to other morphological parameters or genes in the PMT-exposed GPIN. The neurite morphological parameters (Neurite_length, Branch_point, Crossing_point and Posi_area) influenced expression of genes in contrast to the NS morphological parameters. Considering the significant increase of total length of Map2-positive neuron (Figure 2A) and no change in the NS morphological parameters by PMT, the PMT-exposed GPIN successfully drew the change of neuronal morphology.

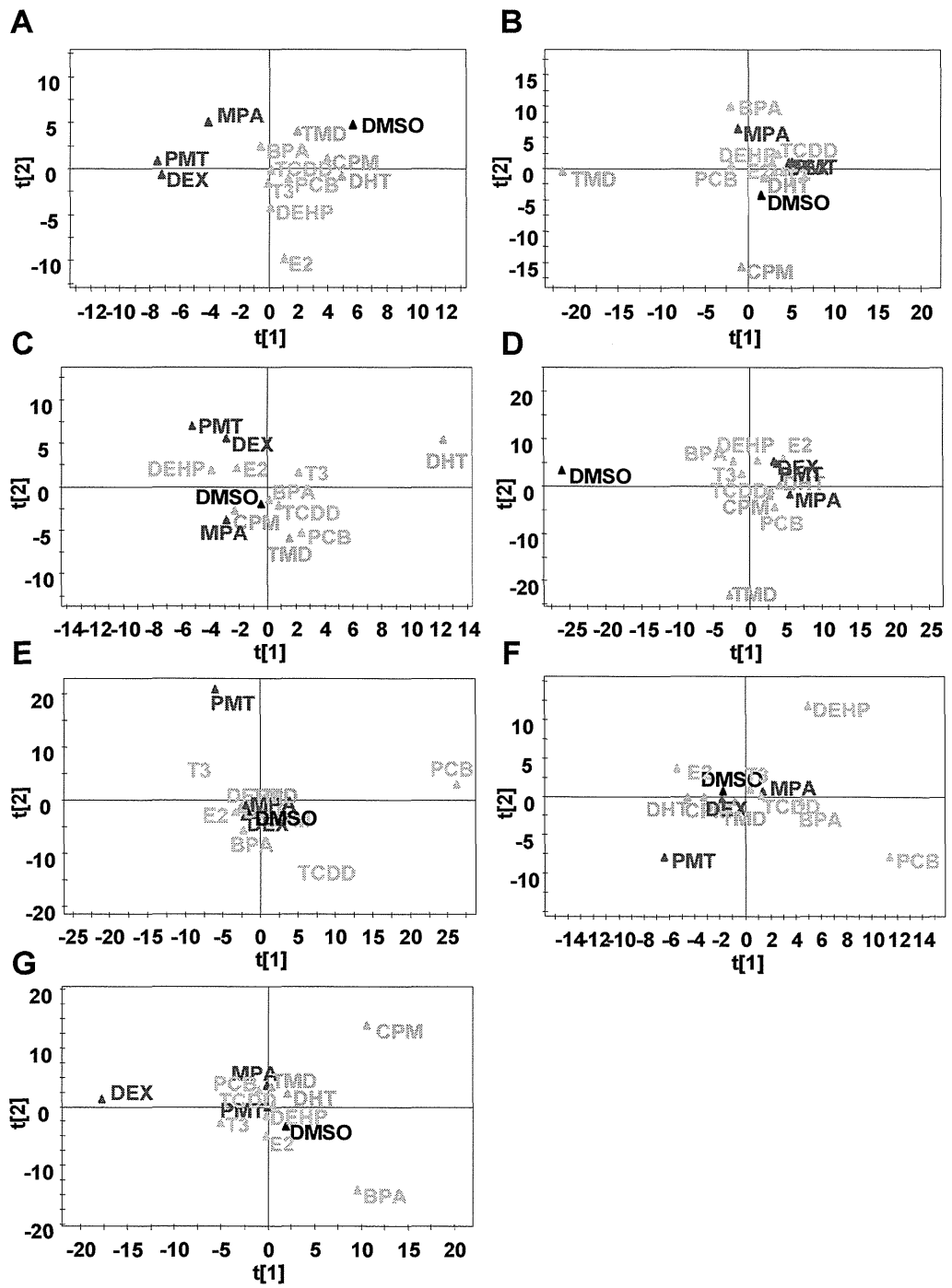
The comparison of TMD-exposed GPIN or PMT-exposed GPIN with DMSO control GPIN for Autism set and Parkinson's disease set could be understood without contradicting known pathological pathways. Therefore, we propose that our MPNs approach could draw out the risk of chemicals. The gene expression profiling data of our study have been published on the Profiles of Chemical Effects on Cells (pCEC) system [27], which is a toxicogenomics database with a toxicoinformatics system for risk evaluation and toxicity prediction of environmental chemicals [28] and produced by the National Institute of Environmental Studies, Japan. The microarray data have also been released on the GEO data base [29].

2.3. Classification of Chemicals Based on the Values of the Parameters of the Comprehensive Networks

The genomic data and cell morphological data were converted to the same matrix vector and were used to analyze GPIN. Principal component analysis (PCA) based on the probabilistic relationship data of the GPIN showed that all variance between the 12 chemicals could be described using the first and second principal components (PCs) (Figure 5). The two dimensional PCA plot showed four different groups: DMSO control (black), TMD group (CPM and DHT, green), BPA group [2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), PCB, T3, bis(2-ethylhexyl) phthalate (DEHP) and E2, blue] and MPA group (PMT and DEX, red) were derived for the Alzheimer's disease related gene set. The same color coding was used for other experiments, which enabled us to visually recognize changes to the grouping of chemicals. When the largest variable variation was placed in the vertical axis (PC1) and the second variation in the horizontal axis (PC2), the two-dimensional plot showed the position of each chemical. PMT and DEX were located near, but separated from, DMSO in Alzheimer set and Parkinson set. The toxic effects of DEX were reported in animal model of Alzheimer' disease [30] and Parkinson' disease [31] although we found no report about PMT in Alzheimer's disease. In Alzheimer set, E2 was located further away from DMSO than DHT and the opposite positioning was detected in Parkinson set. It might reflect the sexual differences of the diseases as the risk of Alzheimer's disease is higher in females [32] and that of Parkinson's disease is higher in males [33]. Because the responsible genes of gender specific Autism spectrum were involved in the Autism set, such gender dependent differences might not be detected in present data. In Autism set, TMD was more separate from DMSO than the others. Indeed other than TMD, the chemicals show no evidence of involvement in autism at present. In the Axon guidance set, all chemicals were almost equally distant from DMSO. As shown in (Figure 2A), all chemicals influenced the total length of Map2-positive neuron at high dose. Therefore, this result is reasonable. In the pluripotent set, PMT and PCB were separated from the others indicating that these chemicals affected the differentiation from ES cells. In fact, PMT and PCB are also located away from DMSO in neural development set. The characterization of chemicals according to their neurotoxic

potential reveals that the method described in this current study—that the MPN analysis based on phenotype and gene expression profiling during neuronal differentiation of mESCs—can provide a useful tool to monitor fetal programming and to predict developmentally neurotoxic compounds.

Figure 5. PCA based on Bayesian network parameters. PCA were applied to the Bayesian network parameters based on phenotypic and global gene expression profiling to evaluate the neurotoxicity of 12 environmental chemicals. Score plots based on (**A**) Alzheimer's disease related gene set; (**B**) Autism related gene set; (**C**) Parkinson's disease related gene set; (**D**) Axon guidance related gene set; (**E**) Pluripotent related gene set; (**F**) Neural development related gene set; and (**G**) Oxidative stress related gene set.



2.4. Discussion for Future Work

ESCs test combined with transcriptomics for the assessment of development toxicity has been well studied in recent years [8,34]. However, studies based on the genotype-phenotype profiling are rare. Cell phenotypes are complex and difficult to quantify in a high throughput fashion. The lack of comprehensive phenotype data can prevent or distort genotype-phenotype profiling. Our study described a unique approach to perform multiple phenotype profiling using gene expression data from the early stage of mESC differentiation and morphological data of neuronal cell differentiation after EB formation. Our method provided numerous advantages: (i) Our method can predict multiple phenotype profiles, which could help researchers to reveal different aspects of complex diseases and facilitate treatment design; (ii) Our method can provide a quantitative phenotype description of the sample characteristics; (iii) Our method can extrapolate the profiling to classes beyond those represented in the training data. This is an advantage over traditional classification methods. In contrast, traditional regression methods cannot be directly applied to microarray datasets from different platforms and cannot predict relationships between early events and late phenomena during the differentiation of ES cells into neuronal cells. However, our method can be applied to other types of genomics data such as proteomics or metabolomics. The present study focuses on linear gene-phenotype associations, but more complex relationships can also be devised depending on the data characteristics. Our multi-parametric profiling method for constructing interfering networks of the gene expression data and cellular phenotypic data is only one of many possible approaches. As mentioned above, our MPN analysis can predict the correlation coefficient for each pair of nodes, regardless of the data types. Therefore, our informatics approach and experimental design is also an efficient tool for data integration, mining and network analysis for the other model systems. However, another important issue for the future will be the validation of a larger set of chemicals at a broad concentration range to identify the specific and mechanistically defined markers for differential environmental chemicals.

ES cell-based assays are a promising platform to assess developmental toxicity, because they are capable of recapitulating many of the differentiation states and rely on signaling pathways present in development. We used a neuronal differentiation assay of mESC to assess the activity of groups of environmental chemicals, most of which have *in vivo* toxicity data. The results of this study demonstrated that a subset of tested chemicals are effective in this assay, and that as a statistical analysis, BNA, identified predictive models of detecting fetal programming in the mESC differentiation for a subset of the tested chemicals. Chandler *et al.* demonstrated evaluation of environmental chemicals using a mESC adherent cell differentiation and cytotoxicity assay, showing that genes involved in reactive oxygen species signaling pathways were strongly associated with decreased ES cell differentiation [35]. However, their approaches are linear regression or categorical approaches and are not identical with our approaches. Our approach is unique in linking early gene expression events to the later cellular phenotype features by BNA, which has become popular among biological scientists [36]. Many studies using BNA focus on basic physiological and developmental phenomenon based on cell proliferation [37]. In contrast, our study targets effects of early exposure on late-onset phenotypes, in accordance with the principles of fetal programming against environmental chemicals. In this regard,

this is the first study to combine gene expression data and morphological data to estimate the mechanistic path of the response during the early embryonic period.

