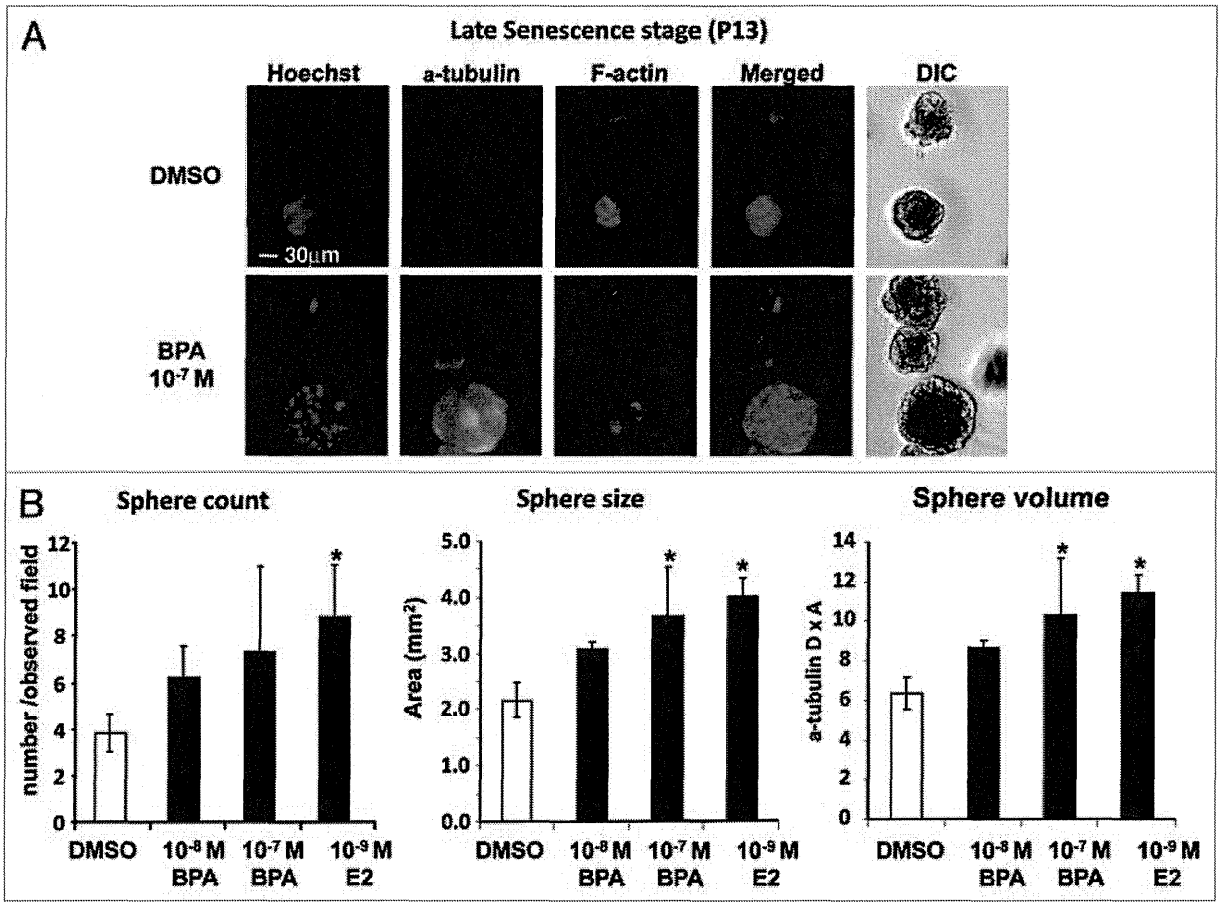


Figure 2. Effects of BPA exposure on colony formation of HMEC. Cells were treated with BPA or E2 at passage 8 (7 d period). (A) Morphology of colonies derived from HMEC at passage 13. Cells were stained with α-tubulin and F-actin antibodies. Cells were also counterstained with Hoechst, and the merged images and differential interference contrast (DIC) images are also shown (30 μm magnification). (B) Statistical analysis of sphere count, size and volume of colonies derived from HMEC at passage 13. *$p < 0.05$ vs. the DMSO control.

passages 10 and 15. As shown in **Figure 1B and C**, proliferation of HMEC was significantly enhanced through treatment with 10^{-7} M BPA and 10^{-9} M E2 at passage 10 and passage 15, respectively. These observations indicated that increased proliferation induced by exposure to BPA or E2 at an early passage persists during later cell divisions.

Effects of BPA exposure on the colony formation of HMEC. We found that BPA and E2 treatment enhanced cell proliferation, possibly also resulting in enhanced cellular senescence. To elucidate whether the enhanced cell proliferation by BPA and E2 had any effect on the carcinogenic status of HMEC, we employed a 3D "on-top" assay that offers the advantages of both 2D and 3D analysis.[27]

Cells covered with Matrigel produced rounded ductal colonies similar to spheres comprising a monolayer of epithelial cells, indicating that ductal formation had successfully occurred through coating with the Matrigel. The morphological status of HMEC at passage 13, which had been treated with BPA at passage 8, was assessed via the 3D "on-top" assay (**Fig. 2A**). Significant increases in sphere size of HMEC were found in the 10^{-7} M BPA and 10^{-9} M E2 treatment groups (**Fig. 2B**). The 3D "on-top" assay demonstrated that exposure to BPA and E2 at passage 8 can affect sphere formation in HMEC at later passages.

Recent studies have described that the balance between proliferation and senescence is important to develop cancer when normal cells are damaged by exogenous stimuli.[14,15] We then wished to determine whether BPA exposure could alter the balance between proliferation and senescence in HMEC. Therefore, we simultaneously examined the expression of cell cycle, proliferation and senescence markers in HMEC at passages 10 and 13, 2 to 5 weeks post-chemical treatment. Three-color fluorescence imaging analyses characterized the distribution of cells of various stages with Hoechst, HP1γ and BrdU (**Fig. 3A**). Hoechst staining indicates the number of nuclei in all cells. BrdU incorporation into nuclei represents identification of cells in the early S phase. HP1γ-positive cells at the early senescence stage (passage 10) were significantly increased by exposure to 10^{-7} M BPA (**Fig. 3B**, left part). At the late senescence stage (passage 13), a significant increase in HP1γ-positive cells was observed by exposure to both 10^{-8} M and 10^{-7} M BPA (**Fig. 3B**, right part). A significant increase in BrdU-positive cells was demonstrated by exposure to 10^{-8} M BPA at passages 10 and 13 (**Fig. 3C**).

