RESEARCH PAPER

Systems Biomedicine 1:3, 1–7; July/August/September 2013; © 2013 Landes Bioscience

The impact of collapsing data on microarray analysis and DILI prediction

Jean-François Pessiot,[1,*] Pui Shan Wong,[1] Toru Maruyama,[2] Ryoko Morioka,[1] Sachiyo Aburatani,[1] Michihiro Tanaka[3] and Wataru Fujibuchi[1,2,3,*]

[1]Computational Biology Research Center; Advanced Industrial Science and Technology; Tokyo, Japan; [2]Faculty of Science and Engineering; Waseda University; Tokyo, Japan; [3]Center of iPS Cell Research and Application (CiRA); Kyoto University; Kyoto, Japan

Keywords: in vivo-in vitro comparison, gene set enrichment analysis, data collapsing, drug-induced liver injury, toxicity prediction

Abbreviations: AUC, area under the curve; DILI, drug-induced liver injury; IVV, in vivo; IVT, in vitro; GSEA, gene set enrichment analysis; ROC, receiver operating characteristic; RBF, radial basis function; SVM, support vector machine

In this work, we focus on two fundamental problems of toxicogenomics using the data provided by the Japanese toxicogenomics project. First, we analyze to what extent animal studies can be replaced by in in vitro assays. We show that the probeset-level representation achieves poor agreement between in vivo and in vitro data. We present a data collapsing approach to resolve poor data agreement between in vivo and in vitro data, as measured by GSEA analysis and AUC scores. Second, we address the difficult problem of predicting DILI using available microarray data. Using a binary classification framework, our results suggest that rat in vivo data are more informative than human in vitro data to predict DILI.

Introduction

In the field of toxicology, animal studies and in vitro experiments are frequently used as surrogates for human studies even though they have shown poor agreement so far. It is still unclear how the results obtained from one animal species, such as rats, can help important biomedical research areas for humans, such as the prediction of drug-induced liver injury (DILI). This work is an attempt to address both issues, using the toxicogenomics data set provided by the Japanese toxicogenomics project.[1]

First, we analyzed to what extent animal studies can be replaced by in in vitro assays. We compared lists of differentially expressed probesets between rat in vivo and rat in vitro data, and found poor agreement between the two. This confirmed previous studies suggesting that probeset-level analysis has major limitations, and motivated us to consider higher levels of data abstractions. Thus, we present a data collapsing approach which improves the agreement between in vivo and in vitro data. We collapsed probesets and evaluated the in vivo–in vitro agreement using Gene Set Enrichment Analysis (GSEA). We also collapsed time points and evaluated the in vivo–in vitro agreement using the binary classification framework.

Second, we addressed the problem of predicting DILI using available microarray data. DILI is the primary cause for the failure of drug candidates during clinical trials, and for the withdrawal of drugs from the market. Because of the potentially dangerous adverse effects of DILI, and because of its cost for the healthcare system and the pharmaceutical industry, the study of DILI has become an important research area in drug development.[2-5] In particular, the ability to accurately predict the DILI level of a candidate drug early in its development would be highly desirable. To this end, it is necessary to pre-process and analyze the results of toxicity assessments for various drugs and various conditions.

Intuitively, we would expect that unprocessed in vitro data are too noisy for DILI prediction and would need to be collapsed in order to achieve a better signal-to-noise ratio. In contrast, we would also expect unprocessed in vivo data to contain information that is important for predicting DILI. This information could be lost during data collapsing, resulting in a lower prediction performance. Our prediction results tend to confirm these assumptions, and suggest to use unprocessed in vivo data simultaneously with collapsed in vitro data to improve the prediction performance of DILI.

Results

Comparison between in vivo and in vitro data. *Probeset-level differential expression analysis.* We studied the similarity between in vivo and in vitro data at the probeset-level using differential expression analysis. Using empirical Bayes statistics, we tested for agreement between in vivo and in vitro probesets at corresponding time points. In our experiments, we obtained AUC = 0.56 for t_{IVV} = 3 h and t_{IVT} = 2 h, AUC = 0.56 at t_{IVV} = 9 h and t_{IVT} = 8 h and AUC = 0.60 at t_{IVV} = 24 h and t_{IVT} = 24 h (**Fig. 1**).

*Correspondence to: Jean-François Pessiot and Wataru Fujibuchi; Email: jfk.pessiot@aist.go.jp and w.fujibuchi@cira.kyoto-u.ac.jp
Submitted: 10/23/12; Revised: 12/10/12; Accepted: 12/31/12
http://dx.doi.org/10.4161/sysbiomad.24255

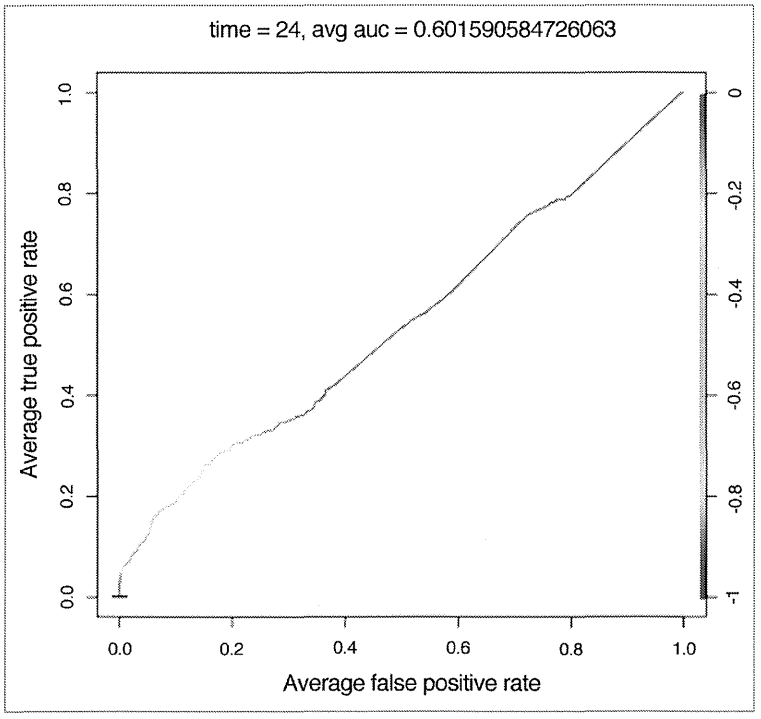

Figure 1. Comparison between in vivo and in vitro differentially expressed probesets at t_{IVV} = 24 h and t_{IVT} = 24 h.

Gene set enrichment analysis procedure. Data overlap between IVT and IVV. In our initial GSEA analyses, we assessed whether expression patterns between the two data sets were significantly similar by comparing in vivo and in vitro data at two time points for each of the 131 drugs in both directions. For each drug, there were four possible outcomes: all four analyses were significantly enriched ($p < 0.05$), meaning there was full agreement with expression patterns between in vivo and in vitro data; three analyses were significantly enriched, meaning there was full agreement with expression patterns at one time point but not the other; two analyses were significantly enriched, meaning either there was full agreement at only one time point or there was only agreement in one direction at each time point; and one analysis was significantly enriched, meaning the expression pattern did not agree except at one time point in a particular direction. Of the 131 drugs analyzed, 55% were significantly enriched in all four tests, 34% were enriched in three tests, 10% were enriched in two tests and 2% were enriched for 1 test.

To check that the lists were not simply correlated, Spearman's rank correlation was calculated between each pair of in vivo and in vitro lists. The correlation coefficients were all under 0.25, indicating that there is little correlation between the lists of fold changes between in vivo and in vitro data.

Data overlap between IVT and liver function, and IVV and liver function. For the GSEA analyses against liver gene ontologies, in vivo and in vitro data were compared with predefined lists of liver gene ontologies at two time points for each of the 131 drugs. To assess whether enrichment analyses against gene ontologies were similar between in vivo and in vitro data, results were compared between in vivo data at 2 h and in vitro data at 3 h, and in vivo and in vitro data at 24 h. For each drug, there were two possible outcomes: enrichment significance was the same between in vivo and in vitro at the two time points; enrichment significance was the same at only one of the two time points; and enrichment significant was different across the two time points. Of the 131 drugs analyzed, 50% had the same enrichment results between in vivo and in vitro at the two time points, 42% had the same enrichment results at one of the time points and 8% did not share the same enrichment results at the two time points. This suggests that liver related genes are highly overexpressed or under expressed in response to 50% of the drugs in both in vivo and in vitro conditions.

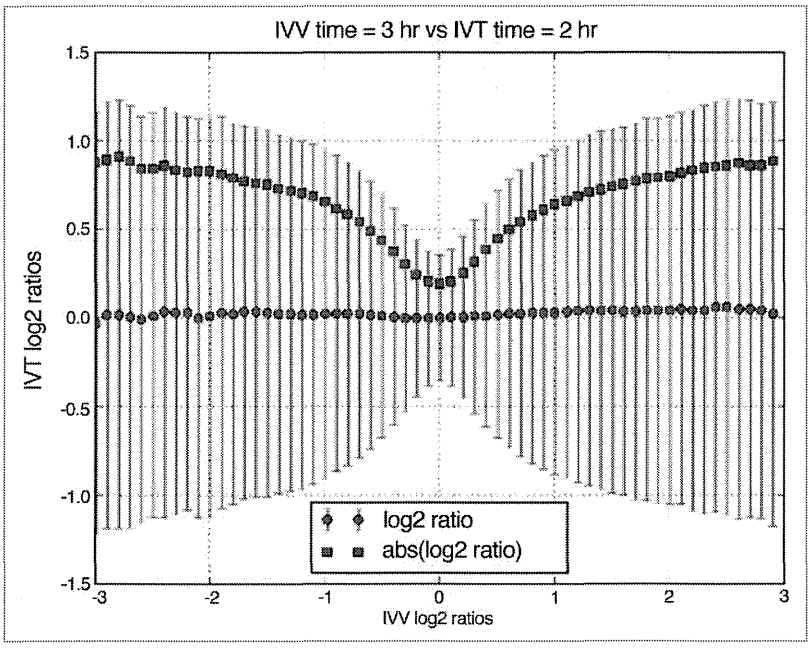

Figure 2. In vivo fold changes at t = 3 h vs. in vitro fold changes at t = 2 h.

Comparison between in vivo fold changes and in vitro fold changes. We analyzed the relationships between in vivo fold changes and in vitro fold changes. We plotted the in vitro fold change (averaged over all drugs and all doses) as a function of the in vivo fold change, at t_{IVV} = 3 h and t_{IVT} = 2 h (**Fig. 2**), and at t_{IVV} = 24 h and t_{IVT} = 24 h (**Fig. 3**).

Collapsing strategies for probesets and time points. In order to improve the signal-to-noise

ratio of the data, we considered a data collapsing approach. **Table 1** shows the various collapsing strategies and their corresponding AUC scores. The highest score AUC = 0.85 ± 0.04 was achieved when using gene-level representation, absolute fold changes, and average values over time points.

In **Figure 4**, we show the ROC curves for three data collapsing strategies. All three approaches use absolute values. The blue curve corresponds to the probeset level and time series integral strategy (AUC = 0.79). The green curve corresponds to the gene level and time series integral strategy (AUC = 0.84). The red curve corresponds to the optimal strategy: gene level, time series average (AUC = 0.85).