3. Experimental Section

3.1. Selection of Test Chemicals

Twelve chemicals, mostly well-characterized medical drugs, pesticides or plastic materials, which have been previously tested by traditional *in vivo* toxicology methods, were used in this study (Table 1). T3, DEX, E2, DHT and MPA are the agonists of the nuclear receptors, ThRs, GR, ERs, AR and RXRs respectively and regulate expression of target genes of each receptor. TCDD also is the agonist of a transcription factor termed AhR [38]. Therefore, these chemicals influence differentiation and development many tissues including neural tissues. CPM, a well characterized teratogen, is the inhibitor of sonic hedgehog (Shh) signal [39]. It can inhibit the acquisition of ventral identity in mESCs-derived neural stem cells [40]. TMD is also well known teratogen of human but not rodents although the toxicological mechanism remains to be unclear [41]. Human epidemiological studies suggested the involvement of TMD in the appearance of autism [42,43]. The studies using rats showed that prenatal exposure to TMD could cause autism-like symptoms in rodents [44]. Prenatal or postnatal exposure to PCB showed long term effects on brain development and behavior in rat [45]. PMT, BPA and DEHP have also shown neurotoxicity in animal models [46–48]. Recently, the TestSmart DNT II meeting to discuss about development of alternative testing methods and models for DNT showed a list of the candidate chemicals for positive control in DNT [5]. 4 chemicals of our list, TMD, PCB, PMT and DEHP are involved in the list. Therefore, the choice of chemicals in present study can be adequate.

3.2. Design of Multi-Parametric Profiling Networks Analysis for Detecting Developmental Neuronal Toxicity of Chemicals That Effects Fetal Programming

To evaluate developmental neurotoxicity of these chemicals, we designed a MPN analysis based on gene expression and cellular phenotypic data. The process of MPN analysis was composed of 5 steps (Figure 1). Step 1 involves the exposure of mESCs to chemicals and then the differentiation of mESCs into neuronal cells. Cells were exposed to chemicals for 2 days during Day 0 to Day 2 when initial EBs were formed. Gene expression determination using microarray analysis was performed on RNAs that were sampled immediately after cells were exposed to chemicals. EBs of Day 8 were transferred to poly-DL-ornithine/laminine-coated 24 wells plate and cultured until Day 20 when cells had adequately differentiated to neuronal phenotypes. Differentiated neuronal cells were visualized by immunofluorescence staining. Cell images were acquired automatically using a 10× objective. Gene expression sets selected from microarray data and morphological data of neuronal cells were collected into the same matrix (Step 2). Seven gene expression signatures (pluripotent, neural development, axon guidance, autism, Parkinson's disease, Alzheimer's disease and oxidative stress) of biological events and neuronal disease were selected manually and are shown in Table 2. The genes in autism set were chosen based on some reviews [49–51]. The gene in pluripotent set were chosen based on Wang *et al.* [52] and Müller *et al.* [53] The KEGG pathway database was referred to choose genes in other sets. Sex steroid receptors (ESR1,

ESR2 and AR) and retinoic acid receptors (RARα, RARβ and RARγ) were added into the autism set, Parkinson set, Alzheimer set to consider the gender depending differences and to consider the effects of neuronal induction by RA *in vitro*, respectively. Once transition matrices were made from gene expression and neuronal cell phenotypes, phenotypic networks and MPNs were derived by BNA. Namely, nodes in the generated GPIN included each of the genes contained in the gene list or each of the morphologic parameters, such as neural cell count or neurite length (Step 3 and 4). We then applied PCA to classify the generated MPN for 7 gene-signature sets of each test-chemical. The values of linkage probability between two nodes in the MPN were used as the parameters in PCA (Step5).

3.3. mESC Culture and Maintenance

mESC (B6G-2) derived from Green mouse FM131, a mouse constantly expressing GFP, were cultured on deactivated mouse fibroblast cells (RIKEN, Japan). The proliferated cells were replated on 0.1% gelatin coated dishes with DMEM (phenol red free, Invitrogen, Carlsbad, CA, USA) containing 15% FBS (fetal bovine serum, Invitrogen), 100 μM NEAA (Non-essential amino acids, Invitrogen), 100 μM 2-ME (2-mercaptoethanol, Invitrogen) and 1000 U/mL LIF (Leukemia inhibitory factor, ESGRO, Invitrogen).

3.4. EB Formation from mESC and Chemical Treatment

The microsphere array used in this study is a frame separated type (chip 300, STEM Biomethod Corporation, Kitakyushu, Japan), which is made of acrylic resin and the surface has been coated with PDMS resin that is not structured for direct cell adhesion. 1024 wells (diameter 300 μm) were arranged on the surface of the microsphere array. EB formation was performed in the three dimension culture based on the microsphere array. After removal of mouse fibroblast cells, aggregated ES cells were counted and 250 μL cell suspension solution (2×10^5 cells) were put on the microsphere array. Six hours later, the medium was exchanged for each chemical containing medium and incubation continued for 48 h. After that, RNA was isolated for gene expression analysis and culture medium was exchanged for EB medium with add 10 nM retinoic acid for the further morphological analysis. EBs were cultured for 6 days with EB medium replaced every two days. Eight days after chemical exposure, aggregated EBs were replated on Ornithine/Laminine coated 24 wells plate (83 EBs/well). Twenty-four hours later, EB medium was exchanged for neural differentiation medium (DMEM/F12 (1:1), N2 (×100), and 10 ng/mL bFGF) and EBs were cultured for another 20 days, exchanging the medium every 3 days. DMSO was used as the primary solvent for all chemicals, and the DMSO solutions were further diluted in cell culture media for treatments. The final concentrations of DMSO in the media did not exceed 0.1% (vol/vol). The concentrations of chemicals used in this study were: 1 pM and 100 pM for BPA; 1 nM and 10 nM for T3, DEX, E2, DHT, PCB and TCDD; 0.1 μM and 10 μM for CPM, PMT and TMD; 1 μM and 100 μM for MPA and DEHP. The neuronal differentiation parallel to development *in vivo* was confirmed by quantitative RT-PCR of stage specific markers, Oct3, Nanog, Pax6 and Map2 (data not shown).

3.5. Immunofluorescence

On Day 20, EBs and differentiated cells were immunostained with Mouse anti-MAP2 antibody (1:200 dilution; Sigma-Aldrich, St. Louis, MO, USA), Mouse anti-GFAP monoclonal antibody (1:200 dilution; Chemicon, GA, USA) and Hoechst 33342 solution (Dojindo, Tokyo, Japan). In brief, cells were fixed with 4% PFA in PBS for 15 min and then blocked for 30 min in PBT buffer (PBS with 5% Goat serum and 0.1% Triton). Cells with primary antibodies were incubated overnight at 4 °C. Cells were washed and blocked in BBT-BSA and then incubated with Alexa-conjugated secondary antibodies (1:1000 dilution, Alexa Fluor 546, Invitrogen). Hoechst 33342 staining was used for counter staining.

3.6. Morphological Analysis of mESC, EB and Neuronal Cell Lineages

The immunofluorescence images were acquired using the IN Cell Analyzer 1000 (GE Healthcare, Buckinghamshire, UK). Each neural cell image was analyzed using image analysis software IN Cell Developer Tool Box 1.7 (GE Healthcare). The following 10 parameters were measured: number of all cells (Nuc_count), nucleus area (Nuc_area), the number, area, perimeter and formation of neurospheres (NS), (NS_count, NS_area, NS_perimeter and NS_formfactor), and the shape of nerve cells and the size of neural marker positive cells (posi_area, Neurite_length, Branch_point and Crossing_point).

3.7. Gene Expression Analysis and Creation of Candidate Gene Sets

Total RNA on Day 2 of cells derived from mESCs were applied to Illumina beads array systems with the Illumina Mouse WG-6 v1.0 expression beadchip (Illumina, San Diego, CA, USA). The amounts, purity and integrity of RNA were evaluated by UV spectrophotometry and an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). Genes were normalized with analytical software GeneSpring GX10.02 (Agilent Technologies) [54]. 7 sets of genes were created with reference to the literature to assess the impact on neural development. These categories were Pluripotent, Neural development, Axon guidance, Autism, Parkinson's disease, Alzheimer's disease, and Oxidative stress.

3.8. Gene and Morphology Interaction Network Analysis

GPIN was quantified to calculate the posterior probability distribution for the strength of the linkages based on gene expression, morphological and chemical exposure dose datasets. Briefly, a GPIN consists of a collection of P nodes, denoted G_1, G_2, … G_P, with observed values n_1, n_2, …. n_p. Define ij (i,j = 1,2, …, P) as parameters in the log-linear function form describing the linkage from node i to node j. Mathematically, this is written as:

$$E[\log(G_j)] = \sum_{i=1,\neq j}^{P} I_{ij}\beta_{ij}\log(g_i) \quad (1)$$

where E[log(G_j)] represents the expectation for the natural logarithm of G_j and I_{ij} (i,j = 1,2, …, P) is an indicator function that equals 1 if node G_i has a link to node G_j, otherwise it equals 0. If a node has a regulatory effect on node G_i, then that node is referred to as a "Parent of node G_i" and we refer to it as

belonging to the set $Pa(G_i)$. The prior distribution for the variance is assumed to be inverse Gamma and assuming that the natural log of G_j follows a normal distribution with mean and standard deviation, posterior distributions for each parameter can be estimated. The posterior distributions for the linkages were derived using Gibbs sampling. Gibbs sampling has no limitation on the number of possible parents and is easy to cooperate with knowledge information or past experimental results by taking the information into the prior distribution. The goal of the method is to examine the posterior distribution of the linkages between genes. In this study, we applied 20 sets of gene expression data ($N = 30$) and morphological data ($N = 162$). Network was used to evaluate the ability of the algorithm to have higher posterior probability (*P*-value) at the correct linkage in GPIN. In each simulation, Gibbs sampling was performed between 33,000 and 48,000 times. The initial Gibbs sampling was considered to be the burn-in period and was removed in estimating and the last 18,000 to 26,000 iterations were used to establish. *P*-value threshold was set to between 0.995 and 1.0 for up-regulation, 0.47 and 1.0 for down-regulation. Three categories were classified out of the12 GPs depending on network structures.

Class 1: Thick and elongated neurons, but with a small amount of neurite branching. Class 1 could be distinguished from other classes in terms of loading the "Neurite_length" parameter on the top of the PN, such that "Neurite_length" controlled "Branch_point" and "Crossing_point". The node located towards the bottom seems to suppress neurite growth. The neurite becomes a parent node, which dominates all the other parameters in the PN in order to facilitate its own growth. Namely, the branching points and intersections are increased in parallel with neurite elongation. The parameters of "EB_Area", "EB_Perimeter" and "EB_FormFactor" are also related to "Neurite_length", which perhaps suggests that neurites have differentiated normally from EBs and that the shape of NSs is not a circle (*i.e.*, NS becomes flattened during differentiation).