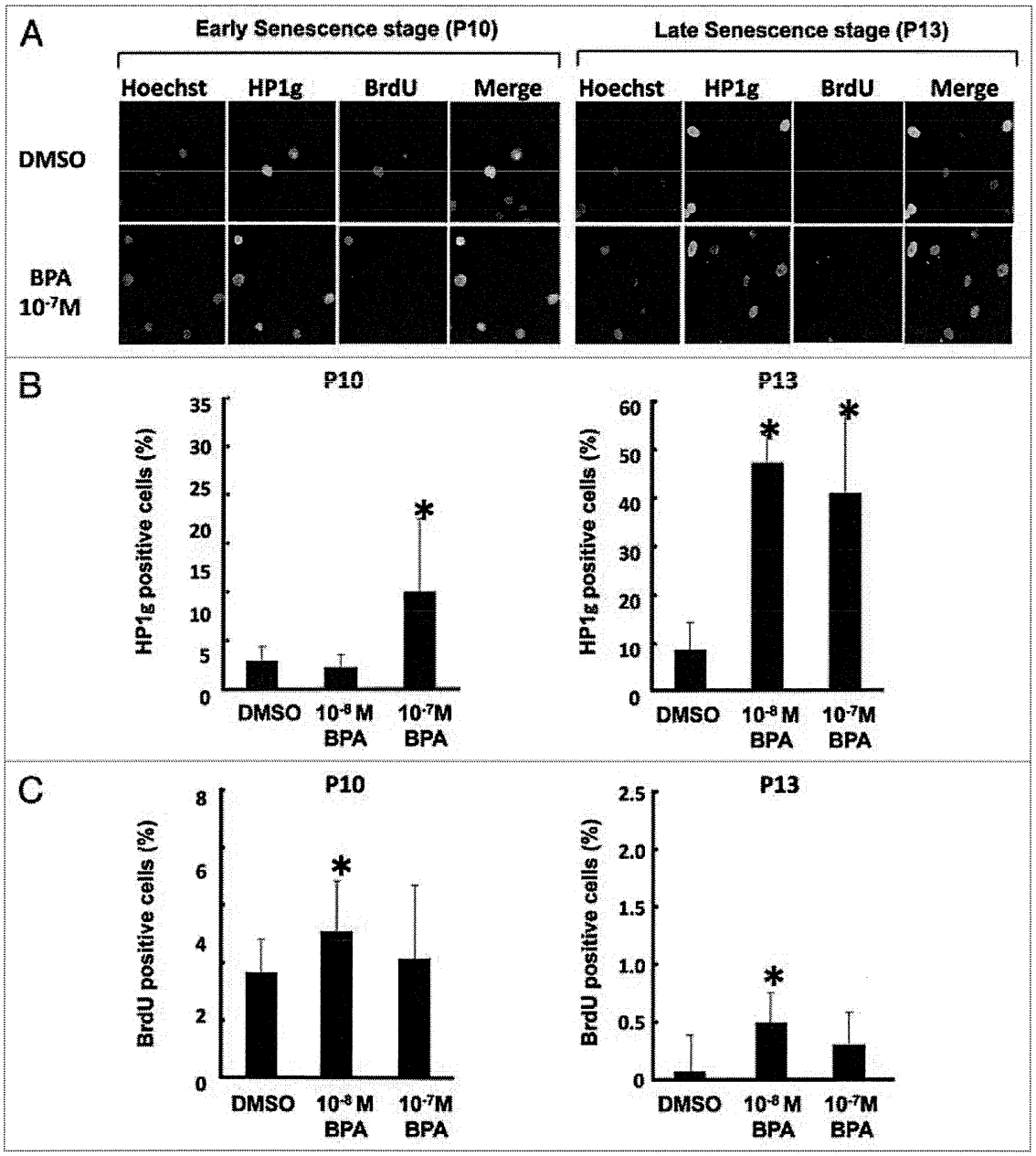


Figure 3. Effects of BPA exposure on cellular senescence of HMEC. Cells were treated with BPA at passage 8 (7 d period). (A) Staining of HMEC at passages 10 and 13 with HP1 γ or BrdU antibodies. Cells were also counterstained with Hoechst, and the merged images are also shown (200x magnification). (B) Number of HP1γ-positive cells. (C) Number of BrdU positive cells. *$p < 0.05$ vs. the DMSO control.

Gene and protein expression analysis. To determine the effects of BPA on the cellular growth of HMEC at the transcriptional level, gene expression analyses using a PCR array for mammary cancer-related genes were performed in HMEC at passage 11. Genes showing differential expression with BPA exposure are summarized in **Figure 4A**. It is noteworthy that the downregulated genes are associated with cell cycle control in many cases. A knowledge-based gene interaction network was then analyzed to determine how BPA at a dose of 10^{-8} M plays a role in signaling associated with cell cycle control (**Fig. 4B**). *CCNE1*, *CCNA2* and *CDKN2A*, which are among the key molecules underlying G_1-S control during the cell cycle, were downregulated or unchanged in BPA-treated HMEC. Other factors related to cell growth such as *EGFR*, *ERBB2*, *PTGS2* and *IGFBP2* were increased by BPA treatment at 10^{-8} M or 10^{-7} M (**Fig. 4A**). These observations indicate that the increased proliferation we observed in the BPA-treated cells may have been due to enhanced G_1-S progression resulting from the decreased expression of negative cell cycle regulators.

Knowledge-based network analysis was then performed to further clarify these findings. Network analysis for gene expression profiling at 10^{-8} M of BPA found that Tp53 upregulated CDKN2A and CCNA2, and indirectly interacted with CCNE1, and that the

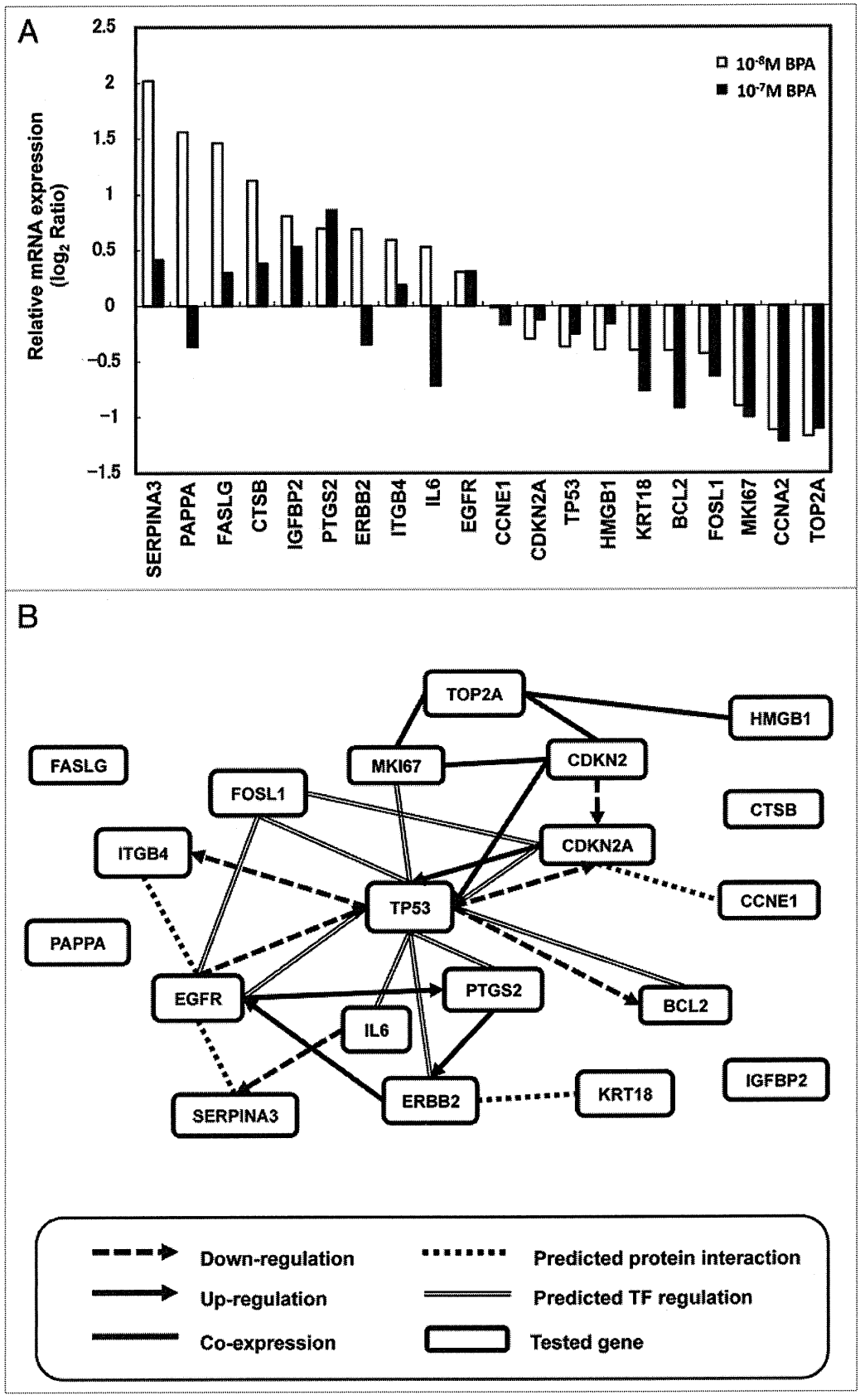


Figure 4. For figure legend, see page 301.

Figure 4 (See previous page). Gene expression and network analysis. HEMC were treated with BPA or E2 at passage 8 (7 d period) and the relative mRNA expression of selected genes was measured at passage 11. (A) Effects of BPA exposure on cancer signaling gene expression. The results are expressed as the average of two independent experiments. Relative mRNA expression normalized to β-actin is shown as the log2 ratio, with the fold-change referring to the DMSO control cells. (B) Gene networks representing key genes for BPA exposure at 10^{-8} M were identified using GNCPro (SA Biosciences).

upregulation of FASLG, CTSB, IGFBP2 and PAPPA were not associated with TP53.

We also investigated the effects of BPA exposure on protein expression of p16 (CDKN2A), p53 (TP53) and Cyclin E (CCNE1) in HMEC using western blot analysis. As shown in **Figure 5**, BPA exposure did not have a significant effect on p53 protein expression, while a dose-dependent increase in p16 protein expression was observed. Downregulation at 10^{-8} M BPA and upregulation at 10^{-7} M BPA of Cyclin E protein expression were also observed.

DNA methylation patterns. Alterations in DNA methylation patterns are associated with the development of a variety of human cancers, including breast cancer.[28,29] Previous studies have found promoter hypermethylation of in situ lesions and identified aberrant methylation at the promoters of candidate genes, which include *GSTP1, CCND2, RARB2, TWIST1, RASSF1A, HIC1, CDKN2A, SFN (TP53), BRCA1, CCNA1, THBS1, TNFRSF* and *APC1*.[30-33] Therefore, BPA-induced changes in the methylation status of 24 gene promoters were investigated via quantitative real-time PCR arrays in this study. **Table 1** shows that seven gene promoters exhibited changes of more than 10% in hypermethylation status between DMSO control and BPA-exposed cells (**Table 1**). These genes exhibited an increased percentage of promoter hypermethylation by BPA exposure including *BRCA1, CCNA1, CDKN2A, THBS1, TNFRSF10C* and *TNFRSF10D*, while *HIC1* had decreased promoter hypermethylation by BPA exposure.

