then the cross correlation is given by

$$r_{xy} = \frac{\sum_{t=1}^{N}\{x_t - \bar{x}\}\{y_{t+d} - \bar{y}\}}{\sqrt{\sum_{t=1}^{N}\{x_t - \bar{x}\}^2}\sqrt{\sum_{t=1}^{N}\{y_{t+d} - \bar{y}\}^2}} \quad (2)$$

where d is the time-lag between variables X and Y. In this case, the expression profiles were measured at 4 time points, and thus three cross correlations of each gene pair were calculated with $d = -1, 0, 1$.

C. Construction of the initial model

In this study, we focused on the chemical-specific regulatory network, and thus the differences between times and concentrations could be merged for the construction of the initial model. Figure 1 shows the new method developed for constructing an initial model of each chemical, with the merging of several conditions. The time difference was summarized by the cross correlations among genes. The time lag, which was defined for the calculation of the cross correlation, was used for the extraction of causality between all gene pairs. According to the time lags, three cross correlations were calculated between each gene pair, and we compared them with the absolute values of the cross correlations. The value *d*, with the highest cross correlation, was selected as the causal information between the gene pairs, and a matrix composed of the selected *d*s was constructed as the time lag matrix of each chemical at one concentration. Thus, five time lag matrices were constructed for each chemical (Fig. 1a).

To obtain the chemical-specific interactions among genes, we extracted the binomial relationships between gene pairs from the five constructed time lag matrices for each chemical (Fig. 1b). From the binomial relationships, we constructed a frequency matrix for each chemical, composed of the frequencies of all gene pairs (Fig. 1c). In this step, the difference in the concentration is merged as the frequency in the matrix. We extracted the gene pairs with frequency matrix values greater than or equal to two, as the chemical-specific regulation (Fig. 1d). From the extracted relationships between the genes, we reconstructed an initial model for each chemical (Fig. 1e). These initial models included the time series information as the directions of edges, and the different concentrations of each chemical were summarized as the existence of edges in the model.

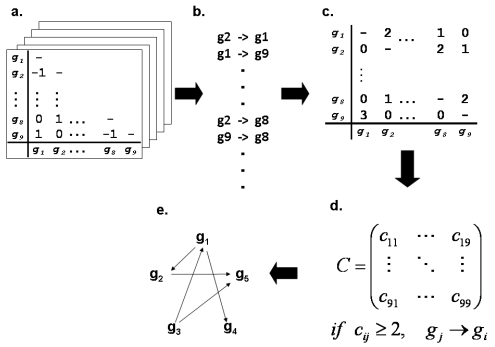


Figure 1. Developed procedure for initial model construction.

The procedure for constructing an initial model from the time-lag information of the cross correlation coefficients. (a) Time-lag matrices for each chemical. In this study, three time-lags were selected for the calculation of the cross correlation coefficients. Thus, three cross correlation coefficient values were obtained between all gene pairs. The time-lag value with the highest absolute value among the cross correlation coefficients was selected. Time-lag matrices were constructed for each concentration, so five time-lag matrices were obtained for each chemical. (b) Binomial relationships. These relationships were extracted from the five time-lag matrices. If the same relationships exist in several concentration matrices, then the extracted binomial relationships are duplicated in this step. (c) Frequency matrix of causal relationships between all gene pairs. From the binomial relationship, we can count the frequency of relationships between gene pairs. (d) Selection of possible causal relationships from the frequency matrix. The possible relationships between genes are considered to persist at several chemical concentrations. Thus, we selected the relationships with two or more values in the frequency matrix. (e) Construction of an initial model with selected causal relationships. By this approach, an initial model can include cyclic structures.

  ISBN

D. Structural Equation Modeling without Latent Variables (SEM without LV)

In general, SEM is a comprehensive statistical model that includes two types of variables: observed and latent. These variables constitute the structural models that consider the relationships between the latent variables and the measurement models that consider the relationships between the observed variables and the latent variables. These relationships can be presented both algebraically, as a system of equations, and graphically, as path diagrams.