Class 2: Neurite elongation and branching are extensive. In this case, "Branch_point" is located on the top, suggesting that the "Branch_point" controls "Neurite_length" and "Crossing_point". "Neurite_length" is expressed as the total length of all neurites per cell. "Branch_point" becomes the parent node in this PN because there are many random short neurites and the total length of all the branching short neurites at their branch points is regarded as the neurite length. Therefore, the promotion arrow from the branch point tends to be the parameter of neurites. Because there are so many random branch points, it is very likely that there are many short crossing intersections. Furthermore, since there are so many branches from the neurites which perhaps did not differentiate from EBs, the parameter of "Branch_point" might not be related to EB shape. Consequently, the EB shape tends to be round compared with that of Class 1 EBs.

Class 3: larger NSs and less neurites. Different from classes 1 and 2, "Nuc_count" and "Nuc_area" are localized at the top in this PN. This suggests that cell proliferation in NSs is more predominant than neural cell outward migration. Common to these three classes, in case of that differentiated neural cell expanded outside of EB and neural differentiation related morphological parameters emerged above of PN. These parameters exert influence on the number of cells and the shape of the EB. Furthermore, when the differentiation is advanced, the PN tends to become complex. In fact, neural differentiation is not too advanced like as Class 3, it became the result of locating the parameter related to number of cells in the high rank from the parameter of the neuronal cell. The parameter concerning the EB is always

located in the subordinate position of the PN on any PN and this tendency corresponded to the theory that the shape changed depending on the number of cells and the progression of neuronal differentiation.

3.9. Statistical Analysis

All experiments in this study were performed in triplicate to test the reproducibility of the results. Statistical analysis was performed by two-tailed Student's *t*-test. Relationships were considered statistically significant with $p < 0.05$.

4. Conclusions

Our study provides an advanced framework to integrate the gene expression and neuronal cell phenotypes for target prediction. Thus a combination of BNA and PCA clustering could provide compound-target prediction efficiency. We believe this method has considerable potential. For example, new markers could be implemented that enable predictive toxicology of active lead compounds. Combined with chemical structure knowledge and ligand-target prediction, such approaches could provide detailed mechanistic insight to help guide medicinal chemists early in the lead optimization process. Dealing with complexities of predictive toxicology will require breakthroughs in cellular image analysis, target prediction schemes and data mining. Our integration analysis of cellular phenotypes with gene expression represents a step forward in solving the DNT for environmental chemical assessment.

Acknowledgments

This study was supported in part by the Environmental Technology Development Fund (to H.S.) from the Ministry of the Environment and a Grant in Aid for Scientific Research from the Ministry of the Health, Labour and Welfare, Japan (to S.O.). The authors gratefully acknowledge the technical support of Noriko Oshima (GE Healthcare Japan Corporation) for analysis using the IN Cell Analyzer 1000 and Shigeru Koikegami (Second Lab, LLC) in constructing software for a Bayesian algorithm.

References

1. Guilloteau, P.; Zabielski, R.; Hammon, H.M.; Metges, C.C. Adverse effects of nutritional programming during prenatal and early postnatal life, some aspects of regulation and potential prevention and treatments. *J. Physiol. Pharmacol.* **2009**, *60*, 17–35.
2. Hales, C.N.; Barker, D.J. The thrifty phenotype hypothesis. *Br. Med. Bull.* **2001**, *60*, 5–20.
3. Heindel, J.J. Role of exposure to environmental chemicals in the developmental basis of reproductive disease and dysfunction. *Semin. Reprod. Med.* **2006**, *24*, 168–177.
4. McMillen, I.C.; Robinson, J.S. Developmental origins of the metabolic syndrome: prediction, plasticity, and programming. *Physiol. Rev.* **2005**, *85*, 571–633.
5. Crofton, K.M.; Mundy, W.R.; Lein, P.J.; Bal-Price, A.; Coecke, S.; Seiler, A.E.; Knaut, H.; Buzanska, L.; Goldberg, A. Developmental neurotoxicity testing: Recommendations for developing alternative methods for the screening and prioritization of chemicals. *ALTEX* **2011**, *28*, 9–15.

6. Chapin, R.E.; Stedman, D.B. Endless possibilities: Stem cells and the vision for toxicology testing in the 21st century. *Toxicol. Sci.* **2009**, *112*, 17–22.
7. Seiler, A.E.; Buesen, R.; Visan, A.; Spielmann, H. Use of murine embryonic stem cells in embryotoxicity assays: the embryonic stem cell test. *Methods Mol. Biol.* **2006**, *329*, 371–395.
8. van Dartel, D.A.; Piersma, A.H. The embryonic stem cell test combined with toxicogenomics as an alternative testing model for the assessment of developmental toxicity. *Reprod. Toxicol.* **2011**, *32*, 235–244.
9. Theunissen, P.T.; Schulpen, S.H.W.; van Dartel, D.A.M.; Hermsen, S.A.B.; van Schooten, F.J.; Piersma, A.H. An abbreviated protocol for multilineage neural differentiation of murine embryonic stem cells and its perturbation by methyl mercury. *Reprod. Toxicol.* **2010**, *29*, 383–392.
10. Zimmer, B.; Kuegler, P.B.; Baudis, B.; Genewsky, A.; Tanavde, V.; Koh, W.; Tan, B.; Waldmann, T.; Kadereit, S.; Leist, M. Coordinated waves of gene expression during neuronal differentiation of embryonic stem cells as basis for novel approaches to developmental neurotoxicity testing. *Cell Death Differ.* **2011**, *18*, 383–395.
11. Coecke, S.; Goldberg, A.M.; Allen, S.; Buzanska, L.; Calamandrei, G.; Crofton, K.; Hareng, L.; Hartung, T.; Knaut, H.; Honegger, P.; *et al.* Workgroup report: incorporating *in vitro* alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies. *Environ. Health Perspect.* **2007**, *115*, 924–931.
12. Friedman, N.; Linial, M.; Nachman, I.; Pe'er, D. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **2000**; *7*, 601–620.
13. Husmeier, D. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics* **2003**, *19*, 2271–2282.
14. Rogers, S.; Girolami, M. A Bayesian regression approach to the inference of regulatory networks from gene expression data. *Bioinformatics* **2005**, *21*, 3131–3137.
15. Toyoshiba, H.; Yamanaka, T.; Sone, H.; Parham, F.M.; Walker, N.J.; Martinez, J.; Portier, C.J. Gene interaction network suggests dioxin induces a significant linkage between aryl hydrocarbon receptor and retinoic acid receptor beta. *Environ. Health Perspect.* **2004**, *112*, 1217–1224.
16. Yamanaka, T.; Toyoshiba, H.; Sone, H.; Parham, F.M.; Portier, C.J. The TAO-Gen algorithm for identifying gene interaction networks with application to SOS repair in *E. coli*. *Environ. Health Perspect.* **2004**, *112*, 1614–1621.
17. Jayawardhana, B.; Kell, D.B.; Rattray, M. Bayesian inference of the sites of perturbations in metabolic pathways via Markov chain Monte Carlo. *Bioinformatics* **2008**, 24, 1191–1197.
18. Tang, W.; Wu, X.; Jiang, R.; Li, Y. Epistatic module detection for case-control studies: A Bayesian model with a Gibbs sampling strategy. *PLoS Genet.* **2009**, *5*, doi:10.1371/ journal.pgen.1000464.
19. Anchang, B.; Sadeh, M.J.; Jacob, J.; Tresch, A.; Vlad, M.O.; Oefner, P.J.; Spang, R. Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6447–6452.

20. Harikrishnan, K.N.; Chow, M.Z.; Baker, E.K.; Pal, S.; Bassal, S.; Brasacchio, D.; Wang, L.; Craig, J.M.; Jones, P.L.; Sif, S.; *et al.* Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat. Genet.* **2005**, *37*, 254–264.
21. Zhang, A.; Shen, C.H.; Ma, S.Y.; Ke, Y.; El Idrissi, A. Altered expression of Autism-associated genes in the brain of Fragile X mouse model. *Biochem. Biophys. Res. Commun.* **2009**, *379*, 920–923.
22. Chao, H.T.; Chen, H.; Samaco, R.C.; Xue, M.; Chahrour, M.; Yoo, J.; Neul, J.L.; Gong, S.; Lu, H.C.; Heintz, N.; *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **2010**, *468*, 263–269.
23. Goncharova, E.A.; Goncharov, D.A.; Li, H.; Pimtong, W.; Lu, S.; Khavin, I.; Krymskay, V.P. mTORC2 is required for proliferation and survival of TSC2-null cells. *Mol. Cell. Biol.* **2011**, *31*, 2484–2498.
24. Nasuti, C.; Gabbianelli, R.; Falcioni, M.L.; di Stefano, A.; Sozio, P.; Cantalamessa, F. Dopaminergic system modulation, behavioral changes, and oxidative stress after neonatal administration of pyrethroids. *Toxicology* **2007**, *229*, 194–205.
25. Elwan, M.A.; Richardson, J.R.; Guillot, T.S.; Caudle, W.M.; Miller, G.W. Pyrethroid pesticide-induced alterations in dopamine transporter function. *Toxicol. Appl. Pharmacol.* **2006**, *211*, 188–197.
26. Kyoto Encyclopedia of Genes and Genomes. Available online: http://www.genome.jp/kegg/pathway/hsa/hsa05012.html (accessed on 9 January 2011).
27. Profiles of Chemical Effects on Cells system. Available online: http://project.nies.go.jp/eCA/cgi-bin/index.cgi (accessed on 28 February 2007).
28. Sone, H.; Okura, M.; Zaha, H.; Fujibuchi, W.; Taniguchi, T.; Akanuma, H.; Nagano, R.; Ohsako, S.; Yonemoto, J. Profiles of Chemical Effects on Cells (pCEC): A toxicogenomics database with a toxicoinformatics system for risk evaluation and toxicity prediction of environmental chemicals. *J. Toxicol. Sci.* **2010**, *35*, 115–123.
29. Gene Expression Omnibus. Available online: http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE18503 (accessed on 10 October 2009).
30. Green, K.N.; Billings, L.M.; Roozendaal, B.; McGaugh, J.L.; LaFerla, F.M. Glucocorticoids increase amyloid-beta and tau pathology in a mouse model of Alzheimer's disease. *J. Neurosci.* **2006**, *26*, 9047–9056.
31. Arguelles, S.; Herrera, A.J.; Carreno-Muller, E.; de Pablos, R.M.; Villaran, R.F.; Espinosa-Oliva, A.M.; Machado, A.; Cano, J. Degeneration of dopaminergic neurons induced by thrombin injection in the substantia nigra of the rat is enhanced by dexamethasone: Role of monoamine oxidase enzyme. *Neurotoxicology* **2010**, *31*, 55–66.
32. Cummings, J.L.; Vinters, H.V.; Cole, G.M.; Khachaturian, Z.S. Alzheimer's disease: Etiologies, pathophysiology, cognitive reserve, and treatment opportunities. *Neurology* **1998**, *51*, S2–S17; discussion S65–S67.