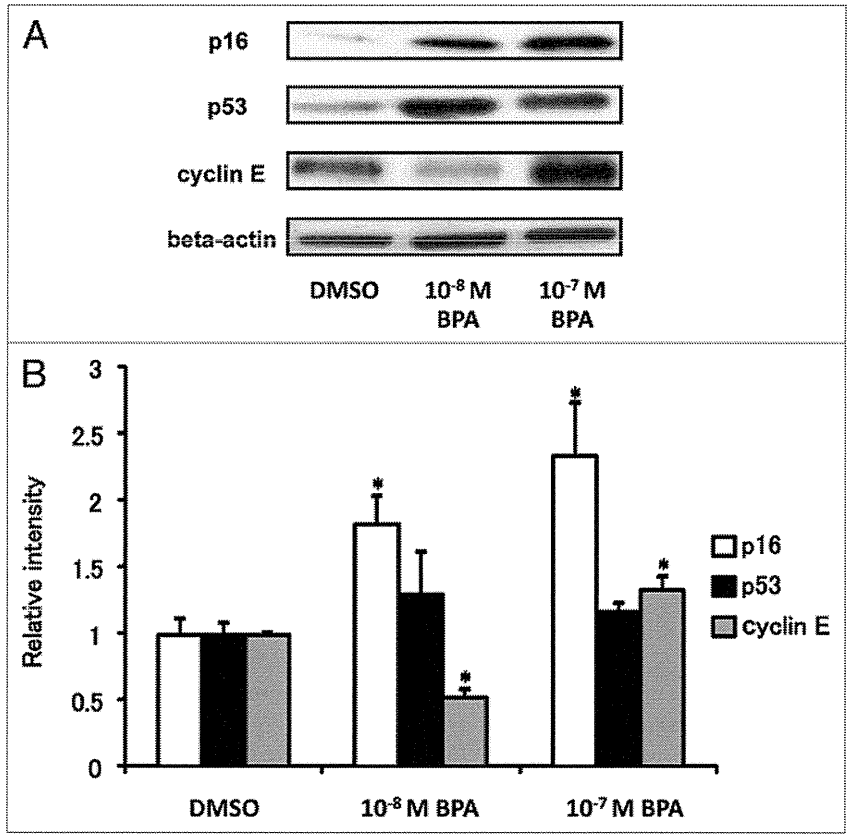


Figure 5. Modulation of the protein expression of p16, p53 and cyclin E in HEMC by BPA exposure. (A) Cells were treated with 10^{-8} M or 10^{-7} M BPA at passage 8 (7 d period) and protein expression of p16, p53 and cyclin E was measured at passage 11. (B) The cellular protein levels were calculated using ImageJ densitometry software and are expressed as the mean ± SD relative to DMSO control after normalizing the bands to β-actin. *$p < 0.05$ vs. the DMSO control.

Discussion

To investigate how BPA affects the carcinogenesis process in normal breast cells, we exposed HMEC to this agent at an early passage for a duration of 1 week and examined subsequent effects on cell proliferation, gene expression and DNA methylation at the stage of later passages. HMEC are a model system for studying early events in mammary tumorigenesis, and they have a normal karyotype and enter senescence after a lengthy culture, while neoplastic cells are allowed to continuously grow, thus overcoming the barrier of cellular senescence.[34] HMEC morphology and growth status in in vitro culture assay have been shown to be closely connected with malignancy in a comparison involving 25 mammary-gland-derived cell lines.[27]

In the current study, we focused on the effects of early-passage exposure to BPA in later-passage HMEC. In non-treated HMEC, the rate of cell growth slowed at approximately passage 15 (**Fig. 1A**), indicating the onset of cellular senescence, as HMEC are not immortalized. In contrast, HMEC treated with E2 or BPA did not show a reduced cell proliferation rate at approximately passage 15 (**Fig. 1A**). Our previous study showed that the telomerase activity of BeWo cells was enhanced by their exposure to E2 or 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, a known endocrine disrupting agent.[35] This suggests that BPA might also upregulate telomerase activity in HMEC, resulting in an extended lifespan until a crisis point is eventually reached and cell division ceases.

To clarify the effects of BPA and E2 treatment on cell growth, a 3D "on-top" assay was performed using Matrigel, which is the optimal coating material for HMEC in this assay. Our optimized 3D "on-top" assay showed that BPA and E2 enhanced the

Table 1. Promotor methylation status of genes in HMEC at passage 12

Symbol	Ref Seq	DMSO			BPA 10^{-8} M			Differences (BPA-DMSO)		
		HM	UM	IM	HM	UM	IM	HM	UM	IM
BRCA1	NM_007294	3.4%	96.6%	0.0%	24.5%	75.5%	0.0%	21.2%	-21.2%	0.0%
CCNA1	NM_003914	7.2%	92.8%	0.0%	36.0%	64.0%	0.0%	28.8%	-28.8%	0.0%
CCND2	NM_001759	11.5%	29.2%	59.3%	12.8%	47.3%	40.0%	1.3%	18.0%	-19.3%
CDKN2A[#]	NM_058195	3.8%	96.2%	0.0%	16.7%	83.3%	0.0%	12.9%	-12.9%	0.0%
CDKN2A[#]	NM_000077	5.0%	8.3%	86.7%	19.3%	28.3%	52.4%	14.3%	20.0%	-34.3%
GSTP1	NM_000852	3.0%	97.0%	0.0%	2.3%	55.9%	41.9%	-0.8%	-41.1%	41.9%
HIC1	NM_006497	97.2%	2.8%	0.0%	68.2%	31.8%	0.0%	-29.0%	29.0%	0.0%
THBS1	NM_003246	5.5%	94.5%	0.0%	36.9%	63.1%	0.0%	31.4%	-31.4%	0.0%
TNFRSF10C	NM_003841	12.5%	54.9%	32.6%	24.8%	75.2%	0.0%	12.2%	20.4%	-32.6%
TNFRSF10D	NM_003840	5.1%	94.9%	0.0%	17.7%	82.3%	0.0%	12.6%	-12.6%	0.0%

HM, hypermethylated; UM, un-methylated, IM intermediately methylation. Cells were treated with vehicle (DMSO) and 10^{-8} M BPA at passage 8 (7 d period). [#]indicated that different probes of the promoter region of *CDKN2A* were used in this assay.

nuclear count for each HMEC colony and increased the area of these colonies at passage 13. Kenny et al.[27] reported morphological differences among colonies under 3D "on-top" analysis of 25 mammary cancer-derived cell lines, and they could be classified into four groups of "round," "mass," "grape-like" and "satellite." Relatively fewer malignant cells show round-shaped colonies in the 3D "on-top" assay, and most malignant grades result in grape-like or satellite shapes.[27] In our study, HMEC showed a "mass" shape (**Fig. 2**), and treatment with BPA or E2 did not change the shape of the colonies, but increased the size and cell numbers for each colony. These results indicated that BPA exposure at passage 8 had a subsequent effect on the cell growth of HMEC. The increased nuclear count and body area of spheres indicated overgrowth of the differentiated colony, possibly indicating a hyperplastic state. This may provide a key insight into the potential adverse effects of BPA upon the status of the mammary gland, which may result in carcinogenesis.

BPA exposure was found to increase the number of HP1γ-positive cells in HMEC at passages 10 and 13 in our study (**Fig. 3A and B**). At the late senescence stage (passage 13), both 10^{-8} M and 10^{-7} M BPA exposure increased HP1γ-positive cells but not BrdU-positive cells. A previous study reported that the onset of senescence induces an increase in the number of positive nuclear bodies that contain HP1γ.[36,37] HP1γ protein is also known to be positive for the entire nuclear area in the early S stage[38] and the G_2/S stage of the cell cycle.[39,40] When non-immortalized cells are close to their cell division limit, they often show positive SA-βgal activity in the cytoplasm and senescence-associated DNA foci in nuclei containing HP1γ.[37] Therefore, there are several possible explanations for the increased nuclear HP1γ-positive nuclei by BPA exposure in this study, such as enhanced cellular senescence or a delayed cell cycle progression during the G_1/S stage.

To elucidate the effects of BPA on enhanced cellular senescence or a delayed cell cycle progression during the G_1/S stage in HMEC, gene and protein expressions of mammary cancer-related genes were investigated in this study. Cyclin E protein expression was increased by exposure to BPA at the high dose (10^{-7} M), but not at the low dose (10^{-8} M) (**Fig. 5**). This might partly explain our findings that BPA exposure at the high dose but not the low dose at passage 8 significantly increased HEMC growth at passages 10 and 13 (**Fig. 1B and C**). Cyclin E is cyclically expressed during the cell cycle, and it binds and activates the cyclin-dependent kinase Cdk2 and catalyzes the transition from the G_1 phase to the S phase.[41,42] Our findings suggest that Cyclin E might play an important role in BPA-induced cell growth of HMEC. However, a discrepancy between Cyclin E gene and protein expression following BPA treatment was observed in our study (**Figs. 4A and 5**). The regulation of Cyclin E is through transcriptional regulatory mechanisms or through protein degradation by the proteasome pathway.[43] It is known that SCF (Skp1-Cullin-F-box) ubiquitin ligases regulate the degradation of many proteins involved in the control of cell division and growth.[44,45] Indeed, it has been reported that the amount of Cyclin E protein present in the cell is tightly controlled by ubiquitin-mediated proteolysis and one ubiquitin ligase responsible for Cyclin E ubiquitination is known as SCF^{Fbw7}.[46,47] Furthermore, it was found that phosphorylation within N- and C-terminal legions of Cyclin E plays a critical role in the binding of Cyclin E to SCF^{Fbw7} and thus its ubiquitination and proteasomal degradation.[48] Therefore, one possible explanation of our results is that BPA exposure might affect the phosphorylation of Cyclin E. Interestingly, a recent study that analyzed the effects of a low dose of BPA in a testicular cell line revealed that this compound induces the activation of cAMP response-element-binding protein and strongly induces the phosphorylation of retinoblastoma protein (Rb).[17] These findings support our hypothesis that the effect of BPA exposure on mammary cell proliferation is related with the dysregulation of cell cycle regulatory genes, since the activation of Rb is also known to enhance the cell cycle, particularly at the G_1-S transition. An animal study recently found that BPA exposure can significantly accelerate mammary tumorigenesis and metastasis in MMTV-erbB2 mice, and one of the underlying mechanisms includes the regulation of phosphorylation of proteins involved in the Akt pathway.[49] Because the overexpression of Cyclin E has been related to progression of a variety of cancers and constitutive expression of Cyclin E leads to genomic instability, further

study on the mechanism by which long-term BPA exposure could mediate Cyclin E expression might provide an insight into the potential carcinogenic activity of BPA in the mammary gland.