Correlation matrix analysis. We used the correlation matrix distances to assess the agreement between in vivo and in vitro gene expression data. **Table 2** shows the top ten genes which show the most similar behaviors between in vivo and in vitro conditions.

Predicting drug-induced liver injury in humans. We considered the DILI prediction problem as a binary classification of "Most DILI" against "Less DILI or no DILI." We used two support vector machine classifiers and evaluated their classification performance using a 10-fold cross validation. **Tables 3 and 4** show the AUC scores of the linear SVM and the RBF kernel SVM, respectively.

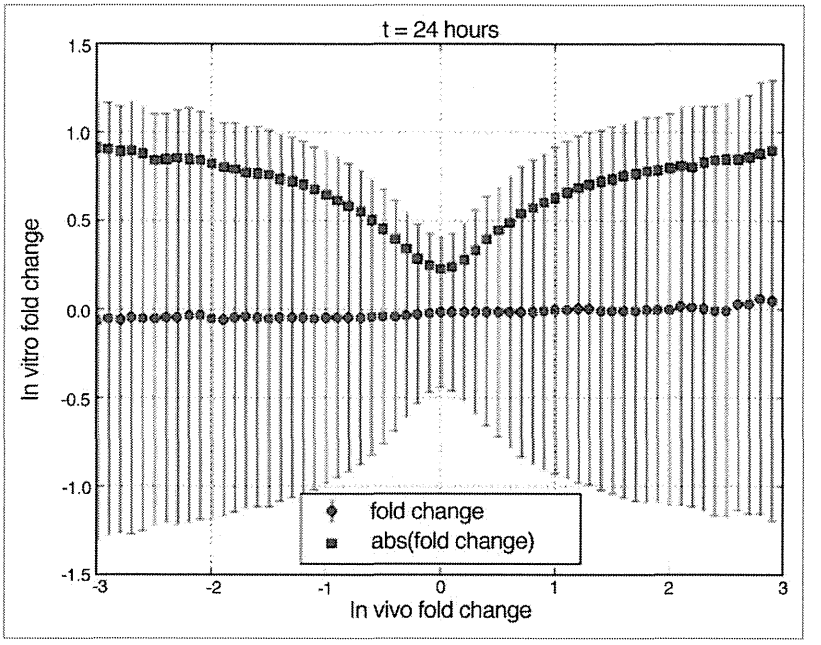

Figure 3. In vivo fold changes at t = 24 h vs. in vitro fold changes at t = 24 h.

Table 1. Data collapsing strategies and their AUC scores

Collapse probesets	Collapse time points	Collapsing strategy	Absolute values	Average AUC
False	False	First	False	0.51 ± 0.04
False	False	Last	False	0.51 ± 0.04
False	True	Integral	False	0.52 ± 0.05
False	True	Average	False	0.52 ± 0.04
True	False	First	False	0.52 ± 0.05
True	False	Last	False	0.51 ± 0.06
True	True	Integral	False	0.52 ± 0.06
True	True	Average	False	0.53 ± 0.06
False	False	First	True	0.72 ± 0.02
False	False	Last	True	0.71 ± 0.04
False	True	Integral	True	0.79 ± 0.03
False	True	Average	True	0.81 ± 0.03
True	False	First	True	0.77 ± 0.03
True	False	Last	True	0.74 ± 0.05
True	True	Integral	True	0.84 ± 0.04
True	True	Average	True	**0.85 ± 0.04**

Discussion

Comparison between in vivo and in vitro data. *Probeset-level differential expression analysis.* The results show that predicting differentially expressed probesets in vivo using in vitro experiments is a difficult task. We have AUC = 0.56 for the first two pairs of time points (t_{IVV} = 3 h and t_{IVT} = 2 h, and t_{IVV} = 9 h and t_{IVT} = 8 h), which is close to random prediction (AUC = 0.5). There is only a slight improvement at t_{IVV} = 24 h and t_{IVT} = 24 h, with AUC = 0.60.

Overall, the empirical Bayes statistics shows that the agreement between in vivo and in vitro is poor, at the probeset-level, with respect to differential expression analysis. This can be explained by the discrepancy between its expressions in the in vivo and in vitro conditions for each biological process. This can also be caused by the limitations of a probeset-level analysis. Therefore, we need to consider higher levels of data abstractions, and compare the in vivo and in vitro data at the gene level.

Gene set enrichment analysis procedure. Although the original GSEA concept used signed fold change values, we decided to perform GSEA on absolute fold changes due to the observations in **Figures 2 and 3**. Our GSEA results show that most highly expressed or under expressed probe expression in vitro respond to drugs similarly in an equivalent in vivo experiment and vice versa. Even though we didn't find universal agreement among all time points or direction of enrichment, more than half of the drugs caused the same probes to show extremes of expression values. This may be due to the different number of samples between the in vivo and in vitro data sets, leading to some tests being enriched in one direction but not the other. If the analyses only involved interest in highly over or under expressed genes, in vitro data can be an adequate model for in vivo data for rats. We also showed that in gene ontology enrichment analyses, in vitro data only gives the same results as in vivo half the time but that the confidence of the results may be dependent on the time point.

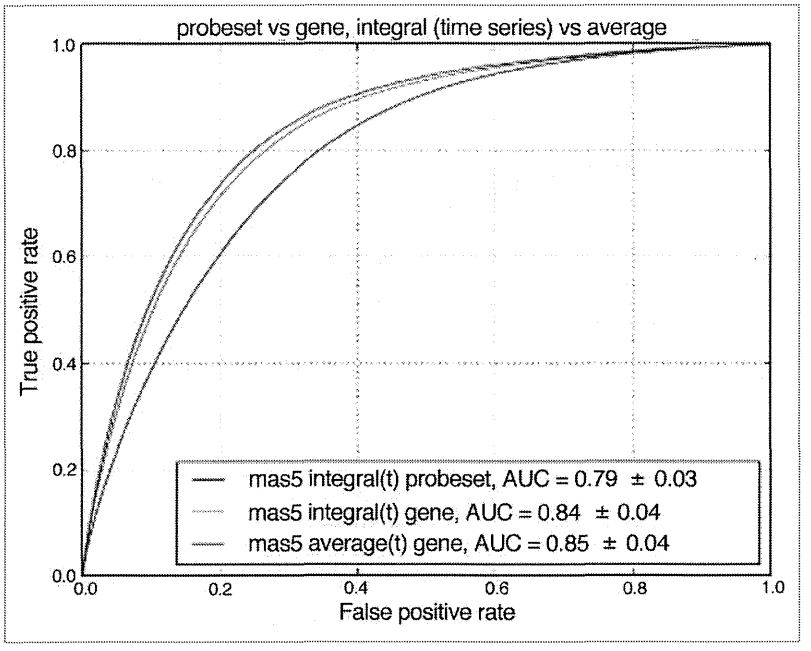

Figure 4. ROC curves for three data collapsing strategies: probeset level + time series integral (blue), gene level + time series integral (green) and gene level + time series average (red).

Table 2. Genes which behave similarly in vivo and in vitro, and their corresponding distance measures

Gene symbol	In vivo–in vitro distance	Gene Symbol	In vivo–in vitro distance
Dazap2	0.4884	Actr2	0.5808
Arf1	0.5446	Gdi2	0.5854
Aamp	0.5573	Cmpk1	0.6030
Ube2l3	0.5608	Morf4l1	0.6121
Cdc42	0.5629	Arpc2	0.6284

Comparison between in vivo fold changes and in vitro fold changes. **Figures 2 and 3** show the in vitro fold change (averaged over all drugs and all doses) as a function of the in vivo fold change. These plots show that while there is no obvious sign correlation between in vivo and in vitro data, there is a correlation between their absolute values. In other words, even if a gene has a highly positive fold change in vivo, we cannot always expect a highly positive fold change in vitro. However, a gene with a high in vivo absolute fold change tends to have a high in vitro absolute fold change as well.

To discern a biological reason behind our observations, we divided genes into two groups: genes that are congruent at signed values and genes that congruent at absolute values. More specifically, for each gene, we computed two correlation coefficients. The first one measured the correlation between the gene's fold changes in vivo and in vitro, while the second one measured the correlation between their absolute values. After filtering significant correlations ($p < 0.05$), we created two groups of genes. The first group contained genes with correlation coefficients of identical signs, while the second group contained genes with correlation coefficients of opposite signs. Then, we compared the gene ontologies for these two groups of genes. However, we did not find any distinguishable trend between the two sets of genes using gene ontology. Our observations may be due to experimental variance or further biological variance which we could not detect.

Collapsing strategies for probesets and time points. In order to improve the signal-to-noise ratio of the data, we also considered various data collapsing strategies. **Table 1** shows the details of these strategies. "Collapse probesets" is set to "True" when probesets values are averaged to their corresponding genes values, "False" otherwise. "Collapse time points" is set to "True" when each time series is collapsed to a single value. In that case, "Collapsing strategy" specifies different time collapsing strategies. With "Integral," we compute the signed integral of the time series. With "Average," we compute the average value of the time series. "Collapse time points" is set to "False" when each time series is reduced to a single time point (i.e., all other time points are discarded). In that case, "Collapsing strategy" specifies which time point of the time series is considered: the first or the last time point. "Absolute values" is set to "True" when we consider absolute fold change values, "False" when we consider the signed values.

For all strategies using signed fold change values, the corresponding AUC scores fall between 0.51 and 0.53. This confirms what we observed in **Figures 2 and 3**: the sign of a gene's fold change in vitro cannot be used to predict its sign in vivo. Thus, our following strategies use the absolute fold changes only. When we collapse probesets into genes (from 0.72 to 0.77) and time series into single values (from 0.72 to 0.79), the AUC scores improve. This is expected, as averaging helps removing the measurement noise and leads to better signal to noise ratios.

As we explained earlier, probeset-level fold changes were collapsed into gene-level fold changes by taking the average value. Another possible strategy consists in considering the maximum value instead, as noise in gene expression measurements tend to occur in probesets with low expression values. However, our experiments show no improvements with respect to AUC scores. For example, we considered the best two strategies in **Table 1**, which achieve AUC scores of 0.84 and 0.85 when considering the average fold change values. When considering the maximum values instead, the same strategies achieve 0.83 and 0.84, respectively. Therefore, when collapsing probesets into genes, we chose to focus on average values.