33. van den Eeden, S.K.; Tanner, C.M.; Bernstein, A.L.; Fross, R.D.; Leimpeter, A.; Bloch, D.A.; Nelson, L.M. Incidence of Parkinson's disease: Variation by age, gender, and race/ethnicity. *Am. J. Epidemiol.* **2003**, *157*, 1015–1022.
34. van Dartel, D.A.; Pennings, J.L.; de la Fonteyne, L.J.; van Herwijnen, M.H.; van Delft, J.H.; van Schooten, F.J.; Piersma, A.H. Monitoring developmental toxicity in the embryonic stem cell test using differential gene expression of differentiation-related genes. *Toxicol. Sci.* **2010**, *116*, 130–139.
35. Chandler, K.J.; Barrier, M.; Jeffay, S.; Nichols, H.P.; Kleinstreuer, N.C.; Singh, A.V.; Reif, D.M.; Sipes, N.S.; Judson, R.S.; Dix, D.J.; *et al.* Evaluation of 309 environmental chemicals using a mouse embryonic stem cell adherent cell differentiation and cytotoxicity assay. *PLoS One* **2011**, *6*, doi:10.1371/journal.pone.0018540.
36. Gohlke, J.M.; Armant, O.; Parham, F.M.; Smith, M.V.; Zimmer, C.; Castro, D.S.; Nguyen, L.; Parker, J.S.; Gradwohl, G.; Portier, C.J.; *et al.* Characterization of the proneural gene regulatory network during mouse telencephalon development. *BMC Biol.* **2008**, *6*, doi:10.1186/1741-7007-6-15.
37. Zhao, L.; Morgan, M.A.; Parsels, L.A.; Maybaum, J.; Lawrence, T.S.; Normolle, D. Bayesian hierarchical changepoint methods in modeling the tumor growth profiles in xenograft experiments. *Clin. Cancer Res.* **2011**, *17*, 1057–1064.
38. Landers, J.P.; Bunce, N.J. The Ah receptor and the mechanism of dioxin toxicity. *Biochem. J.* **1991**, *276*, 273–287.
39. Gould, A.; Missailidis, S. Targeting the hedgehog pathway: The development of cyclopamine and the development of anti-cancer drugs targeting the hedgehog pathway. *Mini. Rev. Med. Chem.* **2011**, *11*, 200–213.
40. Okada, Y.; Shimazaki, T.; Sobue, G.; Okano, H. Retinoic-acid-concentration-dependent acquisition of neural cell identity during *in vitro* differentiation of mouse embryonic stem cells. *Dev. Biol.* **2004**, *275*, 124–142.
41. Vargesson, N. Thalidomide-induced limb defects: Resolving a 50-year-old puzzle. *Bioessays* **2009**, *31*, 1327–1336.
42. Rodier, P.M.; Ingram, J.L.; Tisdale, B.; Nelson, S.; Romano, J. Embryological origin for autism: Developmental anomalies of the cranial nerve motor nuclei. *J. Comp. Neurol.* **1996**, *370*, 247–261.
43. Stromland, K.; Nordin, V.; Miller, M.; Akerstrom, B.; Gillberg, C. Autism in thalidomide embryopathy: A population study. *Dev. Med. Child Neurol.* **1994**, *36*, 351–356.
44. Narita, M.; Oyabu, A.; Imura, Y.; Kamada, N.; Yokoyama, T.; Tano, K.; Uchida, A.; Narita, N. Nonexploratory movement and behavioral alterations in a thalidomide or valproic acid-induced autism model rat. *Neurosci. Res.* **2010**, *66*, 2–6.
45. Meerts, I.A.; Lilienthal, H.; Hoving, S.; van den Berg, J.H.; Weijers, B.M.; Bergman, A.; Koeman, J.H.; Brouwer, A. Developmental exposure to 4-hydroxy-2,3,3',4',5-pentachlorobiphenyl (4-OH-CB107): Long-term effects on brain development, behavior, and brain stem auditory evoked potentials in rats. *Toxicol. Sci.* **2004**, *82*, 207–218.

46. Stump, D.G.; Beck, M.J.; Radovsky, A.; Garman, R.H.; Freshwater, L.L.; Sheets, L.P.; Marty, M.S.; Waechter, J.M., Jr.; Dimond, S.S.; van Miller, J.P.; *et al.* Developmental neurotoxicity study of dietary bisphenol A in Sprague-Dawley rats. *Toxicol. Sci.* **2010**, *115*, 167–182.
47. Dalgaard, M.; Ostergaard, G.; Lam, H.R.; Hansen, E.V.; Ladefoged, O. Toxicity study of di(2-ethylhexyl) phthalate (DEHP) in combination with acetone in rats. *Pharmacol. Toxicol.* **2000**, *86*, 92–100.
48. Kakko, I.; Toimela, T.; Tahti, H. The synaptosomal membrane bound ATPase as a target for the neurotoxic effects of pyrethroids, permethrin and cypermethrin. *Chemosphere* **2003**, *51*, 475–480.
49. Numis, A.L.; Major, P.; Montenegro, M.A.; Muzykewicz, D.A.; Pulsifer, M.B.; Thiele, E.A. Identification of risk factors for autism spectrum disorders in tuberous sclerosis complex. *Neurology* **2011**, *76*, 981–987.
50. O'Roak, B.J.; State, M.W. Autism genetics: Strategies, challenges, and opportunities. *Autism Res.* **2008**, *1*, 4–17.
51. Muhle, R.; Trentacoste, S.V.; Rapin, I. The genetics of autism. *Pediatrics* **2004**, *113*, e472–e486.
52. Wang, J.; Rao, S.; Chu, J.; Shen, X.; Levasseur, D.N.; Theunissen, T.W.; Orkin, S.H. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **2006**, *444*, 364–368.
53. Muller, F.J.; Laurent, L.C.; Kostka, D.; Ulitsky, I.; Williams, R.; Lu, C.; Park, I.H.; Rao, M.S.; Shamir, R.; Schwartz, P.H.; *et al.* Regulatory networks define phenotypic classes of human stem cell lines. *Nature* **2008**, *455*, 401–405.
54. *GeneSpring*, version GX10.02; Agilent Technologies: Palo Alto, CA, USA, 2010.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 07 August 2012
doi: 10.3389/fgene.2012.00141

Identification of stage-specific gene expression signatures in response to retinoic acid during the neural differentiation of mouse embryonic stem cells

Hiromi Akanuma[1], Xian-Yang Qin[1,2], Reiko Nagano[1], Tin-Tin Win-Shwe[3], Satoshi Imanishi[1], Hiroko Zaha[1], Jun Yoshinaga[2], Tomokazu Fukuda[4], Seiichiroh Ohsako[5] and Hideko Sone[1]*

[1] *Health Risk Research Section, Center for Environmental Risk Research, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan*
[2] *Department of Environmental Studies, Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Chiba, Japan*
[3] *Biological Impact Research Section, Center for Environmental Health Sciences, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan*
[4] *Department of Animal Production Science, Graduate School of Agricultural Science, Tohoku University, Sendai, Miyagi, Japan*
[5] *Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*

Edited by:
Pierre R. Bushel, National Institute of Environmental Health Sciences, USA

Reviewed by:
Ryou Fukushima, Shionogi & Co., Ltd., Japan
Manikandan Jayapal, King Abdulaziz University, Saudi Arabia

***Correspondence:**
Hideko Sone, Health Risk Research Section, Center for Environmental Risk Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba 305-8506, Japan.
e-mail: hsone@nies.go.jp

We have previously established a protocol for the neural differentiation of mouse embryonic stem cells (mESCs) as an efficient tool to evaluate the neurodevelopmental toxicity of environmental chemicals. Here, we described a multivariate bioinformatic approach to identify the stage-specific gene sets associated with neural differentiation of mESCs. We exposed mESCs (B6G-2 cells) to 10^{-8} or 10^{-7} M of retinoic acid (RA) for 4 days during embryoid body formation and then performed morphological analysis on day of differentiation (DoD) 8 and 36, or genomic microarray analysis on DoD 0, 2, 8, and 36. Three gene sets, namely a literature-based gene set (set 1), an analysis-based gene set (set 2) using self-organizing map and principal component analysis, and an enrichment gene set (set 3), were selected by the combined use of knowledge from literatures and gene information selected from the microarray data. A gene network analysis for each gene set was then performed using Bayesian statistics to identify stage-specific gene expression signatures in response to RA during mESC neural differentiation. Our results showed that RA significantly increased the size of neurosphere, neuronal cells, and glial cells on DoD 36. In addition, the gene network analysis showed that glial fibrillary acidic protein, a neural marker, remarkably up-regulates the other genes in gene set 1 and 3, and *Gbx2*, a neural development marker, significantly up-regulates the other genes in gene set 2 on DoD 36 in the presence of RA. These findings suggest that our protocol for identification of developmental stage-specific gene expression and interaction is a useful method for the screening of environmental chemical toxicity during neurodevelopmental periods.

Keywords: mouse embryonic stem cells, neural differentiation, Bayesian network, retinoic acid, toxicity screening

INTRODUCTION

Humans are exposed to environmental chemicals on a daily basis; however, many effects of these chemicals on human health are unclear. Currently, assessment of developmental toxicity on children's health is a large and rapidly growing research field. Children are not "little adults" and have special vulnerabilities to the toxic effects of environmental chemicals. For example, brain development during embryonic stages is an important period when microstructures are formed and axon guidance and synapse formation are induced by neuronal signaling (Lamoury et al., 2006; Ligon et al., 2006). These processes are regulated by stage-specific gene expression during embryonic development. Therefore, it is necessary to develop a more comprehensive and efficient system to identify the stage-specific gene expression signatures in embryonic development and to evaluate the toxicity of environmental chemicals on neural development.