Another interesting finding of our study is that BPA exposure appeared to promote cellular senescence and proliferation of HEMC simultaneously (**Figs. 1–3**). This was supported by our findings that BPA exposure increased p16 protein expression in a dose-dependent manner; p16 has recently been found to promote aging in murine cells.[50] Indeed, there has been increased attention on the balance between convergent and divergent mechanisms of cancer and aging.[14,15] Both cancer and aging are fuelled by the accumulation of cellular damage. Additionally, one concern regarding BPA is that differing doses of BPA may function in distinctly different, and sometimes opposing, manners at both the tissue and molecular level.[49] Our study suggests that effects of BPA exposure at differing doses on the balance of cancer and aging in the mammary gland might partly contribute to this issue.

Eckhardt et al.[51] reported that approximately one-third of the differentially methylated 50-UTRs were inversely correlated with transcription in normal tissues. Radpour et al. and other researchers also reported that 10 hypermethylated genes (*APC*, *BIN1*, *BMP6*, *BRCA1*, *CST6*, *ESRb*, *GSTP1*, *CDKN2A*, *CDKN1A* and *TIMP3*) were identified for distinguishing between cancerous and normal tissues.[52-54] Therefore, hypermethylation of tumor suppressor genes causes the inactivation of genes that are important in suppressing the development of most or all tumor types. In our study, we observed increases in DNA hypermethylation of *BRCA1*, *CCNA1*, *CDKN2A*, *THBS1* and *TNFRSF* in BPA-exposed HMEC (**Table 1**). We were interested in *CDKN2A* hypermethylation status by exposure to BPA because *CDKN2A* and *GSTP1* showed significantly ($p < 0.002$) higher mean methylation levels in increasing grades (I, II, III) of invasive ductal breast cancer in a study by Moelan et al. Interestingly, both of the two *CDKN2A* probes increased levels of DNA hypermethylation by exposure to BPA, although DNA hypermethylation levels of *GSTP1* were not altered in our study. *CDKN2A* (p16), an inhibitor of the cyclin D-dependent protein kinases, is a tumor suppressor gene, and is altered in several tumor types.[56] Correlation of *CDKN2A* hypermethylation with CDKN2A protein loss has been previously reported and its loss of function has been associated with the development of a variety of cancers.[56-58] Our finding is inconsistent with these previous studies, which found that exposure to BPA in HMEC at passage 8 induced CDKN2A protein expression at passage 11 and promoter hypermethylation at passage 12. However, a study using tumor tissues derived from patients diagnosed with endometrial carcinoma found that loss of nuclear p16 protein expression is not associated with promoter methylation.[59] Our study indicates that the mechanism by which BPA upregulates CDKN2A protein expression appears to be complicated and further study is required for clarification.

In summary, our present study is the first to address the long-lasting effects of BPA exposure over multiple cellular passages of HMEC. The underlying mechanism might include genetic and epigenetic dysregulation of cell cycle regulatory genes or tumor suppressor genes.

Materials and Methods

Chemicals. Dimethyl sulfoxide (DMSO) and E2 were obtained from Sigma Chemical Co. BPA was obtained from Wako Industries. DMSO was used as the primary solvent for all chemicals, and DMSO solutions were further diluted in cell culture media for treatment. The final concentrations of DMSO in the media did not exceed 0.1% (vol/vol).

Cell culture and chemical treatment. HMEC were obtained from Cambrex Bio Science and maintained in accordance with the supplier's instructions. Briefly, the cells were cultured in plastic dishes with MEGM SingleQuots medium in an incubator at 37°C in 5% CO_2. HMEC were supplied at passage 7, and were grown to passage 8 prior to use in the experiments. BPA at final concentrations of 10^{-7} M and 10^{-8} M or E2 at 10^{-9} M was added to the culture media of passage 8 in HMECs and maintained for a period of 1 week. During this week, the culture media were changed twice with media containing the original BPA concentrations. The culture was then further grown until it was harvested for phenotype and gene expression analysis in media without chemicals. The cumulative population doublings (PDs) as a determinant of cellular lifespan were measured as previously described in reference 26. Briefly, the total number of cells harvested from each subculture was calculated and the number of accumulated PDs per passage was determined by the equation PD = (A/B)/log2, where A is the number of harvested cells and B is the number of plated cells.[26] Data were obtained from 6 duplicate cultures at each passage. The cumulative PDs from passage 8 to 16 were measured twice to validate the reproducibility of the results. The 3D "on-top" in vitro culture assay was performed essentially as described previously with different biocoated plates[27] (BD BioCoat Cellware Matrigel or Collagen Type I Coated Cellware plate, BD Biosciences).

Immunofluorescence cytochemistry. HMEC were transferred to 3D culture systems at passage 11 and maintained until passage 13. HMEC were then fixed in 4% neutralized paraformaldehyde solution for 60 min and blocked with 3% normal goat serum (NGS)/0.5% Triton X-100 in phosphate-buffered saline (PBS). The primary antibodies used were mouse monoclonal antibodies for human heterochromatin protein-1γ (HP1γ, S-19; Santa Cruz Biotechnology, sc-101004), and a rabbit polyclonal antiserum to α-tubulin (Abcam, ab15246). The secondary antibodies were anti-mouse Alexa 546 and Alexa Fluor 488 goat anti-rabbit IgG antibodies (Invitrogen). For the measurement of 5-bromo-2'-deoxyuridine (BrdU) incorporation during DNA synthesis, a cell proliferation fluorescence kit (GE Healthcare, 25-9001-89) was used according to the instruction manual. DNA was visualized by Hoechst staining (Wako Industries) and F-actin was visualized using Alexa Fluor 568 phalloidin (Invitrogen). Immunofluorescence staining signals were detected with an IN Cell Analyzer 1000 (GE Healthcare), a multiple-imaging analyzer, and morphological analysis was performed using IN Cell Investigator image analysis software (GE Healthcare).

Gene expression analysis. Total RNA was extracted using an RNeasy kit (Qiagen) when cells at passage 11 were approximately

70% confluent. These preparations were then used to detect the expression of 83 genes (**Table S1**), which have been reported to be frequently expressed in mammary cancers using a real-time RT-PCR method, Superarray-qPCR (Estrogen Receptor Signaling PCR Array, SA Biosciences). Gene expression was normalized by β-actin expression and set to 1 for the control DMSO-treated cells.

Western blot analysis. HMEC treated with BPA for a period of 1 week at passage 8 were lysed at passage 11 using RIPA buffer (Santa Sruz Biotechnology). After boiling at 99°C for 5 min, the protein samples were resolved by sodium dodecyl sulfate (SDS) PAGE on a 4–20% gel and transferred to a polyvinylidene difluoride membrane (Bio-Rad Laboratories). After blotting in tris-buffered saline (TBS) with 5% nonfat dry milk-Tris buffered saline and 0.1% Tween, the membrane was probed with p16 (1:10,000 dilution, Abcam, ab51243), p53 (1:200 dilution, Santa Cruz Biotechnology, sc-126), cyclin E (1:200 dilution, Santa Cruz Biotechnology, sc-198), and β-actin (1:200 dilution, Santa Cruz Biotechnology, sc-7210) primary antibodies. Blots were then incubated with horseradish peroxidase-conjugated anti-rabbit or anti-mouse secondary antibodies (ECL plus western blotting reagent pack, 1:10,000 dilution, GE Healthcare, RPN2124). The immune complex was detected with the Amersham ECL Plus western blotting Detection System (GE Healthcare, RPN2132). The blots were exposed to Hyperfilm (Amersham Pharmacia Biotech), and bands were quantified with ImageJ densitometry software (National Institutes of Health).