In this study, the selected genes (GATA2, Lmx1A, MAP2, Nanog, Nestin, Nodal, Oct3/4, Pax6 and Tuj1), which are related to neurogenesis, were defined as the observed variables. Meanwhile, none were defined as latent variables. All observed variables were categorized into one of two types of variables, exogenous and endogenous, according to their interactions with other variables. Exogenous variables are those that are not regulated by the other variables, and endogenous variables are regulated by the others. In the initial model, the starting genes are defined as exogenous variables, while all other genes are defined as endogenous variables. Regulatory relationships exist between the observed variables in the network models. The model is defined as follows:

$$y = \Lambda y + \varepsilon \tag{3}$$

Here, y is a vector of p observed variables (measured gene expression patterns), and Λ is a $p \times p$ matrix representing the regulatory relationships between the observed variables. Errors that affect the observed endogenous variables are denoted by ε. The SEM software package SPSS AMOS 17.0 (IBM, USA) was used to fit the model to the data.

E. Parameter Estimation

Parameter estimation was performed by comparing the actual covariance matrix, calculated from the measured data, and the estimated covariance matrices of the constructed model. Maximum likelihood is commonly used as a fitting function to estimate SEM parameters:

$$F_{ML}(S, \Sigma(\theta)) = \log|\Sigma(\theta)| - \log|S| + tr(\Sigma(\theta)^{-1} S) - p \tag{4}$$

Here, $\Sigma(\theta)$ is the estimated covariance matrix, S is the sample covariance matrix, $|\Sigma|$ is the determinant of matrix Σ, $tr(\Sigma)$ is the trace of matrix Σ, and p is the number of observed variables. The principal objective of SEM is to minimize $F_{ML}(S, \Sigma(\theta))$, which is the objective function and is used to obtain the maximum likelihood. Generally, $F_{ML}(S, \Sigma(\theta))$ is a nonlinear function. Therefore, iterative optimization is required to minimize $F_{ML}(S, \Sigma(\theta))$ and to find the solutions [20].

F. Iteration for Optimal Model

The regulatory network analysis by SEM consists of two parts: parameter fitting and structure fitting. After the parameters of the constructed model are estimated by maximum likelihood, the network structures are evaluated according to the goodness of fit between the constructed model and the measured data. Through acceptance or rejection of the models, the optimal model that describes measured data can be selected.

In the network model, the covariance matrix between variables is calculated by the estimated parameters. The similarity between a constructed model and the actual relationships is predicted by comparing the matrix calculated from the network model to the matrix calculated from the actual data. To detect quantitative similarity between a constructed model and an actual relationship, fitting scores were developed. In this study, the quality of the fit was predicted by four different fitting scores: GFI, AGFI, CFI and RMSEA. Values of GFI, AGFI and CFI above 0.90 are required for a good model fit. RMSEA is one of the most popular parsimony indexes displayed in the table, and RMSEA values below 0.05 represent a good model fit [21]. Furthermore, RMSEA values of 0.10 or more are considered to indicate that the constructed model is far from the actual data.

To optimize the model, an iteration algorithm was developed, as follows:

Step1: Deletion of a non-significant edge from the model. Use 0.05 as the significance level for the determination of the chemical-specific interactions among genes. The output of SEM programs includes the probability of each edge, and thus we deleted the edge with the highest probability.

Step2: Reconstruction of the network model. The structure of the network model without the non-significant edge is different from the former network model. Thus, all parameters should be re-calculated from the reconstructed model, and the similarity of the network structure is also re-calculated.

Step3: Iteration of Steps 1 and 2 until all edges become significant. Since the probabilities of all of the edges in the reconstructed models have also changed, the deletion of the non-significant edges is executed step-by-step.

Step4: Addition of a possible causal edge to the reconstructed model. According to the Modification Index (MI), we add a new causal edge between the observed variables. The MI value indicates the possibility of new causality between the variables, and thus we add a new edge according to the highest MI score.

Step5: Iteration from Steps 1 to 3. By the addition of a new edge to a constructed model, the structure of network model is changed again. In other words, all parameters, including the probabilities of all edges, have also changed again. Thus, we execute the iteration from Step 1 to Step 3 again.

Step6: Determination of significant relationships among error terms. After all of the edges are significant and all of the MI scores are lower than 10.0 in the constructed model, significant relationships between error terms are estimated

  ISBN

by the MI scores. The relationships among the error terms have no direction, and thus they are a correlation between error terms. These relationships were used for the calculations, but were not incorporated into the network.

III. RESULTS AND DISCUSSION

A. The chemical concentrations have no effect

In this study, gene expression was measured in the presence of various chemicals, with several exposure times and at different concentrations. To reveal the most effective factor for gene expression, ANOVA and Tukey's HSD test were applied to the measured data. In statistics, ANOVA is utilized to detect differences between groups in terms of some variables. Since the chance of committing a type I error will be increased by performing multiple two-sample t-tests, a statistical test is needed to determine whether or not the means of more than two groups should be applied. The use of Tukey's HSD test clarified which means are significantly different from one another. Interestingly, the groups that were classified by the concentration of chemicals showed no significant difference in their gene expressions. Thus, the concentration of chemicals had no effect on the expression of the tested genes in the ES cells.