Toxicity testing using embryonic stem cells (ESCs) has been developed as an efficient approach to assess the effect of environmental chemicals on neurodevelopment (Seiler et al., 2006). We have previously reported a mouse embryonic stem cell (mESC) neural differentiation protocol and showed that it could be used as an efficient tool to evaluate the toxic effects of environmental chemicals on neurodevelopment (Nagano et al., 2012). Furthermore, we have previously developed a method to quantitatively and statistically analyze microarray gene expression data using Bayesian networks with a log-linear functional relationship between genes (Toyoshiba et al., 2004, 2006). We proposed that advanced Bayesian network analysis is a necessary tool to understand the accurate linkage in the possible networks and the mechanism of the action of developmentally neurotoxic compounds.

During mammalian fetal development, the most active form of vitamin A, retinoic acid (RA) can pass through the umbilical cord to the fetus and induce axon formation and neural system development. ESCs express high levels of RA receptor (RAR)α in the undifferentiated stage, while RARβ begins to be expressed after

embryoid body (EB) formation (Shiotsugu et al., 2004; Wilson and Maden, 2005; So et al., 2006). A series of RA concentrations were examined to detect neural cell identity during neuronal differentiation from mESC (Okada et al., 2004; Engberg et al., 2010). They reported that the 10^{-8} M of RA would be an optimum dose to induce cerebral and mesencephalic neurons and the 10^{-7} M of RA had capability to induce motor neurons (Kawasaki et al., 2000; Nishimura et al., 2003; Miyazaki et al., 2005).

Therefore, in the present study, we focused on identification of stage-specific gene expressions and analyzed their relationship network during mESC neurodevelopmental period after RA exposure at 10^{-8} and 10^{-7} M, using an advanced Bayesian network analysis.

MATERIALS AND METHODS

CELL CULTURE AND DIFFERENTIATION

B6G-2 mESCs (RIKEN Cell Bank, Tsukuba, Ibaraki, Japan) were maintained in Dulbecco's modified Eagle's medium (Invitrogen, Carlsbad, CA, USA) supplemented with 15% knockout serum replacement (Invitrogen), 100 μM non-essential amino acids (Invitrogen), 100 μM 2-mercaptoethanol (Invitrogen), and 1000 U/ml leukemia inhibitory factor (LIF; Invitrogen) in gelatinized tissue culture dishes. On the first day of differentiation (DoD 0), cells were transferred in to 24 well plate in media without LIF and allowed to form EBs. The media was changed by every 2 days. RA was added during DoD 2–6 to induce neuronal differentiation. On DoD 8, EBs were transferred to L-ornithine/laminin-coated 24 well plates (BD Bio Coat, BD, Franklin Lakes, NJ, USA) and were cultured with neural medium from DoD 22 to DoD 36 to promote further neural differentiation (**Figure 1A**).

IMMUNOCYTOCHEMISTRY AND MORPHOLOGICAL ANALYSIS

On DoD 8 and DoD 36, EBs, and their derivatives were fixed with 4% PFA in PBS for 15 min and then performed immunostaining with the conventional methods. Cells were incubated with primary antibodies overnight at 4°C at the following dilutions: anti-microtubule-associated protein 2 (*Map2*) antibody (Sigma-Aldrich, Poole, UK; 1:200) and mouse anti-glial fibrillary acidic protein (*Gfap*) monoclonal antibody (Chemicon International,

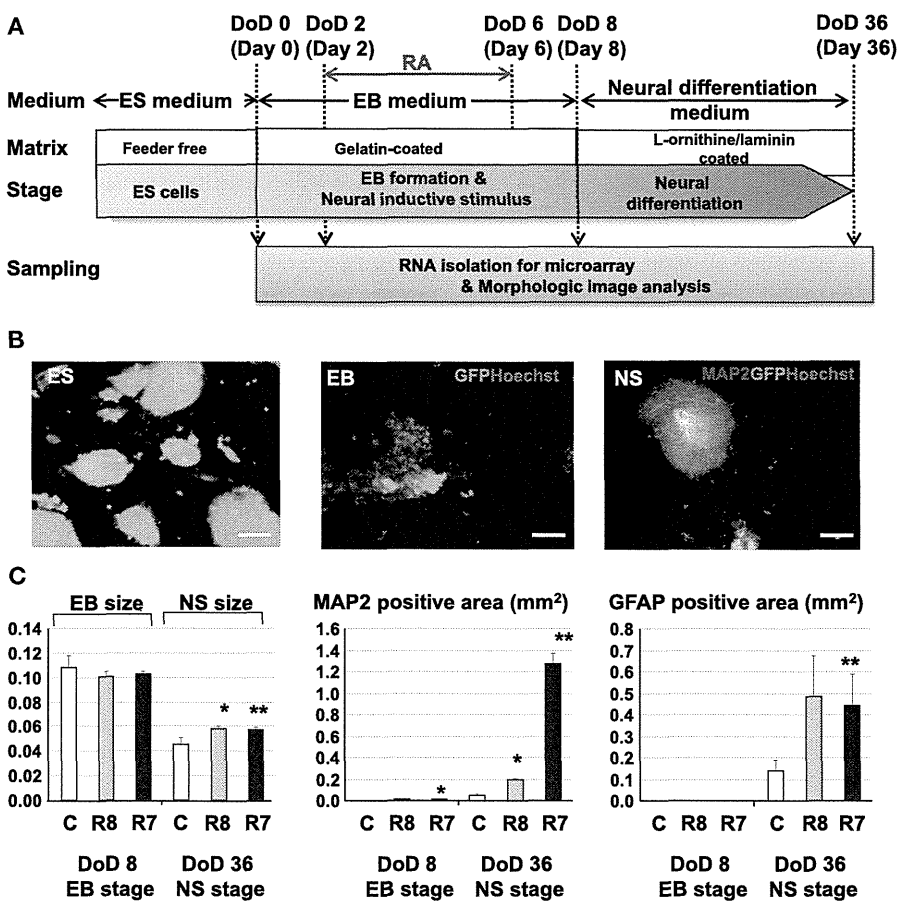


FIGURE 1 | Experimental protocols and morphological analysis of the effect of RA on the neural differentiation of mESCs. (A) A Schematic diagram of cell cultures and chemical exposures. **(B)** Pictures of mESCs, EBs, and neurospheres (NS) with neurite growth derived from neurosphere in fluorescent fields. Green, blue, and red fluorescent images indicate GFP of cytoskeleton proteins, nuclei stained with Hoechst and neuronal cells and neurites stained with the anti *Map2* antibody. Scale bar is 100 μm. **(C)** Morphological analysis of neuronal cell lineages exposed to RAs. Comparison of the EB areas on DoD 8 and DoD 36 showed that the EB area decreased with neuronal cell development. In RA-treated EBs, the numbers of *Map2*-positive axons and *Gfap*-positive cells were increased compared with the control. Symbols of C, R8, and R7 indicate vehicle control, RA 10^{-8} M, and RA 10^{-7} M.

Temecula, CA, USA; 1:200). Cells were rinsed with PBS and then incubated with Alexa-conjugated secondary antibodies (1:1000, Alexa Fluor 546, Invitrogen). Hoechst 33342 solution (Dojindo Laboratories, Kumamoto, Japan) was used for counter-staining. Immunofluorescence images were acquired with six biological replicates per condition using an IN Cell Analyzer 1000 (GE Healthcare, Buckinghamshire, UK) and analyzed using IN Cell Developer Tool Box 1.7 (GE Healthcare). All morphological analysis experiments were performed in triplicate to test the reproducibility of the results. Statistical analysis was performed using two-tailed Student's t-test. Relationships were considered statistically significant with $p < 0.05$.

DNA MICROARRAY ANALYSIS

Total RNA was isolated on DoD 0, 2, 8, and 36 with six biological replicates. And then, single mixed RNA sample per condition was applied to Illumina MouseWG-6v1.0 expression BeadChips covering 46,643 transcripts including 26,766 annotated coding transcripts 2, according to the manufacturer's instructions (Illumina, San Diego, CA, USA). The arrays were scanned in accordance with the manufacturer's directions. Raw expression values of each gene were normalized with median centered by GeneSpring GX10.02 software (Agilent Technologies, Palo Alto, CA, USA). Normalized data were deposited in the National Center for Biotechnology Information Gene Expression Omnibus[1] (accession no. GSE37602).

SELECTION OF GENE SETS

To capture gene expression signatures of stage-specific changes during neural differentiation of mESCs, we performed three approaches to determine gene sets for Bayesian network analysis. Marker genes, which are commonly used to analyze pluripotency and development of neural cells, were selected as the literature-based gene set (set 1) by review of the published literature. The analysis-based gene set (set 2) was selected by the combined use of the knowledge-based database and the following classification methods. Candidate genes involved in axon guidance maps, the nerve growth factor (NGF) pathway, and RA signaling were preliminarily selected from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database[2] and then genes with specific expression patterns were identified using SOM and PCA. Finally, the enrichment gene set (set 3) was selected by clustering expression values of candidate genes contained in the Neurogenesis and Neural Stem Cell PCR Array (SABiosciences, Valencia, CA, USA) using SOM and PCA. SOM and PCA were performed using GeneSpring GX10.02 software (Agilent Technology). Briefly, SOM clustering was done by conditions in which similarity measure: euclidean, maximum number of iterations: 50, numbers of grid rows and columns were 2×4. Then each eight clusters of SOM were analyzed by PCA with four components of eigenvalues (component 1 was more than 40% and component 2 was 10%). To develop set 2 and 3, we collected genes with maximum and minimum values in the PCA component 1 from each SOM cluster.

[1] www.ncbi.nlm.nih.gov/geo

[2] http://www.genome.jp/kegg/pathway.html

GENE INTERACTION NETWORK ANALYSIS

We used a modified gene interaction network (GIN) based on our previous studies (Yamanaka et al., 2004; Toyoshiba et al., 2006; Nagano et al., 2012). The GIN was quantified to calculate the posterior probability distribution for the strength of the linkages based on gene expression and chemical exposure dose datasets. Briefly, a GIN consists of a collection of P nodes, denoted G_1, G_2, ..., G_P, with observed values n_1, n_2, ... n_P. βij ($i, j = 1, 2, \ldots, P$) are parameters in the log-linear function form describing the linkage from node i to node j. Mathematically, this is written as

$$E\left[\log\left(G_j\right)\right] = \sum_{i=1,\neq j}^{P} I_{ij}\beta_{ij}\log\left(n_i\right)$$

where $E[\log(G_j)]$ represents the expectation for the natural logarithm of G_j, and $I_{ij}(i, j = 1, 2, \ldots, P)$ is an indicator function that equals 1 if node G_i has a link to node G_j, otherwise it equals 0. If a node has a regulatory effect on node G_i, then that node is referred to as a "Parent of node G_i," and we refer to it as belonging to the set Pa (G_i). The prior distribution for I_{ij} was assumed to be a Bernoulli distribution with success probability p_{ij} when $Iij = 1$. In the uninformative case, p_{ij} could be set to 0.5 and if there is some expectation that I_{ij} is not equal to zero, the prior probability could be set higher. The posterior distributions for the linkages were derived using Gibbs sampling. The network was used to evaluate the ability of the algorithm to have a higher posterior probability (p-value). Transition matrices were generated at $p > 0.5$.