DNA methylation pattern assay. Genomic DNA from HMEC at passage 12 was isolated using the Qiagen DNeasy kit according to the manufacturer's instructions. Differentially methylated fractions of DNA were then prepared using a Methyl-Profiler DNA Methylation Enzyme Kit (SA Biosciences). After DNA had been digested, digested DNA samples were prepared with real-time PCR using a Methyl-Profiler DNA Methylation PCR Array (SA Biosciences, MEAH-011A). The MEAH-011A was loaded with 24 gene promoters.

Bioinformatics and statistical analysis. Gene networks representing key genes for BPA exposure were identified using GNCPro (SA Biosciences), which is a free online software and an in silico research tool for collating gene and pathway interactions with integrating collective biological knowledge through text mining, data mining, data acquisition and computational prediction. The interactions among a group of genes are represented graphically and are interactive. All experiments in this study were performed independently two or more times to test the reproducibility of the results. Quantitative data are expressed as the means ± SD, except those for mRNA expression levels, which are expressed as the mean of two independent experiments. A non-parametric test, the Mann-Whitney U test, was applied to test for statistical significance. Values of $p < 0.05$ were considered to indicate statistical significance.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of the Health and Labor, Japan. The authors gratefully acknowledge the critical advice of Dr. Tohru Inoue (National Institute of Health Sciences, Japan), and the technical support of Ms. Noriko Oshima (GE Healthcare Japan Corporation) in the analysis using the IN Cell Analyzer 1000. The authors also thank Ms. Yumi Matsumoto for her technical assistance.

Supplementary Material

Supplemental material can be found at:
www.landesbioscience.com/journals/cbt/article/18942/

References

1. Vandenberg LN, Hauser R, Marcus M, Olea N, Welshons WV. Human exposure to bisphenol A (BPA). Reprod Toxicol 2007; 24:139-77; PMID:17825522; http://dx.doi.org/10.1016/j.reprotox.2007.07.010.
2. Brotons JA, Olea-Serrano MF, Villalobos M, Pedraza V, Olea N. Xenoestrogens released from lacquer coatings in food cans. Environ Health Perspect 1995; 103:608-12; PMID:7556016; http://dx.doi.org/10.1289/ehp.95103608.
3. Calafat AM, Kuklenyik Z, Reidy JA, Caudill SP, Ekong J, Needham LL. Urinary concentrations of bisphenol A and 4-nonylphenol in a human reference population. Environ Health Perspect 2005; 113:391-5; PMID:15811827; http://dx.doi.org/10.1289/ehp.7534.
4. Krishnan AV, Stathis P, Permuth SF, Tokes L, Feldman D. Bisphenol-A: an estrogenic substance is released from polycarbonate flasks during autoclaving. Endocrinology 1993; 132:2279-86; PMID:8504731; http://dx.doi.org/10.1210/en.132.6.2279.
5. Keri RA, Ho SM, Hunt PA, Knudsen KE, Soto AM, Prins GS. An evaluation of evidence for the carcinogenic activity of bisphenol A. Reprod Toxicol 2007; 24:240-52; PMID:17706921; http://dx.doi.org/10.1016/j.reprotox.2007.06.008.
6. Markey CM, Luque EH, Munoz De Toro M, Sonnenschein C, Soto AM. In utero exposure to bisphenol A alters the development and tissue organization of the mouse mammary gland. Biol Reprod 2001; 65:1215-23; PMID:11566746.
7. Weber Lozada K, Keri RA. Bisphenol A increases mammary cancer risk in two distinct mouse models of breast cancer. Biol Reprod 2011; 85:490-7; PMID:21636739; http://dx.doi.org/10.1095/biolreprod.110.090431.
8. Lamartiniere CA, Jenkins S, Betancourt AM, Wang J, Russo J. Exposure to the endocrine disruptor bisphenol a alters susceptibility for mammary cancer. Horm Mol Biol Clin Investig 2011; 5:45-52; PMID:21687816; http://dx.doi.org/10.1515/HMBCI.2010.075.
9. Ptak A, Wrobel A, Gregoraszczuk EL. Effect of bisphenol-A on the expression of selected genes involved in cell cycle and apoptosis in the OVCAR-3 cell line. Toxicol Lett 2011; 202:30-5; PMID:21277958; http://dx.doi.org/10.1016/j.toxlet.2011.01.015.
10. Dairkee SH, Seok J, Champion S, Sayeed A, Mindrinos M, Xiao W, et al. Bisphenol A induces a profile of tumor aggressiveness in high-risk cells from breast cancer patients. Cancer Res 2008; 68:2076-80; PMID:18381411; http://dx.doi.org/10.1158/0008-5472.CAN-07-6526.
11. Singleton DW, Feng Y, Yang J, Puga A, Lee AV, Khan SA. Gene expression profiling reveals novel regulation by bisphenol-A in estrogen receptor-alpha-positive human cells. Environ Res 2006; 100:86-92; PMID:16029874; http://dx.doi.org/10.1016/j.envres.2005.05.004.
12. Diel P, Olff S, Schmidt S, Michna H. Effects of the environmental estrogens bisphenol A, o,p'-DDT, p-tert-octylphenol and coumestrol on apoptosis induction, cell proliferation and the expression of estrogen sensitive molecular parameters in the human breast cancer cell line MCF-7. J Steroid Biochem Mol Biol 2002; 80:61-70; PMID:11867264; http://dx.doi.org/10.1016/S0960-0760(01)00173-X.
13. Weng YI, Hsu PY, Liyanarachchi S, Liu J, Deatherage DE, Huang YW, et al. Epigenetic influences of low-dose bisphenol A in primary human breast epithelial cells. Toxicol Appl Pharmacol 2010; 248:111-21; PMID:20678512; http://dx.doi.org/10.1016/j.taap.2010.07.014.
14. Serrano M, Blasco MA. Cancer and ageing: convergent and divergent mechanisms. Nat Rev Mol Cell Biol 2007; 8:715-22; PMID:17717516; http://dx.doi.org/10.1038/nrm2242.
15. Collado M, Blasco MA, Serrano M. Cellular senescence in cancer and aging. Cell 2007; 130:223-33; PMID:17662938; http://dx.doi.org/10.1016/j.cell.2007.07.003.