Surprisingly, we also observe that when collapsing a time series into a single value, computing a simple average value is slighty better than computing the integrals (AUC = 0.81 and 0.85 for the average, AUC = 0.79 and 0.84 for the integral). This result suggests that the time information is also useless, and that we can assume time independence when comparing in vivo and in

Table 3. Average AUC scores of the linear SVM for the DILI prediction problem

Collapse probesets	Collapse time points	Absolute values	Human in vitro	Rat in vitro	Rat in vivo repeated dose	Rat in vivo single dose
False	False	False	0.52 ± 0.17	0.52 ± 0.14	**0.66** ± 0.14	0.61 ± 0.12
False	True	False	0.52 ± 0.13	0.54 ± 0.15	0.56 ± 0.11	0.58 ± 0.17
True	False	False	0.50 ± 0.21	0.47 ± 0.18	0.64 ± 0.13	0.56 ± 0.20
True	True	False	0.49 ± 0.14	0.54 ± 0.13	0.58 ± 0.12	0.50 ± 0.16
False	False	True	**0.59** ± 0.08	**0.58** ± 0.08	0.61 ± 0.17	**0.67** ± 0.15
False	True	True	0.58 ± 0.12	0.55 ± 0.20	0.52 ± 0.19	0.55 ± 0.10
True	False	True	0.56 ± 0.07	0.50 ± 0.16	0.55 ± 0.18	0.62 ± 0.18
True	True	True	**0.59** ± 0.10	0.49 ± 0.12	0.59 ± 0.15	0.63 ± 0.17

Table 4. Average AUC scores of the Gaussian SVM for the DILI prediction problem

Collapse probesets	Collapse time points	Absolute values	Human in vitro	Rat in vitro	Rat in vivo repeated dose	Rat in vivo single dose
False	False	False	0.49 ± 0.19	0.51 ± 0.21	**0.65** ± 0.11	0.61 ± 0.14
False	True	False	0.49 ± 0.10	**0.58** ± 0.12	0.58 ± 0.10	0.63 ± 0.23
True	False	False	0.46 ± 0.14	0.45 ± 0.11	**0.65** ± 0.14	0.58 ± 0.13
True	True	False	0.50 ± 0.23	0.55 ± 0.19	0.61 ± 0.13	0.52 ± 0.17
False	False	True	0.59 ± 0.17	0.50 ± 0.09	0.63 ± 0.15	**0.66** ± 0.11
False	True	True	0.60 ± 0.20	0.55 ± 0.18	0.56 ± 0.21	0.51 ± 0.13
True	False	True	**0.61** ± 0.21	0.47 ± 0.15	0.59 ± 0.16	0.64 ± 0.14
True	True	True	0.60 ± 0.18	0.49 ± 0.14	0.62 ± 0.13	0.59 ± 0.13

vitro data. Overall, the best strategy uses gene-level representation, absolute fold changes, and average values over time points. The corresponding AUC score is 0.85 ± 0.04.

Predicting drug-induced liver injury in humans. Table 3 shows the classification performance of the linear SVM for different data collapsing strategies and different data sets. Overall, AUC scores tend to be low, which shows that predicting DILI using expression data are a difficult problem. We notice that the rat in vivo repeated dose and rat in vivo single dose data reach high AUC scores when no data collapsing is applied ($0.61 \leq AUC \leq 0.67$). However, collapsing either the probesets or the time points tend to decrease the AUC scores. This suggests that in vivo data might contain important information related to DILI prediction that is partially lost during data collapsing.

In contrast, the human in vitro data achieves its lowest AUC score when no pre-processing is applied (AUC = 0.52). Collapsing either the probesets or the time points tends to increase the AUC scores, although not as high as with the rat in vivo data. This suggests that even though the goal is to predict human DILI, using in vivo data from rats is more informative than using in vitro data from humans. In contrast, the rat in vitro data achieves the lowest AUC scores. This is not surprising as it combines the two limitations of the three other data sources: it is not human, and it is not in vivo.

Table 4 shows the AUC scores of the RBF kernel SVM. We observe that the highest AUC scores are achieved with the in vivo data, which confirms our similar observation with the linear SVM. But overall, the classification performance is not better than that of the linear SVM. The highest score of the kernel SVM is AUC = 0.66, compared with 0.67 for the linear SVM. This is caused by an overfitting problem: when dealing with high dimensional problems (on our case, few hundreds samples vs. tens of thousands of features), additional model complexity is not desirable.[6] These results suggest that improvements in prediction performance can be achieved by using appropriate gene filtering strategies,[7] and by combining unprocessed rat in vivo data with processed (collapsed) human in vitro data.

Materials and Methods

Data set description. We used the data set provided by the Japanese toxicogenomics project.[1] Using Affymetrix arrays, the effects of 131 drugs on the liver have been measured both in vivo (using rats) and in vitro (using human and rat hepatocytes). Among them, 101 drugs were assigned into one of the following categories, according to their FDA-approved drug labels: Most DILI concern, less DILI concern, or no DILI concern. Each drug has been tested in various conditions (time points and drug dosages), and the whole data set contains approximately 20,000 arrays. The raw CEL files have been processed using MAS5.[8-10] For the comparison between in vivo and in vitro data, three drugs (adapin, carbon tetrachloride, and chlorpromazine) were ignored because of missing values.

Statistical testing analysis. We applied empirical Bayes statistics (http://bioinf.wehi.edu.au/limma/) to test for agreement between in vivo and in vitro probesets at corresponding time points. In the in vivo data set, we identified which probesets were differentially expressed. In the in vitro data set, we computed the p values obtained from statistical testing. If the in vitro experiments were representative of the in vivo measurements, then these

in vitro p values would be low when the corresponding probesets are differentially expressed in vivo, and high otherwise.[11,12] This is a binary classification problem, and therefore, the agreement between in vivo and in vitro data can be measured by the AUC score. If AUC = 1, then the in vitro data can be used to predict which probesets are differentially expressed in vivo. If AUC = 0.5, then in vitro experiments are useless to predict them.

Gene set enrichment analysis procedure. *Data overlap between IVT and IVV.* As part of our data exploration, GSEA was used to compare the overlap between in vitro and in vivo expression data, and confirm that unsigned expression patterns between them are significantly similar. GSEA was initially used to enrich sets of expression data with biological functions[13] by comparing a ranked list from the expression data with a predefined list associated with the function of interest. During the analysis, GSEA compares the location of the elements in the predefined list with those in the ranked list, calculating the statistical significance of the total representation of the predefined list within the ranked list. The test statistic calculated determines whether the predefined list is enriched within the ranked list or not. The GSEA script used here was downloaded from the Broad Institute (www.broadinstitute.org/gsea). Related geneset-level based approaches include.[14,15]

The analyses compared in vitro and in vivo using absolute ranked lists of probe fold change values (i.e., $\log_2$ ratio). To keep the run time of the analyses reasonably low, only the top 1% of the lists were used as predefined lists. As a result the predefined lists are approximately 200–300 probes in length. The predefined lists were then enriched against the full list of the opposite experimental condition. As the goal of this data exploration was to see similarities between expression patterns, expression data was used regardless if they were significantly differentiated or not. Thus, an example enrichment analysis consisted of a predefined list of the top 1% absolute fold change in vitro data at time 2 h between control and acarbose, and a ranked list of all absolute fold change in vivo data at time 3 h between control and acarbose. GSEA was performed at two time points, 2 h for in vitro and the equivalent 3 h for in vivo, and 24 h for both in vitro and in vivo; and in both directions, in vitro data as the predefined list and in vivo data as the ranked list and vice versa. In total, four GSEA analyses were performed for each of the 131 drugs.

The expression data was normalized using MAS5 in the limma package[16] in R 2.15.1.[17] Due to the low number of replicates, fold change was calculated by combining control and low dosage into condition one, and medium and high dosage into condition two. The lists detailed above were created from the MAS5 output as described and used in the GSEA script. The output is an enrichment score and the associated p value. A p value less than 0.05 indicates that the top 1% fold change probes are expressed similarly, as a set, between in vivo and in vitro data.

Data overlap between IVT and liver function, and IVV and liver function. Following the original usage of GSEA, we also enriched the expression data against a biological function; following the main theme of the study, several liver functions were selected using gene ontology. Our goal was to see if enrichment analyses against gene ontologies would yield the same results in in vivo and in vitro data. AmiGO[18] was used to select 32 gene ontologies associated with liver which was represented by at least one probe in the data set. The 180 probes were ranked by the total number of child nodes their gene ontology had. The resulting list started with the probes of gene ontologies that had no child nodes and ended with the probes of gene ontologies with the most child nodes. These predefined gene ontology lists were then enriched against the full ranked lists in place of the top 1% lists used previously. As per the previous analyses, four GSEA analyses were performed for each of the 131 drugs.

Fold change analysis. From the raw CEL files, we extracted the MAS5 probeset-level values using LIMMA. Then, we averaged those values over biological replicates. We computed the fold changes for each condition (drug, dose and time point), i.e., the $\log_2$ ratios between the sample values and the corresponding control values.

Data collapsing strategies. We collapsed probesets into genes, by computing the average intensity of the probesets in each gene. We also collapsed each time series by computing their average absolute fold change. To evaluate our data collapsing strategy, we considered a binary classification problem where the top 1% genes with the highest in vivo average fold change were defined as true positives, and the remaining genes were defined as the true negatives. The corresponding average in vitro fold changes were used as prediction scores. When the collapsing strategy involved absolute values, we considered the average absolute fold changes instead of the signed values. The classification performance, which reflects the agreement between in vivo and in vitro data, was measured by the AUC. If AUC = 1, then the in vitro experiments can be used to predict high in vivo fold changes. If AUC = 0.5, then the in vitro experiments are useless to predict them.

Correlation matrix analysis. For a given gene, one way to assess the agreement between its in vivo and in vitro expression levels is to define a distance function between the two sets of measurements. We used the correlation matrix distance.[19-21] For each gene, we defined two correlation matrices characterizing the gene's responses to drugs in vivo and in vitro. If the Frobenius norm of the difference between these two correlation matrices is small, then the corresponding gene behaves similarly in vivo and in vitro. When running downstream analysis of in vitro data, this approach can be used to filter out inconsistent genes, and keep the genes that show high correlation between in vivo and in vitro data.

Predicting drug-induced liver injury in humans. We considered the DILI prediction problem as a binary classification of "Most DILI" against "Less DILI or no DILI." For each available data source, we considered all DILI-annotated drugs and doses with no missing data. The resulting human in vitro, rat in vitro, rat in vivo repeated dose, and rat in vivo single dose data contained 223, 303, 303, and 301 samples, respectively. Each sample corresponds to a (drug, dose) pair. The probeset space contained 54,675 probesets for humans and 31,099 probesets for rats. The gene space contained 20,026 genes for humans and 13,878 genes for rats. We used the linear SVM classifier and RBF kernel SVM classifier,[22] and evaluated their classification performance using a 10-fold cross validation. In the case of the RBF kernel, we trained a hard-margin SVM[23] (i.e., $C = +\infty$).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work is supported by the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST program), initiated by the Council for Science and Technology Policy (CSTP). See www.tkl.iis.u-tokyo.ac.jp/FIRST/.