RESULTS

EFFECTS OF RA ON NEURAL DIFFERENTIATION

Exposure to RA at different concentrations during EB formation induced neuronal and glial cell lineages from mESCs (**Figure 1A**). Morphological analysis with immunofluorescent staining showed that RA significantly increased the size of neurosphere, neuronal cells, and glial cells at DoD 36 (**Figures 1B,C**).

GENE SET SELECTION FOR GENE NETWORK ANALYSIS

To investigate transcriptomic changes as a result of neuronal differentiations and influences of RAs, a cDNA microarray was used to compare expression levels with and without the RA treatments in EB formations and neurosphere developments by hierarchical clustering methods (**Figure 2A**). From 22,188 transcripts presented from eight microarrays, 1,157 transcripts with expression differences greater than 2.0-fold in at least 1 microarray were selected for further analysis. From the microarray analysis, *Nanog* as a marker of undifferentiated ESCs and, *Nestin*, *Map2*, and *Gfap* as markers of neural cells were differentially expressed by RA treatments at differential doses during the neural differentiation of mESCs, suggesting that our protocol could detect the effects of RA on neuronal differentiation (**Figure 2B**). A high level of *Nanog* expression on DoD 8 was decreased in a dose-dependent fashion following RA treatments, but not on DoD 36. *Nestin* expression was increased by the 10^{-7} M RA treatment on DoD 8 and DoD 36. *Map2* and *Gfap* expressions were also increased by RA treatments on DoD 8 and DoD 36 (**Figure 2B**).

Three gene sets were selected for the Bayesian network analysis by our strategies as shown in **Figure 2C**. Selected gene sets are

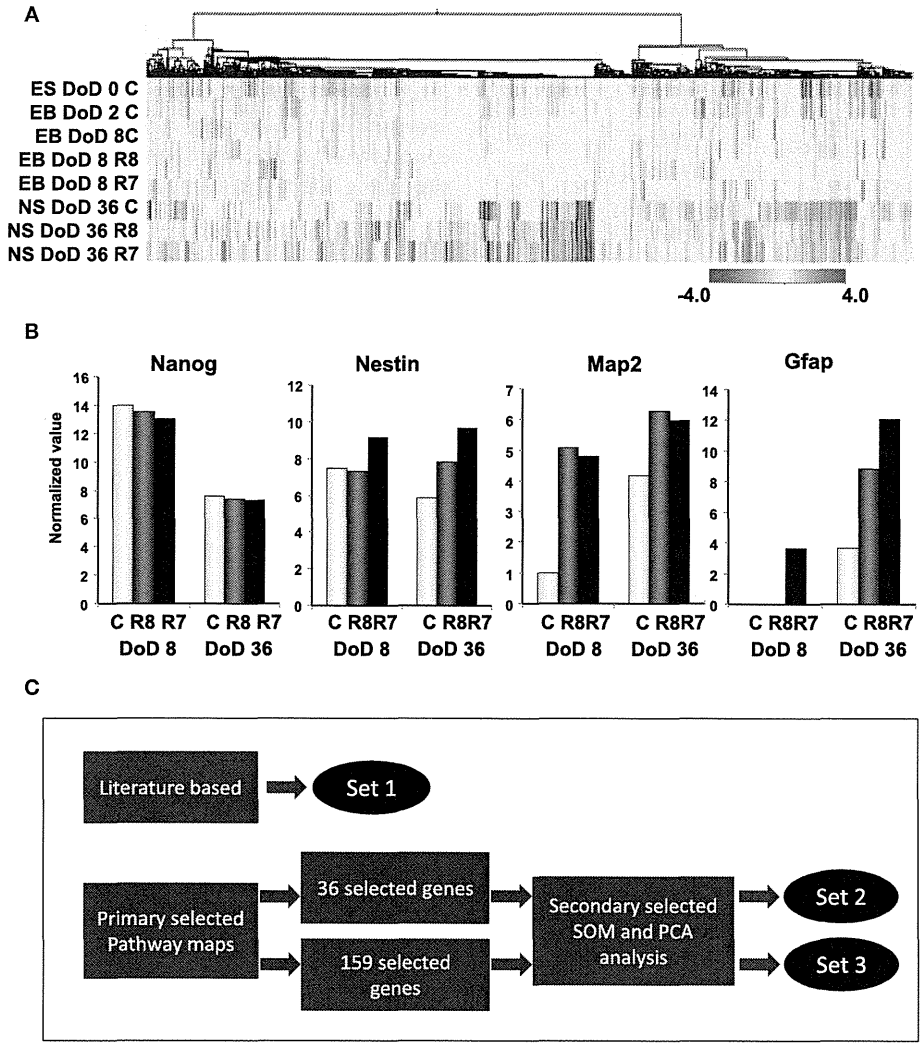


FIGURE 2 | Gene expression analysis by DNA microarray and gene selection strategies for the Bayesian network analysis of differentiation of neuronal cells derived from mESCs. (A) Heat map of hierarchical clustering generated from DNA microarray data. Color-coding in the heat map is that blue from red indicates – 4.0 from 4.0 log2 normalized intensity value by ES values, indicating that red is for up regulation and blue is for down regulation. **(B)** Gene expression of pluripotency and differentiation markers in mESCs, EB, and NS measured in DNA microarray. Symbols of C, R8, and R7 indicate vehicle control, RA 10^{-8} M, and RA 10^{-7} M. **(C)** Stage-specific gene expression signatures in response to RA during the neural differentiation of mESCs were identified as follows: set 1 was a set of genes selected from the literature; set 2 was selected by SOM and PCA after selecting 36 genes from pathway maps; set 3 was selected by SOM and PCA after selecting 159 genes from pathway maps. Expression values of microarray data corresponding to genes in these three sets were used for the Bayesian network analysis.

listed in **Table 1**. Concretely, set 1 was selected by the review of published articles and included *Nanog* (Mitsui et al., 2003; Loh et al., 2006), *Pou5f1* (Okazawa et al., 1991; Catena et al., 2004; Akamatsu et al., 2009), *Zfl42* (Shi et al., 2006; Scotland et al., 2009), *Fgfr1* (Jukkola et al., 2006; Yang et al., 2008; Lee et al., 2009), *Sox2* (Tomioka et al., 2002; Graham et al., 2003; Tanaka et al., 2004; Jin et al., 2009), and *Oligo2* (Ahn et al., 2008). *RARs* were also added to set 1 to assess the effects of RA. Set 2 was selected by the combined use of the KEGG database and SOM and PCA classification methods. Firstly, a list of 36 candidate genes was compiled according to axon guidance, NGF pathway, and RA signaling of KEGG database. It is known that NGF can induce neuronal differentiation of mESCs (Schuldiner et al., 2001) while RA can induce the expression of the NGF receptor (p75) during the neuronal differentiation of PC12 cells (Cosgaya et al., 1996). Therefore, genes in the NGF pathway were selected as indicators to assess the effects of RA on the neural differentiation of mESCs. The 36 candidate genes were then classified in to 17 classes by SOM, and representative genes were selected from each class by PCA. Finally, 16 genes were selected for set 3 by SOM and PCA clustering from 159 candidate genes contained in the Neurogenesis and Neural Stem Cell PCR Array (SABiosciences). Furthermore, specific markers for astrocytes (*Gfap*), mature neurons (*Map2*), neuronal stem cells (*Nestin*), and young neurons (*Tuj1*) were added to sets 2 and 3 to assess the stage of neuronal differentiation.

Table 1 | Lists of gene sets for identifying gene networks.

Category	Set1	Set2	Set3
Pathway signaling		Map2k1	Adora2a
		Mapk1	Drd5
		Mapk3	Fgf13
		Pla2g6	Gnao1
		Rps6ka1	Notch2
		Shc1	Tnr
Transcription/chromatin regulation	RARa	Atbf1	Ascl1
	RARb	Cdyl	Gusb
	RARg	Rhog	Mef2c
	Nanog	Rif1	Pax5
	Pou5f1	Sall1	Pou3f3
	Zfp42	Smarcad1	
Neural development	Fgfr1	Fos	Bdnf
	Olig2	Gbx2	Gdnf
	Sox2	Hras1	Nrp2
		Raf1	Slit2
		Sox2	Ywhah
Neural marker	Gfap	Gfap	Gfap
	Map2	Map2	Map2
	Nestin	Nestin	Nestin
	Tuj1	Tuji1	Tuji1

GENE INTERACTION ANALYSIS

Matrices transferred from gene interaction analysis for set 1, set 2, and set 3 are shown in **Figures 3–5**, respectively (see **Figures A1–A3** in Appendix as references and Tables S1–S6 in Supplementary Material for input data and output raw-results). In the control group of set 1, *Nanog*, and Sox2 (**Figure 3**) which control ESC pluripotency, regulate many other genes on DoD 0, 2, and 8. On DoD 36, Sox2 does not regulate any gene. In RA-treated groups of set 1, linkages of RARs in the matrix indicated that these genes might play principal roles in the regulation of expression of other genes. Briefly, the effect of RA was observed on DoD 8, in which RA 10^{-8} or 10^{-7} M aggravated *Nanog* and *Pou5f1*. On DoD 36, the matrix was more strongly influenced by RA, in which the neural marker genes such as *Gfap* and *Map2* up-regulated the other genes, indicating that RA enhances neural differentiation (**Figure 3**).

Gene interaction matrix analysis for set 2 is shown in **Figure 4**. In the mESC matrix, linkages between genes were concentrated to categories of pathway signaling and neural development, which is similar with those in set 1 on DoD 8 and DoD 36. It is noteworthy that Gbx2 as a neuronal development marker strongly up-regulated *Mapk3*, *Atbf1*, *Rhog*, *Sall1*, *Smarcad*, *Sox2*, and *Map2* on DoD 8 as well as DoD 36 in RA 10^{-8} M matrices. RA-treated matrices showed that linkages shifted to the right side of the matrix with increasing RA concentrations. Finally, linkages on DoD 36 were concentrated to categories of neural development and neural markers.

Gene interaction network analysis for set 3 is shown in **Figure 5**. In the mESC matrix, linkages between genes were concentrated to transcription/chromatin regulation and pathway signaling categories. Linkages between genes in the control matrix on DoD 8 were concentrated to pathway signaling and neural development categories. In the RA-treated matrices, linkages between genes moved to the neural marker category from the neural development category in a dose-dependent manner. Most of the linkages between genes in the RA-treated matrices on DoD 36 were concentrated to pathway signaling and neural marker categories, suggesting that *Gfap* mainly regulates neuronal differentiation.

DISCUSSION

In the present study, a prediction model for the neural differentiation of mESCs was established and stage-specific gene expression signatures in response to RA were identified using Bayesian network analysis. Our present findings showed that RA significantly increased the size of neurosphere, neuronal cells, and glial cells on DoD 36. In addition, neural marker *Gfap* remarkably up-regulated the other genes in gene set 1 and 3, and neural development marker *Gbx2* significantly up-regulated the other genes in gene set 2 on DoD 36 in the presence of RA. These findings suggest that our protocol for identification of developmental stage-specific gene expression and interaction is a useful method for the screening of environmental chemical toxicity during neurodevelopmental periods.

RA is known as a severe teratogen and causes central nervous system malformations. However, *in vivo* study indicated that high dose (70 mg/kg body weight; b.w.) of RA could induce teratogenic effects during gestational day 7–9 in Swiss mice (Veiga Quemelo et al., 2007). In addition, it was reported that the physiological dose that cannot affect RAR level was 1 mg/kg b.w. and minimally teratogenic dose was 10 mg/kg b.w. and completely teratogenic dose was 100 mg/kg b.w. in gestational day 9 of mouse (Harnisha et al., 1990). In the present study, we selected the dose of RA as 10^{-8} and 10^{-7} M because endogenous levels of RA-induced neural differentiation in the early embryo are approximately 1–10 nM (Maden et al., 1998; Mic et al., 2003). Therefore, we considered to use 10^{-8} M as a low dose and 10^{-7} M as a high dose to examine the effect of RA on stage-specific gene expression signature in mESCs.

We have also successfully designed a mESC neural differentiation protocol to evaluate the effect of RA on the neural differentiation of mESCs. Morphological analysis using a high-content image analyzer was able to acquire varying differences of differentiation from mESCs to neural cells by the RA treatment. For instance, neuronal or glial differentiation from neuronal ESCs was delayed in control cells without induction by RA, which was further confirmed by the lower expression levels of *Map2* and *Gfap* detected on DoD 36. RA treatments promoted the loss of pluripotency and differentiation into neural ESCs up to DoD 36 in the present study (**Figures 1B,C**), suggesting that the maturation of *Map2*-positive neurons and *Gfap*-positive astrocytes were accelerated by RA treatment. Our results are consistent with a study showing that RA and LIF enhance the induction of *Gfap*-positive astrocytes from mice neural progenitor cells via epigenetic modifications (Asano et al., 2009).

In restricted sample size analysis like the present study, simulations using Bayesian network analysis have been suggested to be a very effective method (Toyoshiba et al., 2006). Our present study provided a new experimental evidence that Bayesian

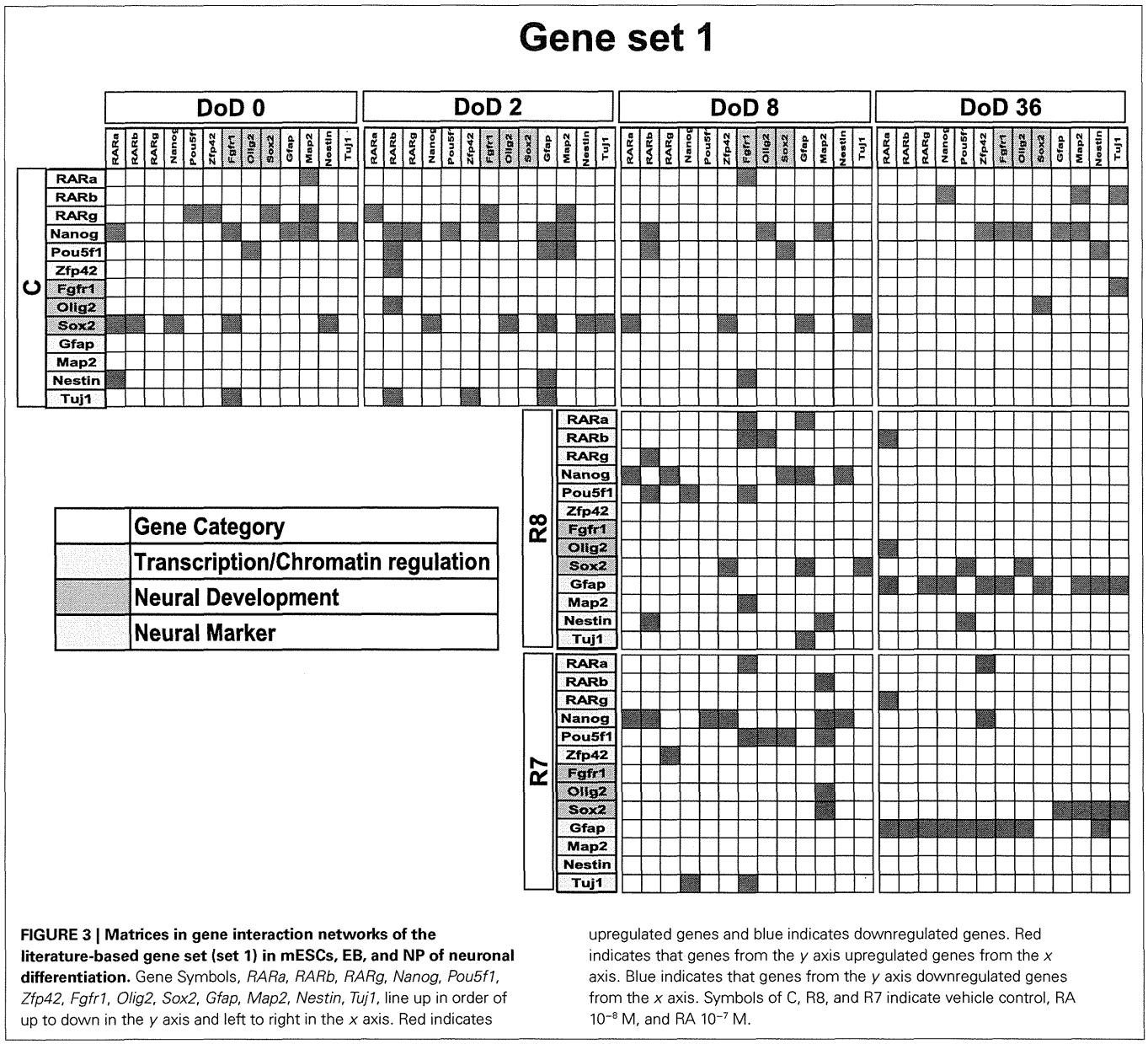


FIGURE 3 | Matrices in gene interaction networks of the literature-based gene set (set 1) in mESCs, EB, and NP of neuronal differentiation. Gene Symbols, *RARa, RARb, RARg, Nanog, Pou5f1, Zfp42, Fgfr1, Olig2, Sox2, Gfap, Map2, Nestin, Tuj1,* line up in order of up to down in the *y* axis and left to right in the *x* axis. Red indicates upregulated genes and blue indicates downregulated genes. Red indicates that genes from the *y* axis upregulated genes from the *x* axis. Blue indicates that genes from the *y* axis downregulated genes from the *x* axis. Symbols of C, R8, and R7 indicate vehicle control, RA 10^{-8} M, and RA 10^{-7} M.

network analysis was effective to identify the functions of the well-known neural development regulators, such as *Gfap* and *SOX2* (**Figures 3–5**), in response to RA during the neural differentiation of mESCs and suggested its further application to predict developmental neurotoxicity of environmental chemicals. However, in the simulation analysis, one major problem is to select genetic markers related with a trait of interest. To perform accurate simulation, it is undesirable to select genes with similar expression patterns. Similar variables could significantly affect the analysis results and potentially lead to biased results. Hence, the selection of genes with distinct expression patterns, which can represent each stage of mESC neural differentiation, seems to be important in the outcome of the GIN analysis. In this study, we selected gene sets for GIN analysis by the combined use of two classification methods, SOM and PCA. SOM is a powerful data mining method, whose algorithm is an unsupervised competitive learning neural network and it maps high-dimensional data into a simple low-dimensional display (Kohonen, 1990; Zhang et al., 2008). Therefore, SOM is able to classify the temporal expression data for each gene. After classification by SOM based on gene expression patterns, the representative genes were further selected from each class by PCA. PCA is a standard technique of pattern recognition and has been widely used as a tool in exploratory data analysis and for making predictive models in many biological systems (Aiba et al., 2006; van Dartel et al., 2010; Qin et al., 2011). In this study, the genes selected by SOM and PCA were shown to have adequate simulation parameters to evaluate the effects of RA on the neural differentiation of mESCs.