16. Iso T, Futami K, Iwamoto T, Furuichi Y. Modulation of the expression of bloom helicase by estrogenic agents. Biol Pharm Bull 2007; 30:266-71; PMID:17268063; http://dx.doi.org/10.1248/bpb.30.266.
17. Bouskine A, Nebout M, Brucker-Davis F, Benahmed M, Fenichel P. Low doses of bisphenol A promote human seminoma cell proliferation by activating PKA and PKG via a membrane G-protein-coupled estrogen receptor. Environ Health Perspect 2009; 117:1053-8; PMID:19654912.
18. Moral R, Wang R, Russo IH, Lamartiniere CA, Pereira J, Russo J. Effect of prenatal exposure to the endocrine disruptor bisphenol A on mammary gland morphology and gene expression signature. J Endocrinol 2008; 196:101-12; PMID:18180321; http://dx.doi.org/10.1677/JOE-07-0056.
19. de Waard WJ, Aarts JM, Peijnenburg AA, Baykus H, Talsma E, Punt A, et al. Gene expression profiling in Caco-2 human colon cells exposed to TCDD, benzo[a]pyrene and natural Ah receptor agonists from cruciferous vegetables and citrus fruits. Toxicol In Vitro 2008; 22:396-410; PMID:18061397; http://dx.doi.org/10.1016/j.tiv.2007.10.007.
20. Buterin T, Koch C, Naegeli H. Convergent transcriptional profiles induced by endogenous estrogen and distinct xenoestrogens in breast cancer cells. Carcinogenesis 2006; 27:1567-78; PMID:16474171; http://dx.doi.org/10.1093/carcin/bgi339.
21. Garbe JC, Holst CR, Bassett E, Tlsty T, Stampfer MR. Inactivation of p53 function in cultured human mammary epithelial cells turns the telomere-length dependent senescence barrier from agonescence into crisis. Cell Cycle 2007; 6:1927-36; PMID:17671422; http://dx.doi.org/10.4161/cc.6.15.4519.
22. Stampfer MR, Bartley JC. Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. Proc Natl Acad Sci USA 1985; 82:2394-8; PMID:3857588; http://dx.doi.org/10.1073/pnas.82.8.2394.
23. Sudo H, Garbe J, Stampfer MR, Barcellos-Hoff MH, Kronenberg A. Karyotypic instability and centrosome aberrations in the progeny of finite life-span human mammary epithelial cells exposed to sparsely or densely ionizing radiation. Radiat Res 2008; 170:23-32; PMID:18582160; http://dx.doi.org/10.1667/RR1317.1.
24. Qin XY, Zaha H, Nagano R, Yoshinaga J, Yonemoto J, Sone H. Xenoestrogens downregulate aryl-hydrocarbon receptor nuclear translocator 2 mRNA expression in human breast cancer cells via an estrogen receptor alpha-dependent mechanism. Toxicol Lett 2011; 206:152-7; PMID:21771643; http://dx.doi.org/10.1016/j.toxlet.2011.07.007.
25. Wetherill YB, Akingbemi BT, Kanno J, McLachlan JA, Nadal A, Sonnenschein C, et al. In vitro molecular mechanisms of bisphenol A action. Reprod Toxicol 2007; 24:178-98; PMID:17628395; http://dx.doi.org/10.1016/j.reprotox.2007.05.010.
26. Nijjar T, Bassett E, Garbe J, Takenaka Y, Stampfer MR, Gilley D, et al. Accumulation and altered localization of telomere-associated protein TRF2 in immortally transformed and tumor-derived human breast cells. Oncogene 2005; 24:3369-76; PMID:15735711; http://dx.doi.org/10.1038/sj.onc.1208482.
27. Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, et al. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. Mol Oncol 2007; 1:84-96; PMID:18516279; http://dx.doi.org/10.1016/j.molonc.2007.02.004.
28. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet 2002; 3:415-28; PMID:12042769.
29. Fleming JM, Miller TC, Meyer MJ, Ginsburg E, Vonderhaar BK. Local regulation of human breast xenograft models. J Cell Physiol 2010; 224:795-806; PMID:20578247; http://dx.doi.org/10.1002/jcp.22190.
30. Bruder ED, Lee JJ, Widmaier EP, Raff H. Microarray and real-time PCR analysis of adrenal gland gene expression in the 7-day-old rat: effects of hypoxia from birth. Physiol Genomics 2007; 29:193-200; PMID:17213367; http://dx.doi.org/10.1152/physiolgenomics.00245.2006.
31. Evron E, Dooley WC, Umbricht CB, Rosenthal D, Sacchi N, Gabrielson E, et al. Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. Lancet 2001; 357:1335-6; PMID:11343741; http://dx.doi.org/10.1016/S0140-6736(00)04501-3.
32. Fackler MJ, McVeigh M, Evron E, Garrett E, Mehrotra J, Polyak K, et al. DNA methylation of RASSF1A, HIN-1, RAR-beta, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma. Int J Cancer 2003; 107:970-5; PMID:14601057; http://dx.doi.org/10.1002/ijc.11508.
33. Lee JS, Fackler MJ, Teo WW, Lee JH, Choi C, Park MH, et al. Quantitative promoter hypermethylation profiles of ductal carcinoma in situ in North American and Korean women: Potential applications for diagnosis. Cancer Biol Ther 2008; 7:1398-406; PMID:18769130; http://dx.doi.org/10.4161/cbt.7.9.6425.
34. Romanov SR, Kozakiewicz BK, Holst CR, Stampfer MR, Haupt LM, Tlsty TD. Normal human mammary epithelial cells spontaneously escape senescence and acquire genomic changes. Nature 2001; 409:633-7; PMID:11214324; http://dx.doi.org/10.1038/35054579.
35. Sarkar P, Shiizaki K, Yonemoto J, Sone H. Activation of telomerase in BeWo cells by estrogen and 2,3,7,8-tetrachlorodibenzo-p-dioxin in co-operation with c-Myc. Int J Oncol 2006; 28:43-51; PMID:16327978.
36. Bandyopadhyay D, Curry JL, Lin Q, Richards HW, Chen D, Hornsby PJ, et al. Dynamic assembly of chromatin complexes during cellular senescence: implications for the growth arrest of human melanocytic nevi. Aging Cell 2007; 6:577-91; PMID:17578512; http://dx.doi.org/10.1111/j.1474-9726.2007.00308.x.
37. Narita M, Nunez S, Heard E, Narita M, Lin AW, Hearn SA, et al. Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. Cell 2003; 113:703-16; PMID:12809602; http://dx.doi.org/10.1016/S0092-8674(03)00401-X.
38. Malyavantham KS, Bhattacharya S, Barbeitos M, Mukherjee L, Xu J, Fackelmayer FO, et al. Identifying functional neighborhoods within the cell nucleus: proximity analysis of early S-phase replicating chromatin domains to sites of transcription, RNA polymerase II, HP1gamma, matrin 3 and SAF-A. J Cell Biochem 2008; 105:391-403; PMID:18618731; http://dx.doi.org/10.1002/jcb.21834.
39. Hayakawa T, Haraguchi T, Masumoto H, Hiraoka Y. Cell cycle behavior of human HP1 subtypes: distinct molecular domains of HP1 are required for their centromeric localization during interphase and metaphase. J Cell Sci 2003; 116:3327-38; PMID:12840071; http://dx.doi.org/10.1242/jcs.00635.
40. Minc E, Allory Y, Worman HJ, Courvalin JC, Buendia B. Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. Chromosoma 1999; 108:220-34; PMID:10460410; http://dx.doi.org/10.1007/s004120050372.
41. Vermeulen K, Van Bockstaele DR, Berneman ZN. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. Cell Prolif 2003; 36:131-49; PMID:12814430; http://dx.doi.org/10.1046/j.1365-2184.2003.00266.x.
42. Sherr CJ. Cancer cell cycles. Science 1996; 274:1672-7; PMID:8939849; http://dx.doi.org/10.1126/science.274.5293.1672.
43. Roussel-Gervais A, Bilodeau S, Vallette S, Berthelet F, Lacroix A, Figarella-Branger D, et al. Cooperation between cyclin E and p27(Kip1) in pituitary tumorigenesis. Mol Endocrinol 2010; 24:1835-45; PMID:20660298; http://dx.doi.org/10.1210/me.2010-0091.
44. Minella AC, Clurman BE. Mechanisms of tumor suppression by the SCF(Fbw7). Cell Cycle 2005; 4:1356-9; PMID:16131838; http://dx.doi.org/10.4161/cc.4.10.2058.
45. Kitagawa K, Kotake Y, Kitagawa M. Ubiquitin-mediated control of oncogene and tumor suppressor gene products. Cancer Sci 2009; 100:1374-81; PMID:19459846; http://dx.doi.org/10.1111/j.1349-7006.2009.01196.x.
46. Hao B, Oehlmann S, Sowa ME, Harper JW, Pavletich NP. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. Mol Cell 2007; 26:131-43; PMID:17434132; http://dx.doi.org/10.1016/j.molcel.2007.02.022.
47. Koepp DM, Schaefer LK, Ye X, Keyomarsi K, Chu C, Harper JW, et al. Phosphorylation-dependent ubiquitination of cyclin E by the SCF(Fbw7) ubiquitin ligase. Science 2001; 294:173-7; PMID:11533444; http://dx.doi.org/10.1126/science.1065203.
48. Minella AC, Loeb KR, Knecht A, Welcker M, Varnum-Finney BJ, Bernstein ID, et al. Cyclin E phosphorylation regulates cell proliferation in hematopoietic and epithelial lineages in vivo. Genes Dev 2008; 22:1677-89; PMID:18559482; http://dx.doi.org/10.1101/gad.1650208.
49. Jenkins S, Wang J, Eltoum I, Desmond R, Lamartiniere CA. Chronic oral exposure to bisphenol a results in a non-monotonic dose response in mammary carcinogenesis and metastasis in mmtv-erbb2 mice. Environ Health Perspect 2011; 119:1604-9; PMID:21988766; http://dx.doi.org/10.1289/ehp.1103850.
50. Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, et al. The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature 2009; 460:1136-9; PMID:19668188; http://dx.doi.org/10.1038/nature08290.
51. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 2006; 38:1378-85; PMID:17072317; http://dx.doi.org/10.1038/ng1909.
52. Radpour R, Kohler C, Haghighi MM, Fan AX, Holzgreve W, Zhong XY. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 2009; 28:2969-78; PMID:19503099; http://dx.doi.org/10.1038/onc.2009.149.
53. Dejeux E, Ronneberg JA, Solvang H, Bukholm I, Geisler S, Aas T, et al. DNA methylation profiling in doxorubicin treated primary locally advanced breast tumours identifies novel genes associated with survival and treatment response. Mol Cancer 2010; 9:68; PMID:20338046; http://dx.doi.org/10.1186/1476-4598-9-68.
54. Pal R, Srivastava N, Chopra R, Gochhait S, Gupta P, Prakash N, et al. Investigation of DNA damage response and apoptotic gene methylation pattern in sporadic breast tumors using high throughput quantitative DNA methylation analysis technology. Mol Cancer 2010; 9:303; PMID:21092294; http://dx.doi.org/10.1186/1476-4598-9-303.
55. Moelans CB, Verschuur-Maes AH, van Diest PJ. Frequent promoter hypermethylation of BRCA2, BRCA1, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer. J Pathol 2011; 225:222-31; PMID:21710692; http://dx.doi.org/10.1002/path.2930.
56. Shim YH, Kang GH, Ro JY. Correlation of p16 hypermethylation with p16 protein loss in sporadic gastric carcinomas. Lab Invest 2000; 80:689-95; PMID:10830779; http://dx.doi.org/10.1038/labinvest.3780072.
57. Taghavi N, Biramijamal F, Sotoudeh M, Khademi H, Malekzadeh R, Moaven O, et al. $p16^{INK4a}$ hypermethylation and p53, p16 and MDM2 protein expression in esophageal squamous cell carcinoma. BMC Cancer 2010; 10:138; PMID:20388212; http://dx.doi.org/10.1186/1471-2407-10-138.

58. Zang JJ, Xie F, Xu JF, Qin YY, Shen RX, Yang JM, et al. P16 gene hypermethylation and hepatocellular carcinoma: a systematic review and meta-analysis. World J Gastroenterol 2011; 17:3043-8; PMID:21799651; http://dx.doi.org/10.3748/wjg.v17.i25.3043.