References

1. Uehara T, Ono A, Maruyama T, Kato I, Yamada H, Ohno Y, et al. The Japanese toxicogenomics project: application of toxicogenomics. Mol Nutr Food Res 2010; 54:218-27; PMID:20041446; http://dx.doi.org/10.1002/mnfr.200900169
2. Chen M, Vijay V, Shi Q, Liu Z, Fang H, Tong W. FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today 2011; 16:697-703; PMID:21624500; http://dx.doi.org/10.1016/j.drudis.2011.05.007
3. Mulrane L, Rexhepaj E, Smart V, Callanan JJ, Orhan D, Eldem T, et al. Creation of a digital slide and tissue microarray resource from a multi-institutional predictive toxicology study in the rat: an initial report from the PredTox group. Exp Toxicol Pathol 2008; 60:235-45; PMID:18479893; http://dx.doi.org/10.1016/j.etp.2007.12.004
4. McBurney RN, Hines WM, Von Tungeln LS, Schnackenberg LK, Beger RD, Moland CL, et al. The liver toxicity biomarker study: phase I design and preliminary results. Toxicol Pathol 2009; 37:52-64; PMID:19171931; http://dx.doi.org/10.1177/0192623308329287
5. Kiyosawa N, Manabe S, Sanbuissho A, Yamoto T. Gene set-level network analysis using a toxicogenomics database. Genomics 2010; 96:39-49; PMID:20363313; http://dx.doi.org/10.1016/j.ygeno.2010.03.014
6. Vapnik V. *The nature of statistical learning theory.* Springer-Verlag New York, Inc., New York, NY USA 1995
7. Lazar C, Taminau J, Meganck S, Steenhoff D, Coletta A, Molter C, et al. A survey on filter techniques for feature selection in gene expression microarray analysis. IEEE/ACM Trans Comput Biol Bioinform 2012; 9:1106-19; PMID:22350210; http://dx.doi.org/10.1109/TCBB.2012.33
8. Hubbell E, Liu WM, Mei R. Robust estimators for expression analysis. Bioinformatics 2002; 18:1585-92; PMID:12490442; http://dx.doi.org/10.1093/bioinformatics/18.12.1585
9. Pepper SD, Saunders EK, Edwards LE, Wilson CL, Miller CJ. The utility of mas5 expression summary and detection call algorithms. BMC Bioinformatics 2007; 8:273; PMID: 17663764; http://dx.doi.org/10.1186/1471-2105-8-273
10. Irizarry R. From cel files to annotated lists of interesting genes. In Gentleman R, Carey VJ, Huber W, Irizarry RA, and Dudoit S, editors, Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Statistics for Biology and Health, pages 431–442. Springer New York, 2005
11. Jeffery IB, Higgins DG, Culhane AC. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics 2006; 7:359; PMID:16872483; http://dx.doi.org/10.1186/1471-2105-7-359
12. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001; 98:5116-21; PMID:11309499; http://dx.doi.org/10.1073/pnas.091062498
13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102:15545-50; PMID:16199517; http://dx.doi.org/10.1073/pnas.0506580102
14. Efron B, Tibshirani R. On testing the significance of sets of genes. Ann Appl Stat 2007; 1:107-29; http://dx.doi.org/10.1214/07-AOAS101
15. Jiang Z, Gentleman R. Extensions to gene set enrichment. Bioinformatics 2007; 23:306-13; PMID:17127676; http://dx.doi.org/10.1093/bioinformatics/btl599
16. Smyth GK. Limma: linear models for microarray data, pages 397–420. Springer, New York, 2005
17. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0
18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al.; The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. Nat Genet 2000; 25:25-9; PMID:10802651; http://dx.doi.org/10.1038/75556
19. Luis A. Degree of polarization for three-dimensional fields as a distance between correlation matrices. Opt Commun 2005; 253:10-4; http://dx.doi.org/10.1016/j.optcom.2005.04.046
20. Luis A. Degree of coherence for vectorial electromagnetic fields as the distance between correlation matrices. J Opt Soc Am A Opt Image Sci Vis 2007; 24:1063-8; PMID:17361292; http://dx.doi.org/10.1364/JOSAA.24.001063
21. Herdin M, Czink N, Özcelik H, Bonek E. Correlation matrix distance, a meaningful measure for evaluation of non-stationary mimo channels. In Vehicular Technology Conference, pages 136–140, 2005
22. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286:531-7; PMID:10521349; http://dx.doi.org/10.1126/science.286.5439.531
23. Mordelet F, Vert JP. Sirene. Bioinformatics 2008; 24:76-82; http://dx.doi.org/10.1093/bioinformatics/btn273

Inference of Gene Regulatory Networks to Detect Toxicity-Specific Effects in Human Embryonic Stem Cells

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

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

Reiko Nagano, Hideko Sone
Research Center for Environmental Risk
National Institute for Environmental Studies
Tsukuba, Japan
nagano.reiko@gmail.com
hsone@nies.go.jp

Satoshi Imanishi, Seiichiroh Ohsako
Center for Disease Biology and Integrative Medicine,
Graduate School of Medicine, The University of Tokyo
Tokyo, Japan
imanishi@m.u-tokyo.ac.jp
ohsako@m.u-tokyo.ac.jp

Abstract—**Environmental chemicals are known to cause serious developmental problems in embryos. To prevent injurious chemical effects, knowledge of the chemical toxicity mechanisms in human embryos is important. To reveal the functional mechanisms in living cells, inferring a gene regulatory network is a useful approach. We applied our developed statistical methods based on Structural Equation Modeling to infer the gene regulatory networks in human embryonic stem cells. In this study, we improved the SEM approach and applied this enhanced version to expression profiles in human embryonic stem cells exposed to various chemicals. For almost all of the tested chemicals, the cell differentiation-related genes and the neuron development-related genes were intermixed in the inferred networks. Since the chemicals' networks displayed diffusion type shapes, the effects of chemical toxicity are considered to affect a few target genes at first, and then ultimately many genes via regulatory mechanisms. Furthermore, the genes that were finally affected were conserved among chemicals with the same toxicity: Tuj1 in Neurotoxic chemicals, Oct3/4 and Pax6 in Genotoxic chemicals, and Oct3/4 in Carcinogenic chemicals. These finally affected genes are considered to be the results of toxicity-specific effects in ES cells, and they reflected the features of the toxicity. We also found that some chemicals shared the same regulatory mechanism. The detected toxicity-specific effects are valuable for developing methods to prevent chemicals from disturbing normal development.**

Keywords-Structural Equation Modeling; Gene Regulatory Network; Embryonic Stem Cell; Environmental Chemicals

I. INTRODUCTION

We are exposed to many chemicals, which are produced by our usual life activities. Since the toxicity of environmental chemicals is known as one of the typical factors causing developmental toxicity, we investigate the specific effects of chemical toxicity [1]. Developmental toxicity is either a structural or functional alteration, and these alterations interfere with the normal developmental programming in early embryos. These interferences can cause abnormal development and diseases [2][3]. For example, Methylmercury is known as a developmental toxin that affects fetal development [4][5]. Furthermore, certain chemicals can cause serious developmental problems and abnormal cell differentiation in embryos [6][7][8].

To prevent the harmful effects of chemicals, elucidation of the toxic stress response in embryonic cells is crucial [9][10]. A gene regulatory network is a useful approach to reveal the regulatory mechanisms in living cells. Using 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 [11][12]. In our previous 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 [13][14]. However, GGM infers only the undirected graph, whereas the Boolean and Bayesian models infer the directed graph, which shows causality. Although all of these approaches are suitable for establishing the relationships among the genes, they cannot reveal the relationships between un-observed factors and genes, due to insufficient information in the gene expression profiles. To clarify the mechanisms of biological processes in living cells, un-observed factors that affect the target gene's expression should also be considered. Thus, an alternative approach that includes un-observed factors should be applied.

Recently, we developed a new statistical approach, based on Structural Equation Modeling (SEM) in combination with factor analysis and a four-step procedure [15][16]. This approach allowed us to reconstruct a model of transcriptional regulation that involves protein-DNA interactions from only the gene expression data, in the absence of protein

information [15]. The significant features of SEM are the inclusion of latent variables within the constructed model and the ability to infer the network, including its cyclic structure. Furthermore, the SEM approach allows us to strictly evaluate the inferred model by using fitting scores. The SEM approach is useful for detecting the causality among selected genes, as the linear relationships between genes are assumed to minimize the difference between the model's covariance matrix and the calculated sample covariance matrix [17][18][19]. Some fitting indices are defined for evaluating the model adaptability, and thus the most suitable model can be selected by SEM [1][19].

Here, we applied the SEM approach to infer the regulatory network among 9 development-related genes. The mRNA levels of these 9 genes were measured in human embryonic stem cells exposed to 15 environmental chemicals. The chemicals were considered to have developmental toxicities that adversely affect the developmental process in human embryos. Thus, inferring the gene regulatory network among development-related genes will help us to elucidate the toxic stress response in the human embryo. Furthermore, we improved our SEM approach for constructing preliminary initial models from the time-series data, in the absence of known regulatory interactions among the genes. We applied this improved SEM approach to infer the chemical-specific regulatory network among the development-related genes.

II. MATERIALS AND METHODS

A. Expression data

We utilized expression data that were measured to clarify the effects of chemical toxicity on neuronal differentiation [7][20]. In these expression data, nine genes considered to be affected by chemicals were measured in human embryonic stem cells: GATA2, Nanog, Oct3/4, Nodal, Lmx1A, MAP2, Nestin, Pax6, and Tuj1 [7][20]. Among the 9 genes, GATA2, Nanog, Oct3/4, and Nodal are mainly related to cell differentiation, and the other genes are related to neuron development. As an internal control, the expression of beta-actin was also measured. The expression data of these 10 genes were obtained from human embryonic stem cells exposed to 15 chemicals: Methylmercury (MeHg), 2-Nitropropane (2-NP), Acrylamide (ACA), p-Nitroaniline (p-NA), 4-hydroxy PCB107 (PCB), Benzo[a] pyrene (BZP), Diethylnitrosamine (DENA), Diethylaminofluorene (DEAF), Phenobarbital (PB), Tamoxifen (TMX), Diethylstilbestrol (DES), TCDD (TCDD), Thalidomide (THAL), Bisphenol-A (BPA), and Permethrin (PER) [7][20]. The toxicity of each chemical was classified into one of four types: Neurotoxic (MeHg, 2-NP, ACA, p-NA, and PCB), Genotoxic (BZP, DENA, and DEAF), Carcinogenic (PB, TMX, DES, and TCDD), and others (THAL, BPA, and PER). The human embryonic cells were exposed to each chemical for several time periods: 24 hours, 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 genes was measured twice under each condition by RT-PCR, and thus 600 (15 chemicals x 4 time periods x 5 concentration types x 2 repeats) expression patterns per gene were measured [20].

First, the expression level of each gene was normalized to the internal beta-actin control and averaged, 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. Multi-factor analysis of variance

In this study, the data contained three factors that affect gene expression: chemicals, exposure times, and concentrations. To detect the significant factors for differences in gene expression, we applied the analysis of variance (ANOVA) for multiple factors [21]. Although the multi-factor ANOVA model includes each factor's effect and all combinations of interactions between the factors, the triple interactions among the factors were confounded with error terms, because the data lacked repetition [21]. Therefore, we used the linear effects model for analysis:

$$E_{ijk}=\mu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+\varepsilon_{ijk} \quad (2)$$

where E_{ijk} is the expression level of each gene under one condition, μ is the averaged value of all measured data, α_i is the effect of factor A, β_j is the effect of factor B, γ_k is the effect of factor C, $(\alpha\beta)_{ij}$ is the interaction between factors A and B, and ε_{ijk} is the error term.

Depending on the linear effects model, the total some of squares, S_{Total} could be decomposed into the following components:

$$S_{Total}=S_A+S_B+S_C+S_{AB}+S_{AC}+S_{BC}+S_e \quad (3)$$

where S_A, S_B, and S_C mean the sum of the squared differences between each factor's marginal mean and the overall mean; S_{AB}, S_{AC}, and S_{BC} mean the sum of the squared differences for particular corresponding data means, marginal means, and overall mean; and S_e measures the difference between S_{Total} and the total sum of squares of all effects. The degree of freedom for S_{Total} was the number of all observed data minus one, and the degrees of freedom for S_A, S_B, and S_C were the number of levels for the factor minus one. The mean square values for S_A, S_B, and S_C were the sums of the squares divided by the numbers of degrees of freedom. In S_e, the degree of freedom was the total degrees of freedom minus the sum of the factor degrees of freedom. The mean square of S_e was the sum of the squares divided by the number of degrees of freedom. In the analysis of variance, S_{Total} accounted for the

factor effects (S_A, S_B, S_C, S_{AB}, S_{AC}, S_{BC}) and the contribution of S_e.

To compare the factor effects, the statistical F-test was used. The F statistic is the mean square for the factor divided by the mean square of the error terms. This F statistic is known to follow an F distribution with degrees of freedom for each factor effect and degrees of freedom for the error terms. Thus, we could calculate the probabilities of the factor effects from the F statistics.

C. Extraction of causalities from expression data

In an SEM analysis, an initial model should be assumed, but no regulations were defined among the selected genes in this study. Thus, we had to construct an initial model among the 9 genes for each chemical. To detect the regulatory relationships between the gene pairs from the measured time series expression data, we applied cross correlation coefficients 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 [22]. The cross correlation values ranged between −1 and +1. In a time series analysis, the cross correlation between two time series describes the normalized cross covariance function. Let $X_t = \{ x_1, \ldots, x_N \}$, $Y_t = \{ y_1, \ldots, y_N \}$ represent two time series data including N time points. The cross correlation is then given by

$$r_{xy} = \frac{\sum_{t=1}^{N}(x_i - \bar{x})(y_{i+d} - \bar{y})}{\sqrt{\sum_{t=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{t=1}^{N}(y_{i+d} - \bar{y})^2} \quad (4)}$$

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

D. Construction of the initial model

In this study, we inferred the chemical-specific regulatory network, and thus the differences between times and concentrations could be merged for the construction of the initial model. Fig. 1 shows the new method developed for constructing an initial model of each chemical, with the merging of several conditions. First, we constructed lag matrices to merge the time difference. The time difference was summarized by the time lag values in the cross correlations among genes. Since the time lags indicated the order of the expression pattern among the gene pairs, the rough causality between all gene pairs could be extracted. In this study, three cross correlations were calculated with three lags, −1, 0, and +1, and the three absolute values of the cross correlations were compared. The value d with the highest absolute value was selected as the causal information between the gene pairs, and the selected lag value d wasarranged as a matrix element in a lag matrix.

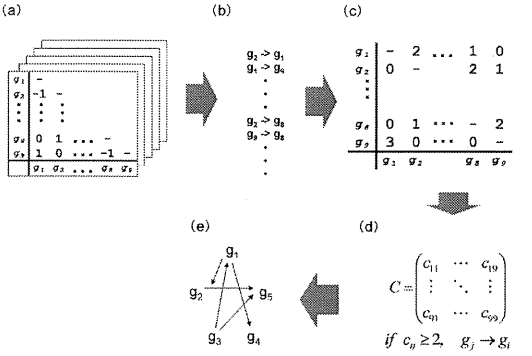

Figure 1. Procedure for initial model construction: (a) Time-lag matrices for each chemical. Five time-lag matrices were obtained for each chemical. (b) Binomial relationships. (c) Frequency matrix of causal relationships between all gene pairs. (d) Selection of possible causal relationships from the frequency matrix. (e) Construction of an initial model with selected causal relationships.

Lag Matrices were constructed for each concentration of a chemical. Thus, five time lag matrices were constructed for each chemical (Fig. 1a).

We subsequently merged the concentration difference of each chemical. For each chemical, there are five lag matrices according to the chemical concentrations, and we considered that the chemical-specific relationships among the genes would be conserved in several lag matrices. To obtain the chemical-specific relationships among the genes, we extracted the binary relationships between gene pairs from the five lag matrices for each chemical. If the same relationships existed in several lag matrices, then the binomial relationships were duplicated (Fig. 1b).

In the next step, we constructed one frequency matrix for each chemical. From the binary relationships, we counted the frequencies of all gene regulatory pairs, and each frequency number was arranged as an element of a frequency matrix (Fig. 1c). In this step, the concentration difference could be merged, since the elements of the frequency matrix indicate the information for the different concentrations. We subsequently selected the gene pairs with frequency matrix values greater than or equal to two, as the chemical-specific regulation (Fig. 1d). At the final step, we constructed an initial model for each chemical from the extracted relationships between the genes (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. By using this approach, an initial model can include cyclic structures.

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

After the construction of an initial model for each chemical, we applied the SEM calculation to infer the network model that fit the measured expression data. 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 9 genes (GATA2, Nanog, Oct3/4, Nodal, Lmx1A, MAP2, Nestin, Pax6, and Tuj1) were defined as the observed variables. Meanwhile, none were defined as latent variables, since considerations about the common regulator of several genes are dispensable for this study. The un-observed factor, which affected each gene's expression, was calculated as an error. 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:

$$y=\Lambda y+e \quad (5)$$

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 e.

The SEM software package SPSS AMOS 17.0 (IBM, USA) was used to fit the model to the data. The quality of the fit was estimated by the Chi-square statistic (CMIN), the goodness-of-fit index (GFI), which measures the relative discrepancy between the empirical data and the inferred model, and the adjusted GFI (AGFI), which is the GFI modified according to the degrees of freedom. Furthermore, we used CFI and RMSEA as fitting scores, to evaluate the model fitting. Since these indices have threshold values, as criteria to decide whether the model is suitable to obtain data independent of a huge sample number, they were considered to be useful to clarify the degree of model fitting in this study.

F. Parameter estimation

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

$$F_{ML}(S, \Sigma(\theta))=\log|\Sigma(\theta)|-\log|S|+tr(\Sigma(\theta)^{-1}S)-p \quad (6)$$

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 [23].

G. Iteration for the 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 the measured data can be selected.

In the network model, the covariance matrix between variables is calculated by the estimated parameters. The similarity between the 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 the quantitative similarity between a constructed model and an actual relationship, fitting scores are usually utilized. In this study, the quality of the fit was predicted by four different fitting scores: CMIN(Prob), GFI, AGFI, CFI, and RMSEA. The value of CMIN(Prob) is calculated by the Chi-square statistic divided by the degrees of freedom, and a CMIN(Prob) value higher than 0.05 is considered as a good model fit. 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 [24]. 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:

Step 1: Deletion of a non-significant edge from the model. Use 0.05 as the significance level for the determination of the significant regulation among the variables. After the parameters are estimated, the inverse matrix of the Fisher information matrix of parameters is calculated. The inverse matrix of Fisher information represents the asymptotic parameters' covariance matrix. The probability of each parameter is calculated by using this asymptotic parameters' matrix, since all of the parameters are usually normally distributed.

Step 2: Reconstruction of the network model. The structure of the network model without the non-significant edge is different from that of the former model. Thus, all parameters should be re-calculated from the reconstructed model, and the similarity of the network structure is also re-calculated.

Step 3: 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.

Step 4: 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 measures how much the chi-square statistic is expected to decrease if a particular parameter setting is constrained [24]. 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.

Step 5: Iteration from Steps 1 to 3. The addition of a new edge to a constructed model changes the structure of the network model. 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.

Step 6: 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 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 within the network.

H. *Extraction of association rules by affinity analysis*

We applied affinity analysis to discover the similar regulatory mechanism models among the 15 chemicals' networks. To detect the relative chemical pairs as association rules, we created a binary dataset with conserved regulations among different chemicals. According to the original definition of association rule mining [25], we defined the problem of association rule mining as follows: Let $I = \{ i_1,\ldots, i_n \}$ be a set of n binary attributes called items. Let $T = \{ t_1,\ldots, t_m \}$ be a set of database transactions. Each transaction t_k is represented by the binary vector $t_k = (t_k^1, t_k^2, \ldots, t_k^n)$, which includes n elements. The value of t_k^1 indicates the appearance of transaction t_k in item i_1. In this study, the 15 chemicals were defined as a set of items, and each conserved gene regulation between the different chemicals was considered as one transaction. Thus, the value of 1 indicated the appearance of the conserved gene regulation in the chemical's network, while the value of 0 indicated its absence.

An association rule is defined as the implication of the form $I_a => I_b$, where I_a and I_b are sets of some items in I, but some of the same items are not present in I_a and I_b. To detect the association rules, we used some constraints: support, confidence and lift. Support is defined as the proportion of transactions that contain the item set to all transactions. Thus, $support(I_a,I_b)=prob(I_a,I_b)$ was calculated as the joint probability of I_a and I_b. The confidence constraint is displayed as $conf(I_a=>I_b)$, and it is defined as the conditional probability $prob(I_a|I_b)$. Thus, we calculated $conf(I_a=>I_b)$ from the proportion of transactions with the item set I_b to the transactions with the item set I_a. The lift constraint is defined as:

$$lift(I_a=>I_b) = conf(I_a=>I_b)/prob(I_b) \quad (7)$$

Lift is a measure of the performance of an association rule with respect to the population as a whole, against the random choice. Thus, lift was obtained by calculating the ratio of the target response to the average response. In general, a lift value over 1 is suitable for association rules.

III. RESULTS AND DISCUSSION

A. *Chemical concentrations had no effect*

In this study, gene expression was measured in the presence of different concentrations of various chemicals, with several exposure times. To reveal the most effective factor for gene expression, multi-factor ANOVA was applied to the measured data. In statistics, ANOVA is utilized to detect differences between groups in terms of some variables. Usually, the chance of committing a type I error will increase by performing multiple two-sample t-tests, and a statistical test is needed to determine whether or not the means of more than two groups should be applied, such as Tukey's HSD test and so on. Although these post-hoc tests are useful for detecting the factor pairs with significant differences between them, the factor pairs are not important in this study. Instead, we wanted to determine factors, which caused gene expression differences, and thus we compared three factors: chemicals, time differences, and concentrations.

The 15 chemicals were divided into 3 categories by their toxicities: Neurotoxic chemicals, Genotoxic chemicals, Carcinogenic chemicals, and other type chemicals. We compared the gene expression differences between these toxicity types. We calculated a p-value from the F statistic for each gene. The p-value is the probability that the variation between conditions may have occurred by chance, so genes with smaller p-values vary more significantly. Thus, the gene's variation is less likely to have occurred by chance, and is conversely more likely to be connected to the difference in conditions. The probabilities of expression differences for each gene, grouped by each factor, are shown in Table I. Interestingly, the expression of all of the genes was significantly different among the chemicals and the time differences. However, the chemical concentrations showed almost no significant differences in terms of the expression of the genes. Thus, the concentrations of the chemicals had no effect on the expression of the tested genes in the ES cells.

TABLE I. RESULTS OF MULTI-FACTOR ANOVA

	Chemical (a)	Concentration (b)	Time (c)	a * b	a * c	b * c
GATA2	<0.01	0.076	<0.01	0.559	<0.01	0.450
Nanog	<0.01	<0.01	<0.01	0.011	<0.01	0.022
Oct34	<0.01	<0.01	<0.01	0.055	<0.01	0.044
Nodal	<0.01	0.130	<0.01	<0.01	<0.01	0.040
Lmx1A	<0.01	0.714	<0.01	<0.01	<0.01	0.787
MAP2	<0.01	0.479	<0.01	<0.01	<0.01	0.576
Nestin	<0.01	<0.01	<0.01	0.012	<0.01	0.548
Pax6	<0.01	0.575	<0.01	<0.01	<0.01	0.881
Tuj1	<0.01	0.011	<0.01	0.810	<0.01	0.097

a. Probabilities were calculated from the F statistics and the degrees of freedom.

b. Significant probabilities are displayed as "<0.01" in this table.

B. The complexities of the initial models are related to the chemical toxicity

We utilized our newly developed method to construct the initial gene regulatory network models under the conditions with 15 chemicals. One of the distinguishing features of our new method is its ability to include a cyclic structure in the network model. Cyclic regulation, such as feedback regulation, is considered to be important for cells to control normal gene expression, and the new method is useful to detect cyclic regulation from the gene expression data. Fig. 2 shows the constructed initial network models.

In Fig. 2, the components of the constructed models were: 9 genes with 22 relationships in MeHg, 9 genes with 23 relationships in 2-NP, 9 genes with 19 relationships in ACA, 9 genes with 23 relationships in p-NA, 9 genes with17 relationships in PCB, 9 genes with 9 relationships in BZP, 8 genes with 14 relationships in DENA, 8 genes with 10 relationships in DEAF, 8 genes with 19 relationships in PB, 9 genes with 23 relationships in TMX, 7 genes with 9 relationships in DES, 9 genes with 23 relationships in TCDD, 8 genes with 10 relationships in THAL, 6 genes with 9 relationships in BPA, and 8 genes with 10 relationships in PER. The distribution of the number of relationships according to the toxicity type is displayed in Fig. 3. In Figs. 2 and 3, the numbers of edges were obviously different, according to the chemicals' toxicity. Neurotoxic and Carcinogenic chemicals contained more relationships than

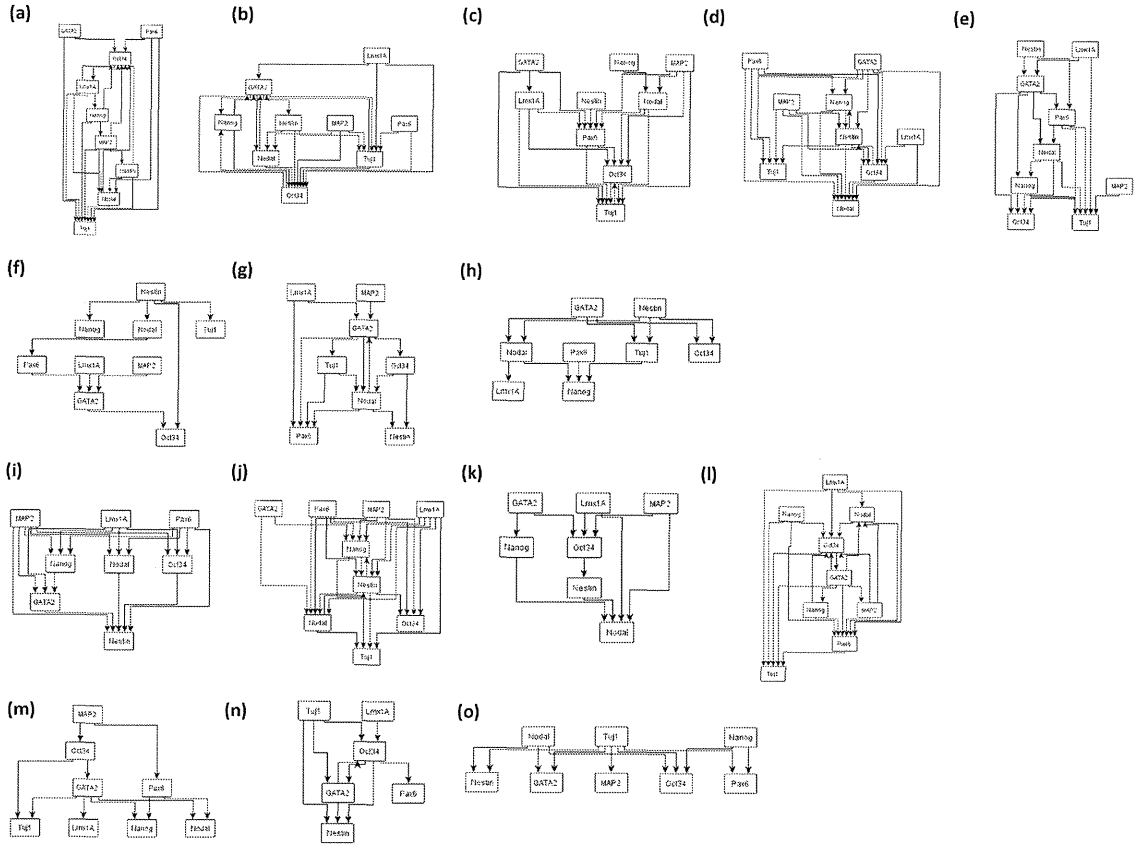

Figure 2. Initial network models: (a) MeHg, (b) 2-Np, (c) ACA, (d) p-NA, (e) PCB, (f) BZP, (g) DENA, (h) DEAF, (i) PB, (j) TMX, (k) DES, (l) TCDD, (m) THAL, (n) BPA, (o) PER. The networks with the same toxicity are arranged on the same line.

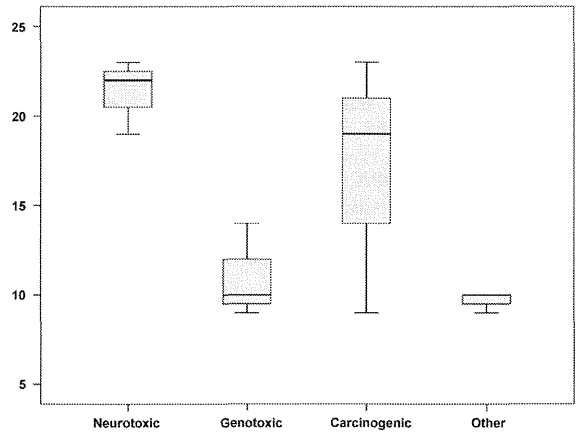

Figure 3. Box plot of edge numbers: Distribution of the number of edges in each initial model.

Genotoxic and other chemicals. Furthermore, only one or two genes were arranged as the last endogenous genes in the initial models with Neurotoxic and Carcinogenic chemicals, as opposed to two or more genes in the initial models of Genotoxic and other chemicals. Thus, the effects of the Neurotoxic or Carcinogenic chemicals were complicated, but could be summarized into only one or two target genes. In contrast, the expressions of many genes were finally affected by Genotoxic and other chemicals, via simple regulatory networks. These differences between chemical toxicity types 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. Before the SEM calculation, we simplified all of the initial models. To simplify these duplicated interactions, we only retained the longest path between two genes. In the initial model, the edges do not represent the direct regulation, but the time provenience information. In other words, the difference between direct and indirect interactions in the initial model is not very important. Thus, the regulation displayed by a direct path could be replaced by indirect paths in the model. By retaining the longest paths, all of the preceding information was included, as the simplest diagram.

C. *Structures of inferred networks*

The final inferred networks for each chemical are depicted in Fig. 4, and the goodness of fit scores are displayed in Table II. From Table II, almost all of the models were considered to fit well with the measured data by some fitting scores, CMIN(Prob), CFI, and RMSEA, except for the DES network. In the DES network, all of the fitting scores indicated that the inferred network could not be judged as a well-fitted model. Since the obtained fitting scores were the best scores in this analysis, we considered the network inference for DES to need more expression data.

The inferred networks of chemicals revealed distinct structures. The cell differentiation-related genes and the neuron development-related genes were intermixed in almost all of the inferred networks, except for MeHg and BPA. In the inferred network of MeHg, the regulations among cell differentiation-related genes and the regulation among neuron development-related genes were separated to the right and left. This specific shape means that the effects of MeHg appeared differently between neuronal and other development. This difference may be related to the two different effects of MeHg: developmental deficits in children [26], and risk of cardiovascular disease in adults [27]. On the other hand, cell differentiation-related genes and neuron development-related genes were separated at the top and bottom in the BPA network. In the BPA network, neuron development-related genes were only disturbed by cell differentiation-related genes.

TABLE II. FITTING SCORES OF INFERRED NETWORKS

	Neurotoxic					Genotoxic			Carcinogenic				Other		
	MeHg	2-NP	ACA	p-NA	PCB	BZP	DENA	DEAF	PB	TMX	DES	TCDD	THAL	BPA	PER
CMIN (*Prob*)	0.50	0.34	0.06	0.26	0.30	0.44	0.16	0.11	0.01	0.27	0.00	0.63	0.31	0.52	0.11
GFI	0.76	0.82	0.83	0.78	0.79	0.79	0.84	0.77	0.75	0.81	0.74	0.83	0.83	0.78	0.78
AGFI	0.60	0.63	0.61	0.59	0.61	0.62	0.65	0.60	0.54	0.61	0.52	0.64	0.60	0.64	0.56
CFI	1.00	0.99	0.96	0.97	0.98	1.00	0.97	0.94	0.90	0.98	0.88	1.00	0.99	1.00	0.96
RMSEA	0.00	0.07	0.15	0.10	0.08	0.03	0.12	0.14	0.21	0.09	0.23	0.00	0.08	0.00	0.14

a. Five fitting scores were utilized for measuring the fitness level between the constructed model and the measured data.

b. The well-fitted threshold of each score is: CMIN(Prob) is P>0.05, GFI > 0.90, AGFI> 0.90, CFI>0.90, RMSEA< 0.05.

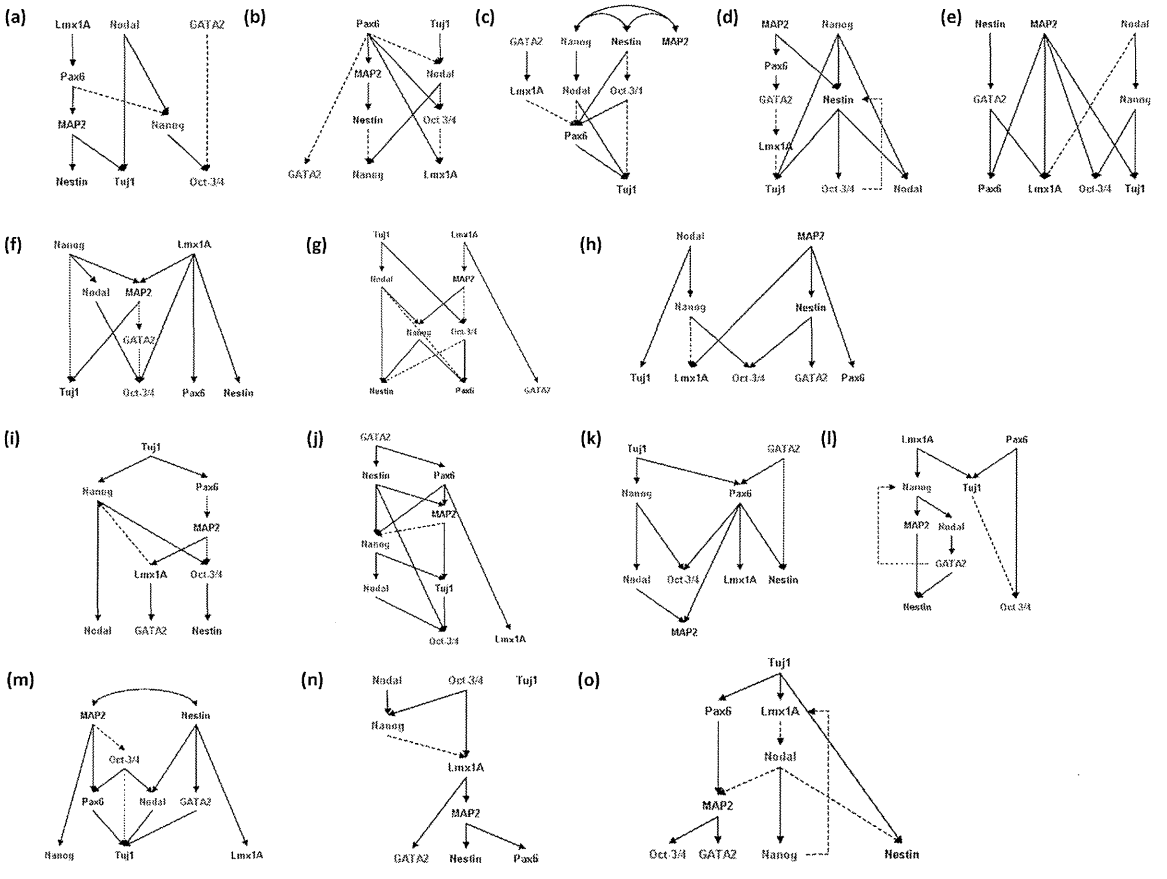

Figure 4. Inferred chemical networks: A positive relationship between genes is displayed with a solid arrow. A negative relationship between genes is displayed with a dashed arrow. Gene names with blue characters indicate "neuron development-related genes", and genes with red characters indicate "cell differentiation-related genes". (a) MeHg, (b) 2-Np, (c) ACA, (d) p-NA, (e) PCB, (f) BZP, (g) DENA, (h) DEAF, (i) PB, (j) TMX, (k) DES, (l) TCDD, (m) THAL, (n) BPA, (o) PER. The networks with the same toxicity are arranged on the same line.

Concerning the shapes of the inferred networks, we defined the network shape by comparing the numbers of genes at the top phase (*N*(*top*)) and the final phase (*N*(*bottom*)) within each chemical network. One of the specific shapes was a centralized model, which was defined as N(*top*)-N(*bottom*)≥2. In this model, many genes were arranged at the top phase, and only a few genes were arranged at the final phase in the network structure. The ACA network was the only network with a centralized model. The other specific shape was a diffusion model. The shape of a diffusion model is defined as N(*bottom*)-N(*top*)≥2. Among the well-fitted models, four networks were classified into diffusion models: BZP, DEAF, PB, and PER. The shape of the BPA network was different from those of the other networks, and resembled a bow-tie like model.

Fundamentally, the genes were hierarchically controlled in the inferred networks, but there were a few recursive relationships. Interestingly, the values of the regression weights of the recursive regulations among all of the inferred networks were negative: regulation from Oct3/4 to Nestin in the p-Na network, regulation from GATA2 to Nanog in TCDD, and regulation from Nanog to Lmx1A in PER. These recursive regulations indicated that feedback regulation exists in ES cells.

D. Detection of Toxicity-Specific Effects

To detect the specific features that were dependent on the toxicity type, we monitored the position of each gene in the inferred networks. Table III displays the number and probability of incoming edges and those of outgoing edges for each gene. Among the Neurotoxic chemicals' networks, Tuj1 has significantly few incoming edges and significantly many outgoing edges. Actually, Tuj1 was arranged as a result of network regulation in almost all of the Neurotoxic networks.

TABLE III. INTERACTING EDGES OF EACH GENE

	Neurotoxic				Genotoxic				Carcinogenic			
	OUTPUT		INPUT		OUTPUT		INPUT		OUTPUT		INPUT	
	Num	P	Num	P	Num	P	Num	P	Num	P	Num	P
Oct3/4	4	0.113	8	0.111	2	0.120	7*	0.032	1*	0.043	9*	0.042
GATA2	5	0.135	3	0.083	1	0.091	3	0.222	6	0.146	2	0.070
Lmx1A	3	0.084	7	0.141	6	0.101	2	0.145	4	0.141	3	0.114
MAP2	9	0.077	2	0.049	7	0.070	3	0.222	5	0.155	6	0.160
Nanog	7	0.132	5	0.146	7	0.070	3	0.222	7	0.118	8	0.079
Newtin	7	0.132	5	0.146	2	0.120	5	0.174	3	0.111	6	0.160
Nodal	8	0.107	5	0.146	6	0.101	2	0.145	4	0.141	4	0.154
Pax6	9	0.077	8	0.111	0	0.060	5	0.174	10*	0.025	4	0.154
Tuj1	1*	0.031	10*	0.043	2	0.120	3	0.222	6	0.146	4	0.154

a. The significant values (P<0.05) are highlighted with *.

This means that the toxicities of Neurotoxic chemicals are considered to finally affect Tuj1, which is known to contribute to microtubule stability in neuronal cells [28]. Although the expression levels of 5 genes were measured as neuron development-related genes, Tuj1 was detected as the final target of Neurotoxicity.

Among the Genotoxic chemicals' networks, Oct3/4 exhibited a significant number of incoming edges. Furthermore, both Pax6 and Oct3/4 were arranged at the lower phase in all Genotoxic networks. Oct3/4 is one of the key regulators of pluripotency [29], and Pax6 is known as a key transcription factor for the development of the cerebral cortex and other sensory organs [30]. Considering the features of both Pax6 and Oct3/4, developmental processes, such as normal cell differentiation, were disturbed by Genotoxic chemicals.

In the Carcinogenic networks, both the incoming and outgoing edges of Oct3/4 were significant, and Oct3/4 was arranged as a result in almost all of the Carcinogenic networks. The chemicals that were classified as either Genotoxic or Carcinogenic are known as carcinogens [31][32]. Thus, the Genotoxic and Carcinogenic features indicated that the chemical disturbance of Oct3/4 is related to cancer. The other feature of the Carcinogenic networks, regulation from Nanog to Nodal, was conserved among all of the Carcinogenic networks. Both Nanog and Nodal are important for normal early embryonic development. Nanog is a key factor for maintaining pluripotency in embryonic stem cells [33][34]. Nodal is related to the development of the left-right axial structure [35][36], and its signaling pathway is known to be important very early in development, for cell fate determination and many other developmental processes [36]. Although the Carcinogenic chemicals do not affect genetic structures, the regulatory mechanisms of these carcinogenic chemicals may be similar.

To compare the conserved gene relationships among chemicals with the same toxicity, we extracted the conserved gene regulations from the chemicals' networks. The numbers of conserved regulations were: 13 within Neurotoxic chemicals, 2 within Genotoxic chemicals, and 11 within Carcinogenic chemicals. Even though the average numbers of edges in the inferred models were similar among the three toxicity types (10.6 in Neurotoxic, 10.1 in Genotoxic, and 12.5 in Carcinogenic), the numbers of conserved regulations were different. From this feature, it is considered that a similar regulatory mechanism controlled the Neurotoxic chemicals' effects and the Carcinogenic chemicals' effects in ES cells, but the gene regulation by each Genotoxic chemical was independent of the toxicity type.

E. Similar mechanisms between chemicals

By utilizing the data mining method, we identified the chemicals with similar regulation. First, we constructed a transaction Table about the conserved regulation for each chemical, as shown in Table IV. Each row of data indicates the conserved regulation between genes, and each column indicates one chemical. In this transaction table, the value of 1 means that the corresponding regulation appeared with the chemical, whereas the value of 0 means that the regulation did not exist in the chemical's network.

In the affinity analysis, we set the thresholds as: Support > 0.5, Confidence > 0.5, and lift > 1. According to these restrictions, 2 rules were extracted. One is *BPA* => *DEAF*, and the other is *DEAF* => *PCB*. These results reflected the finding that the regulations in the *BPA* network were also conserved in the *DEAF* network. Furthermore, the regulations in the *DEAF* network were conserved in the *PCB* network. Although these three chemicals were categorized into different types of toxicities, they may share the same regulatory mechanisms to affect the ES cells.

IV. CONCLUSION

We applied an improved SEM approach to reconstruct a gene regulatory model from gene expression data in human embryonic stem cells. Our results confirmed that SEM is a powerful approach to estimate the gene regulation caused by chemical toxicity. The shapes of the inferred network models for the various chemicals were different, but the inferred networks had a tendency to finally affect the same gene by their toxicity type. One of the neuron development-related genes, Tuj1, was arranged as the result of almost all of the Neurotoxic toxicity networks. Furthermore, Oct3/4 was important for both the Genotoxic and Carcinogenic networks. Since the Genotoxic chemicals are also carcinogenic, Oct3/4 is considered to be carcinogenic in ES cells. We detected some specific features for each toxicity type, and thus the inferred network among genes can be utilized for the estimation of a chemical's effects, 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.

TABLE IV. TRANSACTION TABLE OF CONSERVED REGULATIONS

edge info.		Neurotoxic chemicals					Genotoxic chemicals			Carcinogenic chemicals				Other chemicals		
parent	child	MeHg	2-Np	ACA	p-NA	PCB	BZP	DENA	DEAF	PB	TMX	DES	TCDD	THAL	BPA	PER
Oct34	Lmx1A	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
Oct34	Nestin	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0
Oct34	Pax6	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
Oct34	Tuj1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
GATA2	Oct34	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
GATA2	Lmx1A	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
GATA2	Nestin	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
GATA2	Pax6	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0
Lmx1A	GATA2	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0
Lmx1A	MAP2	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
Lmx1A	Nanog	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
Lmx1A	Pax6	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
Lmx1A	Tuj1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
MAP2	Oct34	0	0	0	0	1	0	1	0	1	0	0	0	1	0	1
MAP2	GATA2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
MAP2	Lmx1A	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0
MAP2	Nanog	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0
MAP2	Nestin	1	1	0	1	0	0	0	1	0	0	0	1	0	1	0
MAP2	Pax6	0	0	0	1	1	0	0	1	0	0	0	0	1	1	0
MAP2	Tuj1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0
Nanog	Oct34	1	0	0	0	1	0	0	1	1	0	1	0	0	0	0
Nanog	Lmx1A	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Nanog	MAP2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
Nanog	Nestin	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
Nanog	Nodal	0	0	1	1	0	1	0	0	1	1	1	1	0	0	0
Nanog	Tuj1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Nestin	Oct34	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0
Nestin	GATA2	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
Nestin	Nanog	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Nestin	Nodal	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
Nodal	Oct34	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
Nodal	MAP2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Nodal	Nanog	1	1	0	0	1	0	1	1	0	0	0	0	0	1	1
Nodal	Nestin	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Nodal	Pax6	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
Nodal	Tuj1	1	0	1	0	0	0	0	1	0	1	0	0	1	0	0
Pax6	Oct34	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0
Pax6	GATA2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
Pax6	Lmx1A	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0
Pax6	MAP2	1	1	0	0	0	0	0	0	1	1	1	0	0	0	1
Pax6	Nanog	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Pax6	Tuj1	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0
Tuj1	Oct34	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0
Tuj1	Nanog	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0
Tuj1	Nodal	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
Tuj1	Pax6	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1

a. The first column indicates the starting gene of one edge, and the second column indicates the end gene of the same edge.

b. The value of 1 means that the corresponding regulation appeared with the chemical, whereas the value of 0 means that the regulation did not exist in the chemical's network.

ACKNOWLEDGMENT

This study was supported in part by the Grant in Aid for Scientific Research from the Ministry of the Health and Labor (to S. O.), Japan. The authors gratefully acknowledge the technical support of Miss Masami Yokoyama for her technical assistance.

REFERENCES

[1] S. Aburatani and W. Fujibuchi, "Application of Structural Equation Modeling for Inferring Toxicity-Dependent Regulation in Human Embryonic Stem Cells," GLOBAL HEALTH 2012 paper, pp.27-32, Oct. 2012.

[2] A. Baccarelli and V. Bollati, "Epigenetics and environmental chemicals," Curr. Opin. Pediatr. vol. 21(2), Apr. 2009, pp. 243-251, doi:10.1097/MOP.0b013e32832925cc.

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

[4] Y. Yuan, "Methylmercury: a potential environmental risk factor contributing to epileptogenesis," Neurotoxicology, vol. 33(1), Jan. 2012, pp. 119–126, doi:10.1016/j.neuro. 2011.12. 014.

[5] N. Tatsuta, et al., "Prenatal exposures to environmental chemicals and birth order as risk factors for child behavior problems," Environ. Res., vol. 114, Apr. 2012, pp. 47–52, doi: 10.1016/j.envres.2012.02.001.

[6] 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), Feb. 2012, pp. 33–40, doi:10.3109/19396368.2011.647381.

[7] X. He, et al., "Effects of methylmercury exposure on neuronal differentiation of mouse and human embryonic stem cells," Toxicol. Lett., vol. 212(1), Apr. 2012, pp. 1–10, doi: 10.1016/j.toxlet.2012.04.011.

[8] 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), Feb. 2011, pp. 368–387, doi:10.1016/j.tiv.2010.10.011.

[9] 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), Dec. 2007, pp. 4541–4554, doi: 10.1002/pmic.200700381

[10] J. Thompson and J. Bannigan, "Cadmium: toxic effects on the reproductive system and the embryo," Reprod. Toxicol., vol. 25(3), Apr. 2008, pp. 304–315, doi: 10.1016/j.reprotox.2008.02.001.

[11] 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, Aug. 2000, pp. 331-343, doi:10.1089/ 106652700750050817.

[12] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian networks to analyze expression data," J. Comput. Biol., vol. 7, Jul. 2000, pp. 601–620, doi:10.1089/ 106652700750050961.

[13] 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, Apr. 2003, pp. 777–788, doi:10.1016/S0165-1684(02)00476-0.

[14] S. Aburatani, "Development of network inference among LexA/RecA-dependent manner genes in the SOS response," J. of Nucleic Acids Investigation, vol. 1, Oct. 2010, pp. e13–e17, doi:10.4081/jnai.2010.e13.

[15] S. Aburatani, "Application of structure equation modeling for inferring a serial transcriptional regulation in yeast," Gene. Regul. Syst. Bio., vol. 5, Nov. 2011, pp. 75–88, doi:10.4137/ GRSB.S7569.

[16] S. Aburatani, "Network Inference of pal-1 Lineage-Specific Regulation in the C. elegans Embryo by Structural Equation Modeling," Bioinformation, vol. 8(14), Aug. 2011, pp. 652–657, doi:10.6026/97320630008652.

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

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

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

[20] W. Fujibuchi, et al., "Prediction of Chemical Toxicity by Network-based SVM on ES-cell Validation System," Proc. of the 2011 Joint Conference of CBI-Society and JSBi, pp.47, Nov. 2011.

[21] J. Neter, M. Kutner, W. Wasserman, and C. Nachtsheim, Applied Linear Statistical Models, 5th ed., Chicago: Irwin, 2006.

[22] L. Li and G. E. Caldwell, "Coefficient of cross correlation and the time domain correspondence," Journal of Electromyography and Kinesiology, vol. 9, Dec. 1999, pp. 385–389, doi:10.1016/S1050-6411(99)00012-7.

[23] 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.

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

[25] R. Agrawal, T. Imieliński, A. Swami, "Mining association rules between sets of items in large databases," Proc. of SIGMOD'93, May 1993, pp. 207–216, doi:10.1145/170035. 170072.

[26] D. C. Rice, R. Schoeny, K. Mahaffey, "Methods and rationale for derivation of a reference dose for methylmercury by the U.S. EPA," Risk analysis: an official publication of the Society for Risk Analysis, vol. 23(1), Feb. 2003, pp. 107–115, doi:10.1111/1539-6924.00294.

[27] A. L. Choi, et al., "Methylmercury Exposure and Adverse Cardiovascular Effects in Faroese Whaling Men," Environmental Health Perspectives, vol. 117(3), 2009, pp. 367–372, doi:10.1289/ehp.11608.

[28] 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.

[29] L. H. Looijenga, et al., "POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors," Cancer Res., vol. 63(9), pp. 2244–2250, May 2003.

[30] Y. Arai, et al., "Role of Fabp7, a Downstream Gene of Pax6, in the Maintenance of Neuroepithelial Cells during Early Embryonic Development of the Rat Cortex," J. Neurosci., vol. 25(42), pp. 9752–9761, Oct. 2005.

[31] N. Ito, et al., "Pathological markers for non-genotoxic agent-associated carcinogenesis," Toxicol. Lett., Vol. 64–65, Dec. 1992, pp. 613-620, doi:10.1016/0378-4274(92)90239-G.

[32] 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), Jul. 2006, pp. 397–404, doi:10.1016/ j.etp.2006.01.005.

[33] K. Mitsui, "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells," Cell, vol. 113(5), May 2003, pp. 631–642, doi:10.1016/ S0092-8674(03)00393-3.

[34] I. Chambers, et al., "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells," Cell, vol. 113(5), May 2003, pp. 643–655, doi:10.1016/ S0092 -8674(03)00392-1.

[35] H. Hamada, C. Meno, D. Watanabe, and Y. Saijoh, "Establishment of vertebrate left-right asymmetry," Nat. Rev. Genet., vol. 3(2), Feb. 2002, pp. 103–113, doi:10.1038/nrg 732.

[36] C. Grandel and N. H. Patel, "Nodal signaling is involved in left-right asymmetry in snails," Nature, vol. 457(7232), Feb. 2009, pp. 1007–1011, doi:10.1038/nature07603.

Original Paper

Infiltration of CD40-Positive Tumor-Associated Macrophages Indicates a Favorable Prognosis in Colorectal Cancer Patients

Makoto Kinouchi[1,2], Koh Miura[2], Takayuki Mizoi[2], Kazuyuki Ishida[3], Wataru Fujibuchi[4], Hiroyuki Sasaki[2], Shinobu Ohnuma[2], Kazuya Saito[5], Yu Katayose[2], Takeshi Naitoh[2], Fuyuhiko Motoi[2], Ken-ichi Shiiba[1,2], Shinichi Egawa[2], Chikashi Shibata[2] and Michiaki Unno[2]

1Department of Surgery, Miyagi Cancer Center, Miyagi, Japan
2Department of Surgery, Tohoku University Graduate School of Medicine, Sendai, Japan
3Department of Pathology, Tohoku University Hospital, Sendai, Japan
4National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
5Department of Surgery, Sendai Postal Services Agency Hospital, Sendai, Japan

Corresponding author: Koh Miura, Department of Surgery, Tohoku University Graduate School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai 980-8574, Japan; Tel.: +81-22-717-7205, Fax: +81-22-717-7209; E-mail: k-miura@surg1.med.tohoku.ac.jp

ABSTRACT

Background/Aims: Recently the role of tumor-associated macrophages (TAMs) on immunity has been variously discussed. We studied a series of cell surface antigens in TAMs in colorectal cancer tissues and their corresponding normal tissues using flow cytometry to find out prognostic indicators of these patients. **Methodology:** We assessed the numbers of CD14+ macrophages positive for each of the cell surface antigens (CD80, CD86, HLA-DR, CD1a, CD40 and CD83) in cancer tissues and corresponding normal tissues among 31 patients with colorectal cancer, and performed the univariate and multivariate analysis to find out prognostic indicators for overall survival among the patients. **Results:** The numbers of CD80+, CD86+ and HLA-DR+ TAMs in the cancer tissues were higher than those in corresponding normal tissues. Inversely CD40+ and CD83+ macrophages in cancer tissues were less than those in normal tissues. With the multivariate analysis, the number of CD40+ TAMs, as well as lymph node metastasis and distant metastasis, was shown to be an independent prognostic factor of colorectal cancer patients. **Conclusions:** The dense infiltration of CD40+ TAM in colorectal cancer tissues indicates a favorable prognosis, which suggests that CD40 plays an important role in the tumor immunity of colorectal cancer.

Key Words: CD40; Tumor-associated macrophage; Colon cancer; Prognostic indicator.

Abbreviations: Dendritic Cell (DC); Natural Killer (NK); Antigen Presenting Cell (APC); Tumor-Infiltrating Lymphocyte (TIL); Tumor-Infiltrating Dendritic Cell (TIDC); T Cell Receptor (TCR); Tumor-Associated Macrophage (TAM); Lymph Node (LN); Colorectal Cancer (CRC); Overall Survival (OS); Monoclonal Antibody (mAb); Phycoerythrin (PE); Fluorescein Isothiocyanate (FITC); Fluorescence-Activated Cell Sorting (FACS).

INTRODUCTION

The anti-tumor immune response is regulated by a variety of innate immune cells, such as lymphocytes, dendritic cells (DCs), natural killer (NK) cells, macrophages and neutrophils (1); among which macrophages are precursor cells of DCs which act as antigen presenting cells (APCs) (2). In the anti-tumor immune response, the number of tumor-infiltrating lymphocytes (TILs) has been reported to correlate with the prognosis of patients with several types of malignancies (3) while the tumor-infiltrating dendritic cells (TIDCs) are reported to have independent positive prognostic relevance in breast cancer (4) and some other malignancies. For the effective anti-tumor immune response, firstly T cell activation is essential which is induced by the binding of MHC class II molecules to T cell receptors (TCRs) (5), and secondly the stimulation of CD28 molecules on cell surfaces of T cells is required which is induced by the binding of CD28 ligands as co-stimulatory molecules (6). CD14 has been used as a marker of tissue macrophages (7,8); in our previous study (9) we reported that CD14+ tumor-associated macrophages (TAMs) were distributed more predominantly at the invasive front of colorectal cancer (CRC) tissues rather than in corresponding normal tissues. We also reported that in CRC patients with lymph node (LN) metastasis, the 5-year overall survival (OS) rate of those with higher number of TAMs (the high CD14 group) was better than that of the low CD14 group. CD1a is a marker of immature DCs (10) and CD83 is a marker of mature DCs (10) and we further identified that in both cancer and corresponding normal tissues the populations of CD1a+CD14- cells (immature DCs) and CD83+CD14- cells (mature DCs) were quite limited compared with the population of CD14+ TAMs (9).

In addition, CD86 and CD80, both of which are known as co-stimulatory factors necessary for T cell activation, were found to be expressed in a large population of TAMs (9); for these reasons, the expressions of CD1a, CD83, CD86, and CD80 may have important roles on enhancing the anti-tumor activity of TAMs. On the other hand, CD40 is a co-stimulatory molecule that should be up-regulated for APC maturation and it plays an important role in promoting

Hepato-Gastroenterology 2013; 60:83-88 doi 10.5754/hge12372