Finally, our prediction model, employing Bayesian network analysis, showed that it is possible to capture genetic correlations

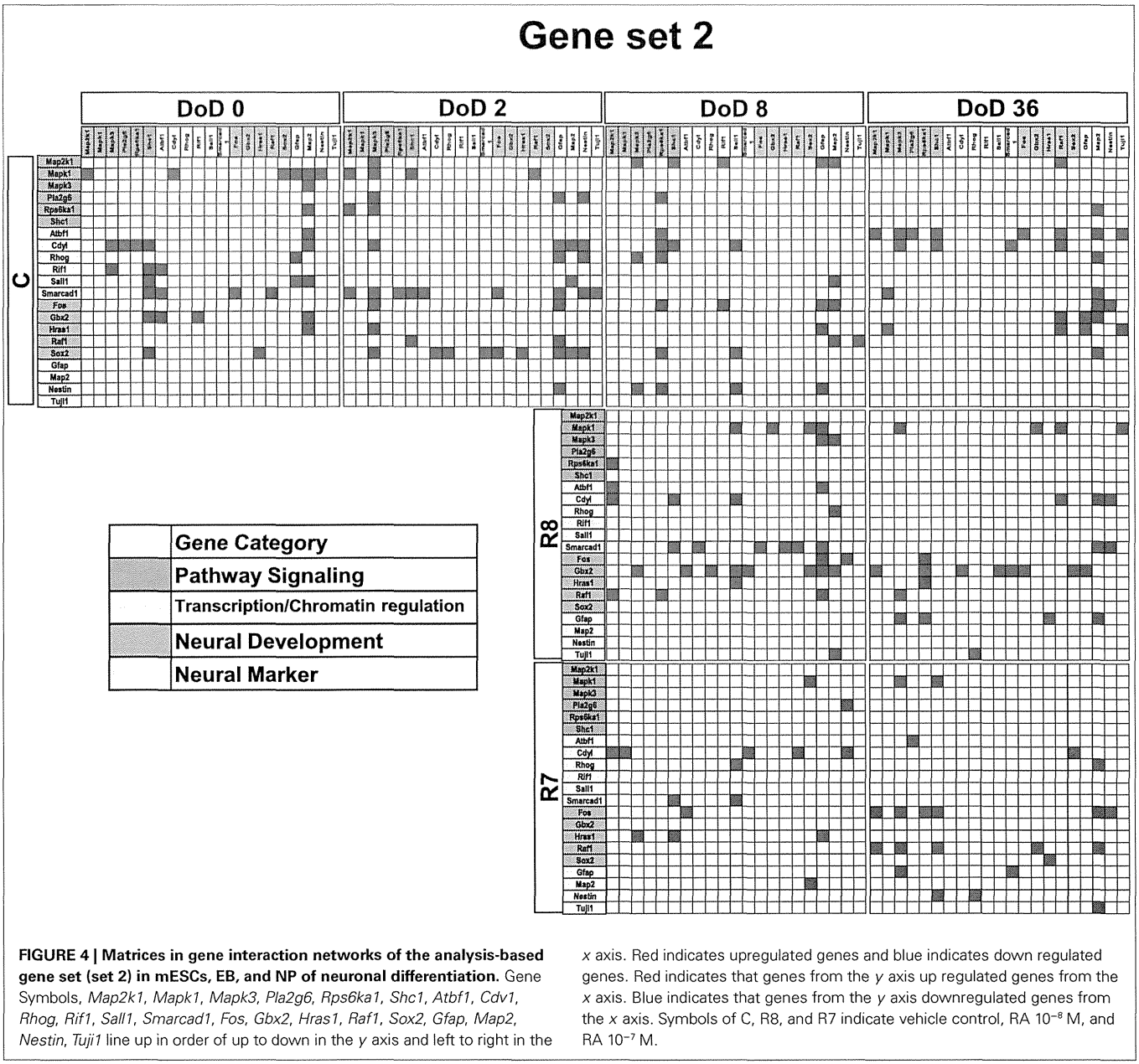


FIGURE 4 | Matrices in gene interaction networks of the analysis-based gene set (set 2) in mESCs, EB, and NP of neuronal differentiation. Gene Symbols, *Map2k1*, *Mapk1*, *Mapk3*, *Pla2g6*, *Rps6ka1*, *Shc1*, *Atbf1*, *Cdv1*, *Rhog*, *Rif1*, *Sall1*, *Smarcad1*, *Fos*, *Gbx2*, *Hras1*, *Raf1*, *Sox2*, *Gfap*, *Map2*, *Nestin*, *Tuji1* line up in order of up to down in the *y* axis and left to right in the *x* axis. Red indicates upregulated genes and blue indicates down regulated genes. Red indicates that genes from the *y* axis up regulated genes from the *x* axis. Blue indicates that genes from the *y* axis downregulated genes from the *x* axis. Symbols of C, R8, and R7 indicate vehicle control, RA 10^{-8} M, and RA 10^{-7} M.

between genes and to identify slight variations for different conditions. We performed the same prediction model for three gene sets of different genetic constitution. Our study indicated that the GIN was able to capture features of each developmental stage during the neural differentiation of mESCs. RA treatment could change the network structure in a dose-dependent manner. In addition, among the three gene sets, set 3 was the best according to the morphological results. We found that the *Gfap* gene was linked with other genes in the RA 10^{-7} M matrix in the GIN analysis, while the number of *Gfap*-positive cells was markedly increased by RA 10^{-7} M treatment on DoD 36 in the morphological analysis. This suggested that the approach used in this study, of the independent selection of gene sets using SOM or PCA, was efficient. This Bayesian model might also be useful to investigate the developmental toxicity of environmental chemicals other than RA.

In summary, to find the optimized GIN that integrated chemical effects, we created three different gene sets and then performed GIN analysis using Bayesian network algorithms to capture the stage-specific gene expression signatures in response to RA treatment during the neural differentiation of mESCs. "*Toxicity Testing in the Twenty First Century – A vision and a strategy*" issued by the US Nuclear Regulatory Commission indicated that the most important issue for toxicity testing is how to connect the extensive body of toxicity information to high-throughput screening to perform chemical risk assessment (Thomas et al., 2007; Davis et al., 2008; Ellinger-Ziegelbauer et al., 2009; Hubal, 2009). Here, we described a novel approach to identify stage-specific

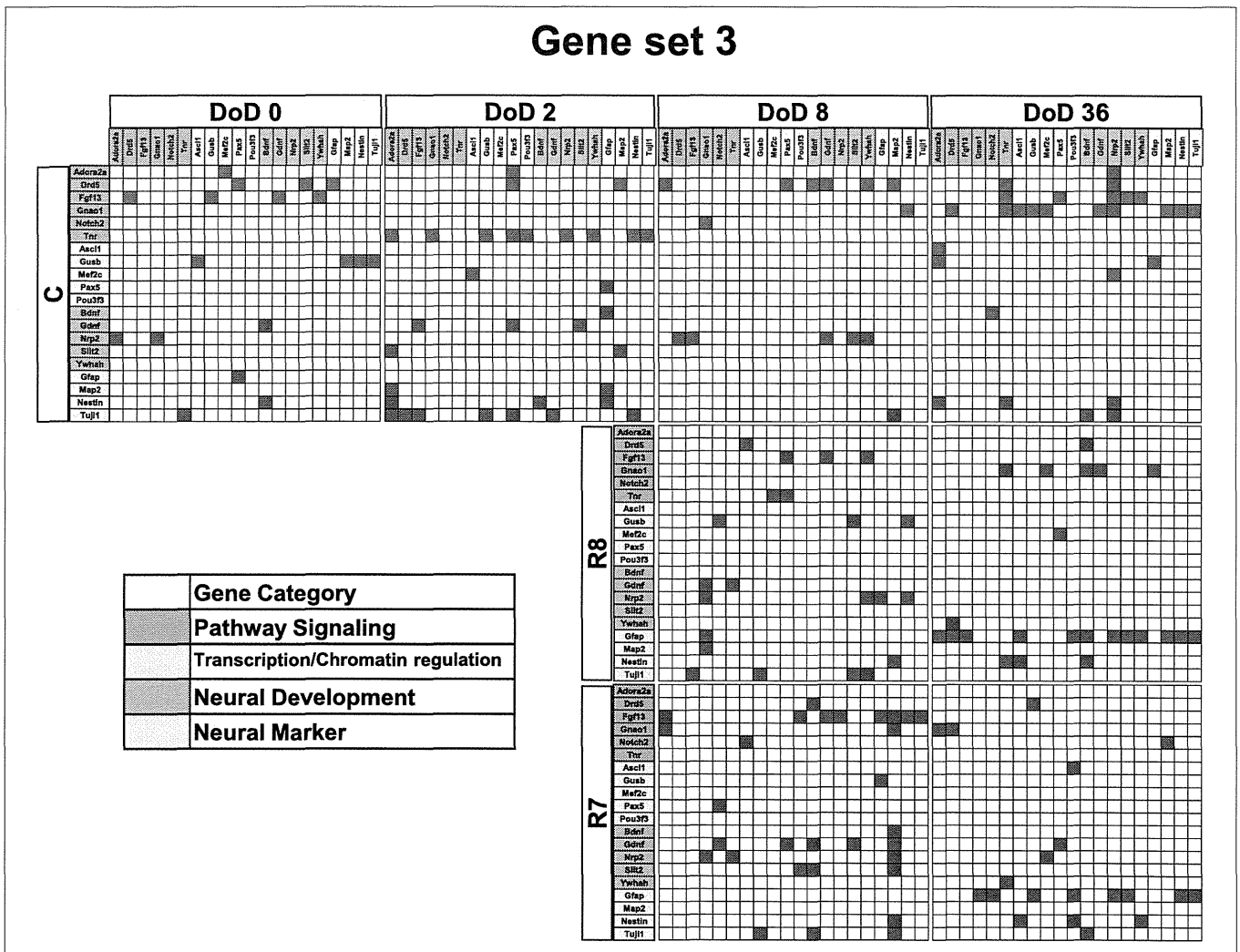


FIGURE 5 | Matrices in gene interaction networks of the enrichment gene set (set 3) in mESCs, EB, and NP of neuronal differentiation. Gene Symbols, *Adora2a, Drd5, Fgf13, Gnao1, Notch2, Tnr, Ascl1, Gusb, Mef2c, Pax5, Pou3f3, Bdnf, Gdnf, Nrp2, Slit2, Ywhah, Gfap, Map2, Nestin, Tuji1* line up in order of up to down in the *y* axis and left to right in the *x* axis. Red indicates up regulated genes and blue indicates down regulated genes. Red indicates that genes from the *y* axis up regulated genes from the *x* axis. Blue indicates that genes from the *y* axis down regulated genes from the *x* axis. Symbols of C, R8, and R7 indicate vehicle control, RA 10^{-8} M, and RA 10^{-7} M.

gene expression in embryonic development and suggested its application to evaluate the neural developmental toxicity of environmental chemicals in future studies.

ACKNOWLEDGMENTS

This study was supported, in part, by the Environmental Technology Development Fund (to Hideko Sone) from the Ministry of the Environment and by a Grant-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare, Japan (to Seiichiroh Ohsako). The authors gratefully acknowledge the technical support of Ms Noriko Oshima (GE Healthcare Corporation, Japan) for analysis using the IN Cell Analyzer 1000.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Toxicogenomics_/10.3389/fgene.2012.00141/abstract

Table S1 | Input data of gene expression values for set 1. Values are normalized expression values in beads of each gene with approximate 30 replicates.

Table S2 | Input data of gene expression values for set 2. Values are normalized expression values in beads of each gene with approximate 30 replicates.

Table S3 | Input data of gene expression values for set 3. Values are normalized expression values in beads of each gene with approximate 30 replicates.

Table S4 | Output results of the Bayesian network analysis with gene expression values for set 1. Values are posterior probabilities.

Table S5 | Output results of the Bayesian network analysis with gene expression values for set 2. Values are posterior probabilities.

Table S6 | Output results of the Bayesian network analysis with gene expression values for set 3. Values are posterior probabilities.