59. Salvesen HB, Das S, Akslen LA. Loss of nuclear p16 protein expression is not associated with promoter methylation but defines a subgroup of aggressive endometrial carcinomas with poor prognosis. Clin Cancer Res 2000; 6:153-9; PMID:10656444.

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

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

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

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

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

I. INTRODUCTION

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

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

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

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

II. MATERIAL AND METHODS

A. Expression data

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

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

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

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

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

B. *Extraction of causalities from expression data*

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

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

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

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

C. *Construction of the initial model*

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

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

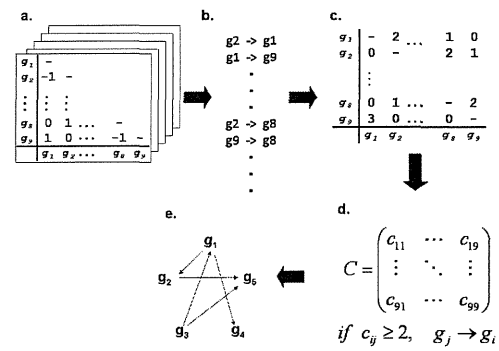


Figure 1. Developed procedure for initial model construction.

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

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

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

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

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

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

E. Parameter Estimation

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

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

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

F. Iteration for Optimal Model

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

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

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

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

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

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

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

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

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

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

III. RESULTS AND DISCUSSION

A. The chemical concentrations have no effect

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

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

TABLE I. NUMBER OF GENES WITH SINIFICANTLY DIFFERENT EXPRESSION

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

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

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

B. Genes are hierarchically controlled by chemical toxicity

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

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

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

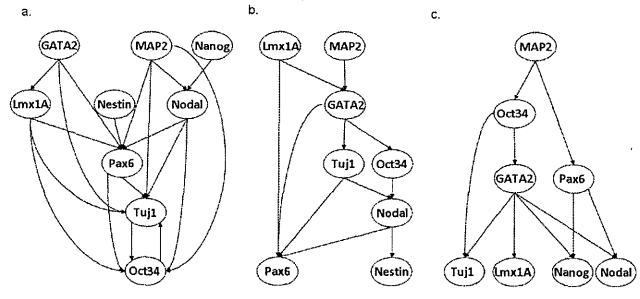


Figure 2. Initial network models.

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

C. *Inferred Network by SEM*

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

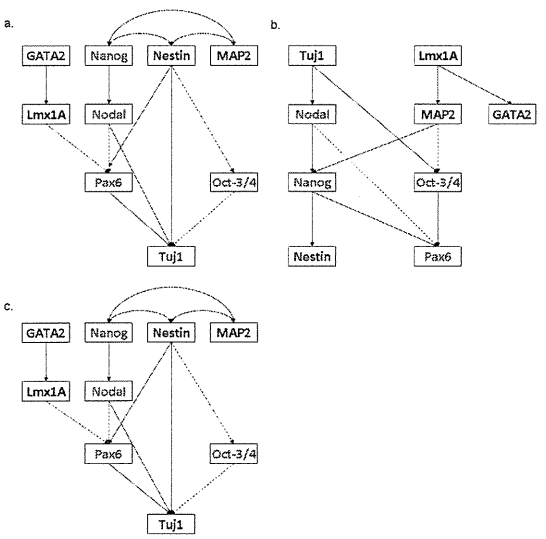
According to our inferred network, the differences between the gene regulation by chemicals were clarified, and the network shapes reflected the features of chemical toxicities. In the inferred network, the effects of Acrylamide toxicity finally aggregated to Tuj1, which is known to contribute to microtubule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

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

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

IV. CONCLUSION

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



d.

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

Figure 3. Inferred network by SEM.

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

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

ACKNOWLEDGMENT

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Journal of Molecular Biology Research; Vol. 2, No. 1; 2012
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

Inference of Specific Gene Regulation by Environmental Chemicals in Human Embryonic Stem Cells

Sachiyo Aburatani[1] & Wataru Fujibuchi[2]

[1] Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

[2] Center for iPS Research and Application, Kyoto University, Kyoto, Japan

Correspondence: Sachiyo Aburatani, Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology, AIST Tokyo Waterfront BIO-IT Research Building, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan. Tel: 81-335-998-712. E-mail: s.aburatani@aist.go.jp

Received: December 2, 2012 Accepted: December 17, 2012 Online Published: December 27, 2012

doi:10.5539/jmbr.v2n1p54 URL: http://dx.doi.org/10.5539/jmbr.v2n1p54

Abstract

We are exposed to many environmental chemicals in our daily life. Certain chemicals threaten our health, especially that of embryos and can cause serious developmental problems. To prevent abnormal development and diseases caused by chemicals, it is important to clarify the mechanisms of chemical toxicity in embryonic cells. The gene regulatory network is one of the useful methods for clarifying functional mechanisms in living cells, so we applied a statistical method to infer the gene regulatory network in human embryonic stem cells. In this study, we improved our previously developed SEM approach for inferring a network model from 9 gene expression profiles in human embryonic stem cells, which were exposed to various chemicals. The estimated regulatory models clarified the differences between chemicals, and the shapes of the inferred models reflected the features of the chemical toxicities. The toxicity of acrylamide affected neuronal cell-related genes, while that of diethylnitrosamine disturbed cell differentiation-related genes. On the other hand, the TCDD network reflected feedback regulation, and finally disturbed neuronal cell-related genes. In the Thalidomide network, cell differentiation genes related to axis formation in embyronic cells were affected by thalidomide toxicity.

Keywords: structural equation modeling, environmental chemical, gene regulatory network, embryonic stem cell

1. Introduction

1.1 Introduction of the Problem

Environmental pollution is a byproduct of our usual life activities. Vehicle exhaust contains gases, including many noxious chemicals. Factories discharge industrial waste in the air, ground, and water. Many rivers are polluted by domestic sewage and wastewater. The emitted chemicals are sometimes trapped in clouds and then contaminate the ground in rainfall. Thus, we are exposed to many chemicals in our daily life, and some environmental chemicals can cause serious developmental toxicity effects. 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 (Baccarelli & Bollati, 2009; Hou et al., 2012). One of the most infamous environmental chemicals is methylmercury, which is known to affect fetal development (Yuan, 2012; Tatsuta et al., 2012). Furthermore, other chemicals are also considered to be toxic, since they can cause abnormal cell differentiation in embryos (Rappolee et al., 2012; He et al., 2012; Harrill et al., 2011).

To prevent chemically-induced developmental abnormalities and diseases, it is important to clarify the mechanisms of chemical toxicity in embryonic cells (Gündel et al., 2007; Thompson & Bannigan, 2008). The gene regulatory network is one of the useful methods to clarify the regulatory mechanisms. To infer the networks among the genes from the mRNA levels, various algorithms, including Boolean and Bayesian networks, have been developed (Akutsu et al., 2000; Friedman et al., 2000). In our previous investigation, we developed an approach based on graphical Gaussian modeling (GGM) in combination with hierarchical clustering, and we could infer the huge network among all of the genes by this approach. (Aburatani et al., 2003; Aburatani &

Horimoto, 2005). 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, because of insufficient information in the gene expression profiles. To clarify the mechanisms of biological processes in living cells, un-observed factors, which 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), to infer the protein-DNA interactions for gene transcriptional control from only the gene expression profiles, in the absence of protein information (Aburatani, 2011; 2012). We applied this approach to reveal the causalities within the well-studied transcriptional regulation system in yeast (Aburatani, 2011). The significant features of SEM are the inclusion of latent variables within the constructed model and the ability to infer the network, including the cycle structure. Furthermore, the SEM approach allows us to strictly evaluate the inferred model, by using fitting scores. The linear relationships between variables are assumed to minimize the differences between the model's covariance matrix and the calculated sample covariance matrix. Some fitting indices are defined for evaluating the model adaptability, and thus the most suitable model can be selected by SEM (Bollen, 1989; Duncan, 1975; Pearl, 2001).

Here, we applied the SEM approach to infer the regulatory relationships among 9 neurodevelopmentally-related genes. The expression profiles of these genes were measured in human embryonic stem cells exposed to four environmental chemicals. The chemicals are known to have harmful toxicities that affect the developmental process in human embryos. Thus, inferring the regulatory network among the developmentally-related genes will help us to reveal the mechanisms of toxicity-dependent responses in the embryo. Furthermore, we improved our SEM approach for assuming preliminary initial models from the time-series data. By using this new approach, we can construct an initial model for the SEM calculation in the absence of known regulatory interactions. The resulting gene expression data clarified the chemical-specific interactions among the developmentally-related genes.