The numbers of significantly expressed genes between each chemical pair are shown in Table 1. From this table, the differences in the gene expression were not significant among the same type of toxic chemicals. Furthermore, the toxicity difference between neurotoxic and carcinogenic did not cause an expression difference for almost all of the genes. However, the exposure to 'other' chemicals, such as Thalidomide, bisphenol A and Permethrin, caused significant expression changes in many genes. To reveal the differences in gene expression due to the type of chemical toxicity, we selected one chemical, which was the most different from those of the other toxicity types, as the representative chemical for each toxicity type.

TABLE I. NUMBER OF GENES WITH SINIFICANTLY DIFFERENT EXPRESSION

		Methylmercury	2-Nitropropane	Acrylamide	p-Nitroaniline	4-hydoroxy PCB107	Benzo[a] pyrene	Diethylnitrosamine	Diethylaminofluorene	Phenobarbital	Tamoxifen	Diethylbestrol	TCDD	Thalidomide	Bisphenol-A	Permethrin
	Methylmercury	-														
	2-Nitropropane	1	-													
Neurotoxicity	Acrylamide	0	4	-												
	p-Nitroaniline	1	2	1	-											
	4-hydoroxy PCB107	2	4	0	0	-										
	Benzo[a] pyrene	4	5	1	2	0	-									
	Diethylnitrosamine	4	3	5	4	5	5	-								
	Diethylaminofluorene	3	3	4	5	5	7	0	-							
Carcinogenic	Phenobarbital	3	2	4	0	1	2	6	6	-						
	Tamoxifen	3	4	3	1	0	0	6	6	0	-					
	Diethylbestrol	3	5	2	0	1	0	5	6	0	0	-				
	TCDD	3	2	1	0	0	1	5	6	0	0	0				
	Thalidomide	5	7	5	5	5	4	7	7	4	4	4	5	-		
others	Bisphenol-A	5	6	5	5	3	4	6	7	6	3	3	3	0	-	
	Permethrin	4	6	4	5	4	4	7	7	3	4	4	4	0	0	-
		Methylmercury	2-Nitropropane	Acrylamide	p-Nitroaniline	4-hydoroxy PCB107	Benzo[a] pyrene	Diethylnitrosamine	Diethylaminofluorene	Phenobarbital	Tamoxifen	Diethylbestrol	TCDD	Thalidomide	Bisphenol-A	Permethrin
		Neurotoxicity					Carcinogenic							others		

Expression profiles were compared between all chemical pairs for each gene by Turkey's HSD test, and the genes that were estimated as exhibiting significantly different expression dependent on chemicals were counted.

The toxicities of the 15 chemicals were divided into 3 categories, Neural, Carcinogenic and others. Thus, three chemicals, Acrylamide, Diethylnitrosamine and Thalidomide, were selected as the representative toxic chemicals for neural, carcinogenic and others, respectively.

B. Genes are hierarchically controlled by chemical toxicity

We utilized the new method to construct the initial gene regulatory network models under the conditions with the three chemicals. Even though the new method can detect the cyclic interactions among genes, such as feed-back regulation, the structures of the constructed initial models indicated the hierarchical regulations among the genes. Figure 2 shows the constructed initial network models. The initial models of Acrylamide, Diethylnitrosamine and Thalidomide were composed of 9 genes with 19 relationships, 8 genes with 14 relationships, and 8 genes with 10 relationships, respectively.

There are some obvious features in the hierarchical diagram of each initial model. The numbers of exogenous and endogenous genes are different from each other. The initial Acrylamide model was composed of 4 genes as exogenous genes, but only Oct3/4 was the last endogenous gene. Thus, the expression profiles of Acrylamide indicated the quick responses of many genes after the chemical exposure, and only one gene was affected later. In contrast, Thalidomide exposure induced the expression of only one gene. These differences between the initial chemical models summarized the distinctive gene expression profiles for each chemical.

All of the initial models included some duplicated gene interactions, such as a direct interaction between two genes and an indirect interaction between them. To simplify these duplicated interactions, we only retained the longest path between two genes, since the regulation displayed by a direct path could be replaced by indirect paths in the model. In the initial model, the edges do not represent the direct regulation, but the preceding information. Thus, the difference between direct interaction and indirect interaction in the initial model is not very important. By retaining the longest paths, all of the preceding information was included, as the simplest diagram.

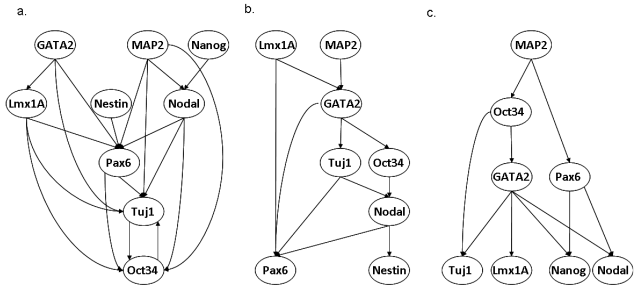


Figure 2. Initial network models.

The initial models which include summarized time-series information and concentration information. (a) Initial model of Acrylamide. (b) Initial model of Diethylnitrosamine. (c) Initial model of Thalidomide.

  ISBN

C. *Inferred Network by SEM*

The final inferred networks for each chemical and the estimated regression weights of the edges are depicted in Figure 3. The inferred networks of chemicals revealed distinct structures. In the inferred network of Acrylamide, many genes were arranged as exogenous objects, and only one gene was arranged as the final result of all regulation in the network. On the other hand, two serial regulations interacted with each other in the Diethylnitrosamine network model. One serial regulation was from Lmx1A to Pax6, and the other was from Tuj1 to Nestin. The signal input genes in the Diethylnitrosamine network were also different from those in the Acrylamide network. Even though Tuj1 was arranged as an output object in the Acrylamide network, Tuj1 was arranged as input in the Diethylnitrosamine network. The inferred network of Thalidomide was also different from both the Acrylamide and Diethylnitrosamine networks. In the Thalidomide network, only two genes were arranged as input objects, but four genes were arranged as output objects. This means that only a few genes will be directly affected by Thalidomide, but finally many genes were affected throughout the gene regulatory network.

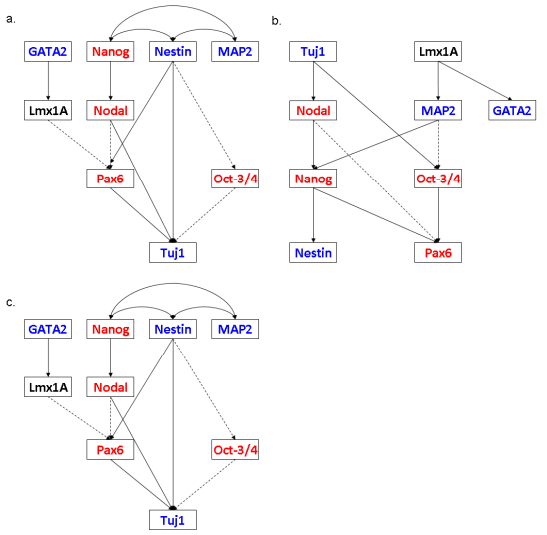
According to our inferred network, the differences between the gene regulation by chemicals were clarified, and the network shapes reflected the features of chemical toxicities. In the inferred network, the effects of Acrylamide toxicity finally aggregated to Tuj1, which is known to contribute to microtubule stability in neuronal cells [22]. Acrylamide is neurotoxic, and thus it is reasonable that the effect of Acrylamide finally aggregated to a neuronal cell-related gene.

As compared with the Acrylamide network, the cell differentiation genes were arranged at downstream steps in the Diethylnitrosamine network. From the carcinogenic features of Diethylnitrosamine [23][24][25], normal cell differentiation in the embryonic stem cell may be disturbed.

The most complicated structure was the Thalidomide network. In the Thalidomide network, several type of genes are finally affected by its chemical toxicity. Particularly, two cell differentiation-related genes, Nodal and Nanog, are important for normal early embryonic development. Nodal is related to the development of the left-right axial structure [26][27], and its signaling pathway is known to be important very early in development for cell fate determination and many other developmental processes [27]. Nanog is known as a key factor for maintaining pluripotency in embryonic stem cells [28][29]. Thus, the unusual expressions of these genes, which occurred due to Thalidomide toxicity, may have caused its harmful side effects.

IV. CONCLUSION

We applied an improved SEM approach to reconstruct a gene regulatory model from gene expression data in human embryonic stem cells, and we have shown that SEM is a powerful approach to estimate the gene regulation caused by chemical toxicity. The inferred networks clarified the differences between the gene regulation by chemicals, and the features of chemical toxicities were well reflected in the network structures. Thus, the network construction by SEM is one of the useful approaches for inferring the regulatory relationships among genes. Furthermore, the inferred



d.

Acrylamide			Diethylnitrosamine			Thalidomide		
parent	child	regression weight	parent	child	regression weight	parent	child	regression weight
GATA2	Lmx1A	0.919	Tuj1	Nodal	0.702	MAP2	Oct-34	-0.448
Nanog	Nodal	0.515	Lmx1A	MAP2	0.537	MAP2	Pax6	0.43
Nestin	Oct-34	-0.442	Tuj1	Oct-34	0.623	Oct-34	Pax6	0.951
Lmx1A	Pax6	-0.129	MAP2	Oct-34	-0.492	Nestin	Nodal	1.026
Nestin	Pax6	0.606	MAP2	Nanog	0.9	Nestin	GATA2	0.664
Nodal	Pax6	-0.844	Nodal	Nanog	0.284	Oct-34	Nodal	0.25
Oct-34	Pax6	0.643	Lmx1A	GATA2	0.636	Oct-34	Tuj1	-0.477
Pax6	Tuj1	1.145	Nodal	Pax6	-0.209	GATA2	Tuj1	0.183
Oct-34	Tuj1	-0.77	Nodal	Nestin	0.269	MAP2	Nanog	0.948
Nodal	Tuj1	0.933	Nanog	Nestin	0.843	Nestin	Lmx1A	0.831
MAP2	Nanog	0.968	Oct-34	Pax6	1.121	Pax6	Tuj1	1.002
MAP2	Nestin	0.944	Nanog	Pax6	0.428	Nodal	Tuj1	0.308
Nanog	Nestin	0.966				Nestin	MAP2	0.922

Figure 3. Inferred network by SEM.

The optimal model for each chemical, obtained by the developed SEM iteration procedure. A positive relationship between genes is displayed with a solid arrow. A negative relationship between genes is displayed with a dashed arrow. Gene names with blue characters indicate "neurodevelopment related genes", genes with red characters indicate "cell differentiation-related genes" and genes with black characters indicate "related to transcription of insulin". (a) Acrylamide model, (b) Diethylnitrosamine model and (c) Thalidomide model. (d) The estimated regression weights of all edges in the optimal models.

  ISBN

network among genes can be utilized for the estimation of a chemical's effect, from experimentally obtained expression profiles. The ability to identify expression profiles and the corresponding biological functions is expected to provide further possibilities for SEM in the inference of regulatory mechanisms by chemical toxicity.

ACKNOWLEDGMENT

We would like to express our gratitude to Dr. Yamane (Kyoto University) and Dr. Ohsako (University of Tokyo) for providing us the expression profiles in human embryonic stem cells exposed to 15 chemicals. Their supports were invaluable for this investigation.

REFERENCES

[1] A. Baccarelli and V. Bollati, "Epigenetics and environmental chemicals," Curr. Opin. Pediatr. vol.21(2), pp. 243-251, Apr. 2009.

[2] L. Hou, X. Zhang, D. Wang, and A. Baccarelli, "Environmental chemical exposures and human epigenetics," Int. J. Epidemiol., vol. 41(1), pp. 79-105, Feb. 2012. doi:10.1093/ije/dyr154.

[3] Y. Yuan, "Methylmercury: a potential environmental risk factor contributing to epileptogenesis," Neurotoxicology, vol. 33(1), pp. 119-126, Jan. 2012.

[4] N. Tatsuta, et al., "Prenatal exposures to environmental chemicals and birth order as risk factors for child behavior problems," Environ. Res., vol.114, pp. 47-52, Apr. 2012.

[5] D. A. Rappolee, Y. Xie, J.A. Slater, S. Zhou, and E. E. Puscheck, "Toxic stress prioritizes and imbalances stem cell differentiation: implications for new biomarkers and in vitro toxicology tests," Syst. Biol. Reprod. Med., vol.58(1), pp. 33-40, Feb. 2012.

[6] X. He, et al., "Effects of methylmercury exposure on neuronal differentiation of mouse and human embryonic stem cells,". Toxicol. Lett., vol.212(1), pp. 1-10, Apr. 2012.

[7] J. A. Harrill, B. L. Robinette, and W. R. Mundy, "Use of high content image analysis to detect chemical-induced changes in synaptogenesis in vitro," Toxicol. In. Vitro., vol.25(1), pp. 368-387, Feb. 2011.

[8] U. Gündel, D. Benndorf, M. von Bergen, R. Altenburger, and E. Küster, "Vitellogenin cleavage products as indicators for toxic stress in zebra fish embryos: A proteomic approach," Proteomics, vol.7(24), pp. 4541-4554, Dec. 2007.

[9] J. Thompson and J.Bannigan, "Cadmium: toxic effects on the reproductive system and the embryo," Reprod. Toxicol., vol.25(3), pp. 304-315, Apr. 2008.

[10] T. Akutsu, S. Miyano, and S. Kuhara, "Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function," J. Comput. Biol., vol.7, pp. 331-343, 2000.

[11] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian networks to analyze expression data," J. Comput. Biol., vol.7, pp. 601-620, 2000.

[12] S. Aburatani, S. Kuhara, H. Toh, and K. Horimoto, "Deduction of a gene regulatory relationship framework from gene expression data by the application of graphical Gaussian modeling," Signal Processing, vol.83, pp. 777-788, Apr. 2003.

[13] S. Aburatani and K. Horimoto, "Elucidation of the Relationships between LexA-Regulated Genes in the SOS response," Genome Informatics, vol.16(1), pp. 95-105, Aug. 2005.

[14] S. Aburatani, "Application of structure equation modeling for inferring a serial transcriptional regulation in yeast," Gene. Regul. Syst. Bio., vol.5, pp. 75-88, Nov. 2011.

[15] K. A. Bollen, Structural Equations with Latent Variables, New York: Wiley-Interscience, 1989.

[16] O. D. Duncan, Introduction to Structural Equation Models, New York: Academic Press, 1975.

[17] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed., Cambridge: Cambridge University Press, 2001.

[18] W. Fujibuchi, et al., "Prediction of Chemical Toxicity by Network-based SVM on ES-cell Validation System," The Proc. of the 2011 Joint Conference of CBI-Society and JSBi, Nov. 2011.

[19] L. Li and G. E. Caldwell, "Coefficient of cross correlation and the time domain correspondence," J. of Electromyography and Kinesiology, vol.9, pp. 385–389, Dec. 1999.

[20] K. G. Joreskog and D. Sorbom, LISREL-VI: Analysis of Linear Structural Relationships By the Method of Maximum Likelihood, Redondo Beach: Doss-Haus Books, 1984.

[21] P. Spirtes, C. Glymour, and R. Scheines, Causation, Prediction, and Search, 2nd ed., Cambridge: The MIT Press, 2001.

[22] J. M. Rosenstein, N. Mani, A. Khaibullina, and J. M. Krum, "Neurotrophic effects of vascular endothelial growth factor on organotypic cortical explants and primary cortical neurons,". J. Neurosci., vol.23(35), pp. 11036-11044, Dec. 2003.

[23] N. Ito, et al., "Pathological markers for non-genotoxic agent-associated carcinogenesis,". Toxicol. Lett., Vol. 64-65, pp. 613-620, Dec. 1992.

[24] R. Puatanachokcha, et al., "Lack of promoting effects of phenobarbital at low dose on diethylnitrosamine-induced hepatocarcinogenesis in TGF-alpha transgenic mice," Asian Pac. J. Cancer Prev., vol.7(2), pp. 274-278, Apr-Jun 2006.

[25] M. J. Iatropoulos, C. X. Wang, K. E. von Keutz, and G. M. Williams, "Assessment of chronic toxicity and carcinogenicity in an accelerated cancer bioassay in rats of Nifurtimox, an antitrypanosomiasis drug," Exp. Toxicol. Pathol., vol.57(5-6), pp. 397-404, Jul. 2006.

[26] H. Hamada, C. Meno, D. Watanabe, and Y. Saijoh, "Establishment of vertebrate left-right asymmetry," Nat. Rev. Genet., vol.3(2), pp. 103-113, Feb. 2002.

[27] C. Grande1 and N. H. Patel, "Nodal signaling is involved in left-right asymmetry in snails," Nature, vol.457(7232), pp. 1007–1011, Feb. 2009.

[28] K. Mitsui, "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells," Cell, vol.113(5), pp. 631-642, May 2003.

[29] I. Chambers, et al., "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells," Cell, vol.113(5), pp. 643-655, May 2003.

  ISBN