2. Methods

2.1 Expression Data

We utilized the expression data that were measured to clarify the effects of environmental chemical exposure on neuronal differentiation (He et al., 2012; Fujibuchi et al., 2011). In these expression data, nine genes considered to be affected by chemicals were measured in human embryonic stem cells: GATA2, Lmx1A, MAP2, Nanog, Nestin, Nodal, Oct3/4, Pax6 and Tuj1 (He et al., 2012; Fujibuchi et al., 2011). The expression of beta-actin was also measured, as an internal control. The expression levels of these 10 genes were measured in human embryonic stem cells exposed to four chemicals: acrylamide, diethylnitrosamine, TCDD and thalidomide (He et al., 2012; Fujibuchi et al., 2011). The toxicities of these chemicals are different: acrylamide is neurotoxic, diethylnitrosamine is genotoxic, TCDD is carcinogenic, and thalidomide has other toxicity. 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, medium, high and very high. The expression of the selected genes was measured twice under each condition by RT-PCR, and thus 160 (4 time periods x 5 concentrations x 2 repeats x 4 chemicals) expression patterns per gene were measured (Fujibuchi et al., 2011).

First, the expression level of each gene was normalized to the internal beta-actin control and averaged, as follows:

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

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. By dividing by the expression level of beta-actin, the intracellular expression level of each gene was normalized. To minimize the experimental error, the logarithms of the normalized expression data were obtained and averaged.

2.2 Extraction of Causalities from Expression Data

Usually, we assume an initial model from previous knowledge for the SEM calculation, but there are no defined regulations among the selected genes in this study. Thus, we had to construct an initial model of each chemical from the regulatory relationships between the gene pairs. To detect the regulatory relationships from the measured time series expression data, cross correlation coefficients were applied to the expression profiles. These cross correlation coefficients were calculated 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 (Li & Caldwell, 1999). In a time series analysis, the cross correlation between two time series describes the normalized cross covariance function. Therefore, the range of cross correlation values is from -1 to +1. If we let $X_t = \{x_1, \cdots, x_N\}, Y_t = \{y_1, \cdots, y_N\}$ represent two time series datasets including N time points, then the cross correlation is given by

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

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, and 1.

2.3 Construction of the Initial Models

To infer the chemical-dependent regulatory networks, the differences between times and concentrations should be merged. In this study, we developed a new method for constructing an initial model of each chemical, with the merging of time and concentration conditions. Figure 1 shows the newly developed method. First, we constructed lag matrices to simplify the information from the time series data. The elements of the lag matrices were the time lags, which were defined for the calculation of the cross correlation. In this study, cross correlations were calculated with three lags, -1, 0, and +1. The absolute values of these three cross correlations were compared, and the lag value d with the highest absolute value was arranged as a matrix element. Lag matrices were constructed for each concentration, and thus five lag matrices were obtained for each chemical (Figure 1a).

In the next step, we merged the difference in the concentrations of each chemical. Binomial relationships were extracted from each lag matrix. For each chemical, there are five lag matrices according to the chemical concentration, and we considered that the chemical-specific relationships among the genes will be conserved in several lag matrices. If the same relationships existed in several lag matrices, then the binomial relationships were duplicated (Figure 1b).

We subsequently constructed one frequency matrix of binary relationships for each chemical. We counted the frequency of the appearance of relationships in binomial relationships. The number representing the frequency of each gene pair was arranged in this matrix, and thus the range was from 0 to 5 (Figure 1c). In the frequency matrix, we can merge the differences in the concentrations, since the elements of the frequency matrix indicate the information for the different concentrations. We selected the possible relationships from the frequency matrix. It is considered that a possible relationship would be indicated by its frequency of appearance. Thus, we selected the relationships with two or more values in the frequency matrix (Figure 1d). At the final step, an initial model was constructed with the selected possible relationships. By this approach, an initial model can include cyclic structures (Figure 1e).

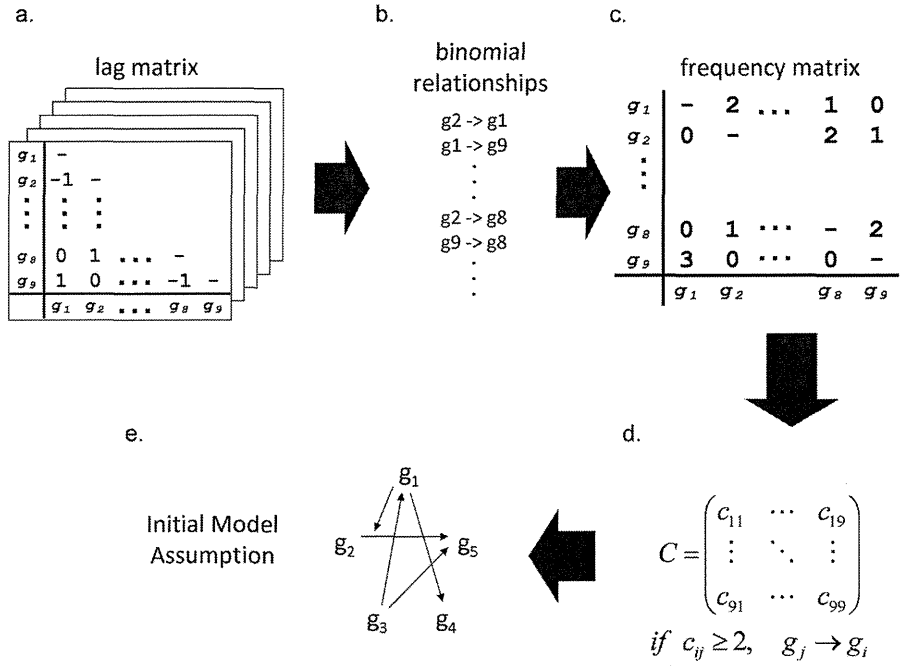


Figure 1. Developed procedure for initial model construction

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

2.4 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. Usually, two types of variables can be included in the SEM model: 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 nine developmentally-related genes (GATA2, Lmx1A, MAP2, Nanog, Nestin, Nodal, Oct3/4, Pax6 and Tuj1) were defined as the observed variables. Meanwhile, none were defined as latent variables, which were common regulators of several genes. The un-observed factor, which affected each gene's expression, was displayed as an error. The observed variables were classified as one of two types: exogenous variables and endogenous variables. Exogenous variables are not regulated by other variables in the system, as opposed to endogenous variables, which are regulated by other variables in the system. In the initial model, the starting genes are defined as exogenous variables without errors, while all other genes are defined as endogenous variables with errors. We inferred the regulatory relationships that exist between the observed variables in the network model. The model is defined as follows:

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

Here, y is a vector of p observed variables (measured gene expression patterns), and Λ is a $p \times p$ matrix representing the regulatory relationships between the observed variables. Errors that affect the observed endogenous variables are denoted by ε. The above equation can be represented in the SEM matrix format as:

$$\begin{bmatrix} O \\ y \end{bmatrix} = \begin{bmatrix} O & O \\ O & \Pi \end{bmatrix} \begin{bmatrix} O \\ y \end{bmatrix} + \begin{bmatrix} O \\ \varepsilon \end{bmatrix} \tag{4}$$

In this study, we did not define the latent variables, and thus Os were arranged as zero partial matrices, which denote no relationships with q latent variables. The SEM is based on a covariance analysis defined as $S = \Sigma(\theta)$, where S is the covariance matrix calculated from the observed data and $\Sigma(\theta)$ is the matrix-valued function of the parameter θ. Let Φ denote the covariance matrix of the error terms ε, and G denote the $p \times (p+q)$ combined matrix of the $p \times q$ zero matrix and the $p \times p$ identity matrix. The covariance matrix of model is given by

$$\Sigma(\theta) = G \begin{bmatrix} I-O & O \\ O & I-\Pi \end{bmatrix}^{-1} \Phi \begin{bmatrix} I-O & O \\ O & I-\Pi \end{bmatrix}^{-1'} G' \tag{5}$$

Each element of the covariance matrix model is expressed as a function of the parameters that appear in the model. The unknown parameters were estimated, in order to minimize the difference between the model covariance matrix and the sample covariance.

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 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 are considered to be useful to clarify the degree of model fitting in this study.

2.5 Parameter Estimation

Parameter estimation was performed by comparing the actual covariance matrix S, calculated from the measured data, and the estimated covariance matrices $\Sigma(\theta)$ of the constructed model. To minimize the difference between S and $\Sigma(\theta)$, the Maximum Likelihood (ML) method is commonly used as a fitting function to estimate the SEM parameters:

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

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 (Joreskog & Sorbom, 1984).

2.6 Iteration for Optimal Model

In the SEM analysis, both the parameters and network structures are fitted to the measured data. The parameters are estimated by maximum likelihood, and the network structures are evaluated by the scores of goodness of fit indices. The goodness of fit scores indicate the similarity between the constructed model and the measured data. Through the acceptance or rejection of the models, the optimal model that describes the measured data can be selected.

By using the estimated parameters, the variance-covariance matrix between the variables could be calculated in the network model. This model's variance-covariance matrix is compared with the actual variance-covariance matrix between observed variables, which is calculated from the measured data. The similarity between a constructed model and the actual data is defined in a quantitative manner by the fitting scores. In this study, four different fitting scores were utilized: GFI, AGFI, CFI and RMSEA. Values of GFI, AGFI and CFI above 0.90 are required for a good model fit. RMSEA is one of the most popular parsimony indexes displayed in the table, and RMSEA values below 0.05 represent a good model fit (Spirtes et al., 2001). 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, we developed an iteration algorithm as follows: