Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells

Junko Yamane[1,2], Sachiyo Aburatani[3], Satoshi Imanishi[2], Hiromi Akanuma[4], Reiko Nagano[4], Tsuyoshi Kato[5], Hideko Sone[4], Seiichiroh Ohsako[2] and Wataru Fujibuchi[1,3,*]

[1]Center for iPS Cell Research and Application, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan, [2]Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan, [3]Computational Biology Research Center, Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan, [4]Research Center for Environmental Risk, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan and [5]Department of Computer Science, Gunma University, 1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan

ABSTRACT

Predictive toxicology using stem cells or their derived tissues has gained increasing importance in biomedical and pharmaceutical research. Here, we show that toxicity category prediction by support vector machines (SVMs), which uses qRT-PCR data from 20 categorized chemicals based on a human embryonic stem cell (hESC) system, is improved by the adoption of gene networks, in which network edge weights are added as feature vectors when noisy qRT-PCR data fail to make accurate predictions. The accuracies of our system were 97.5–100% for three toxicity categories: neurotoxins (NTs), genotoxic carcinogens (GCs) and non-genotoxic carcinogens (NGCs). For two uncategorized chemicals, bisphenol-A and permethrin, our system yielded reasonable results: bisphenol-A was categorized as an NGC, and permethrin was categorized as an NT; both predictions were supported by recently published papers. Our study has two important features: (i) as the first study to employ gene networks without using conventional quantitative structure-activity relationships (QSARs) as input data for SVMs to analyze toxicogenomics data in an hESC validation system, it uses additional information of gene-to-gene interactions to significantly increase prediction accuracies for noisy gene expression data; and (ii) using only undifferentiated hESCs, our study has considerable potential to predict late-onset chemical toxicities, including abnormalities that occur during embryonic development.

INTRODUCTION

In recent years, several massive toxicological analyses have yielded large volumes of microarray data sets. Such undertakings as the chemical effects in biological systems (CEBS) project (1), a study of the effects of 2,4,6-trinitrotoluene (TNT) on the liver (2) and the TGP Consortium in Japan (3) have taken on the challenge of systematically measuring, analyzing and understanding the toxic effects of chemicals on the human body, as well as developing new bioinformatics methods for mining large-scale toxicology data (4,5). However, due to legal, ethical and/or religious concerns that prohibit direct testing on the human body, chemical toxicity studies performed over the last few decades have had to use animals as human models. Unfortunately, animal models are not equivalent to the human system, and some drugs that have been approved for release after successful animal studies have yielded tragic results in humans, including the delayed occurrence of mental deficiencies or cancers. For example, thalidomide [3-(N-phthalimido) glutarimide], a well-known versatile drug used to treat such serious diseases as red nodules in Hansen's disease, multiple myeloma and HIV, was responsible for various forms of birth defects when it was administered to pregnant women, via mechanisms that remain unknown to this day (6). Methylmercury, a well-known chemical pollutant used since ancient times,

has poisoned tens of thousands of individuals in Japan ([7]). The toxicities of those two chemicals were very difficult to assess because they exerted weaker effects in mice than in humans, or different effects altogether. Therefore, in order to overcome the problems of animal test systems, we urgently need to develop a novel 'pseudo-human' system for drug testing.

In this regard, it is likely that toxicological studies based on various types of stem cells, including embryonic, induced pluripotent and mesenchymal stem cells, as well as induced differentiated cells, will facilitate the development of novel drugs that act on various diseases with improved efficiencies. A recent report showed that human and mouse embryonic stem (ES) cell-derived neural cells exhibited completely different responses to methylmercury due to differences in the underlying gene networks in the two species ([8]). Although the authors of that study clearly demonstrated the existence of chemical effects that were specific to the artificially but fully differentiated neural cells, it was not straightforward to reproduce the entire process of neuronal development, during which the effects of the chemicals were expected to be observed.

In this study, we developed an alternative approach that uses human ES cells (hESCs) without further differentiation. Our aim was to predict the 'late-onset' effects of toxic chemicals whose clinical symptoms appear months or years after exposure. Using only 10 genes pre-selected by multivariate statistical analysis of limited microarray data, we performed 9768 qRT-PCR measurements and extensive bioinformatics studies to predict the toxic effects of 20 categorized and two uncategorized chemicals. Using this system, we achieved as high as 97.5–100% prediction accuracies for three representative toxicity categories: neurotoxins (NTs), genotoxic carcinogens (GCs) and non-genotoxic carcinogens (NGCs). We also found that gene network information reinforced prediction accuracy when qRT-PCR data alone had little predictive power. The resultant gene networks exhibited some conserved patterns as well as variations among chemicals belonging to the same toxicity category, indicating that sub-categorical information that is not extracted from raw qRT-PCR data should be included in the gene network architecture.

MATERIALS AND METHODS

Toxicity tests on the hESC system

Ethics statement. The hESC line KhES-3 (XY karyotype) was provided by Dr Hirofumi Suemori, Research Center of Stem Cells, Institute for Frontier Medical Science, Kyoto University ([9]). All experiments using hESCs were approved by the ethics committees of the National Institute for Environmental Studies and The University of Tokyo, in accordance with the guidelines of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Selection of toxic chemicals. Twenty two chemicals were carefully selected according to the literature ([10,11]) from among three representative toxicity categories: 9 neurotoxins (NTs), 5 genotoxic carcinogens (GCs) and 6 non-genotoxic carcinogens (NGCs) and two uncategorized toxins. The NTs were valproic acid, cyclopamine,

phenytoin, methylmercury, acrylamide, 4-OH-2′,3,3′,4′,5′-PCB, 2,5-hexanedione, warfarin and thalidomide; the GCs were benzo[a]anthracene, 3-methylcholanthrene, benzo[a]pyrene, diethylnitrosamine and diethylstilbestrol; the NGCs were 2,3,7,8-Tetrachlorodibenzodioxin (TCDD), lithocholic acid, thioacetamide, butylated hydroxyanisole, methapyrilene hydrochloride and phenobarbital; and the uncategorized toxins were bisphenol-A and permethrin. To guide our selection process, we chose NTs, GCs and NGCs with well-established toxicities in humans (NTs: valproic acid, phenytoin, warfarin and thalidomide; GCs: benzo[a]pyrene, diethylnitrosamine and diethylstilbestrol; NGCs: TCDD) and/or well-established mechanisms (NTs: cyclopamine; GCs: benzo[a]pyrene and diethylstilbestrol; NGCs: TCDD). Other NTs were selected based on data from animal experiments and human epidemiology data ([12,13]), and other GCs and NGCs were selected based on U.S. and European international projects ([14–16]) and the International Agency for Research on Cancer monograph on human carcinogens (http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php). Uncategorized chemicals were selected from among widely used representative materials found in plastics and pesticides. All of the selected chemicals could be handled using normal laboratory equipment with sufficient care for safety. Detailed information of the 22 chemicals, including chemical abstract services registry numbers (CASRNs), suppliers and references, is shown in Supplementary Table S1.

Exposure system. Human KhES-3 ES cells were cultured on Matrigel^{TM} in mTeSR1 (a feeder-free maintenance medium for hESCs). As in conventional toxicology experiments ([17–19]), to avoid drawing only over-dosed responses, such as apoptosis, and to enable the detection of subtle gene expression changes that may happen in human embryos, hESCs were exposed to each of the above 22 chemicals (9 NTs, 5 GCs, 6 NGCs and two uncategorized toxins) at five doses (1, 1/2, 1/4, 1/8 and 1/16, where 1 indicates the highest dose that did not impair cell viability), and mRNA samples were taken at four time points (24, 48, 72 and 96 h after initiation of exposure), to yield 11 genes $\times$ 22 chemicals $\times$ 5 doses $\times$ 4 time points = 4840 measurements. As the negative control, mRNA levels were also measured at the same four time points in cells not exposed to any chemical, thus yielding 11 genes $\times$ 4 time points = 44 measurements. Then, we repeated all the measurements, thereby performing (4840 + 44) $\times$ 2 = 9768 measurements in total.

Correspondence analysis (CA) to narrow down candidate feature genes

Microarray analysis of 10 chemicals (valproic acid, cyclopamine, phenytoin, methylmercury, acrylamide, benzo[a]anthracene, 3-methylcholanthrene, benzo[a]pyrene, diethylnitrosamine and diethylstilbestrol) and one control (no exposure), for a total of 11 experimental conditions, was performed to select marker or feature genes for the gene network and toxicity prediction analysis. Feature genes constitute feature vectors in machine learning approaches; such genes are expected to both characterize and predict toxicity effects. Because we

assumed that the mechanisms of normal development are primarily controlled by transcription factors, the feature genes were selected from genes encoding transcription factors. In order to simply detect genes sensitive to chemicals, or candidate feature genes at the stem cell stage, we performed microarray analysis of undifferentiated KhES-3 ES cells rather than differentiated cells or embryos. The raw data obtained 48 h after exposure to 1/2 dose were log$_2$-transformed and quantile-normalized using the BeadArray (20) library of the Bioconductor (http://bioconductor.org/) package in the R (http://www.r-project.org/) statistical computing language. Among the 48 206 gene probes, 30 990 probes with signal detection *P*-value < 0.05 in at least one of the 11 experiments were analyzed further. Next, we performed a classical unsupervised multivariate statistical method called correspondence analysis (21) using the *corresp* function from the MASS (22) library in R on the matrix data of 11 samples, adopting 3492 probes, which correspond to genes encoding transcription factors. Importantly, this process does not use any toxicity categorical information and thus, it should not affect the subsequent toxicity predictions. We conditionally selected 10 candidate marker or feature genes that satisfied the following three criteria: (i) signal detection *P*-value < 0.1 for three or more samples; (ii) the encoded protein is a transcription factor; and (iii) within the top 20 highest (or lowest for negative) weights in each principal component. Using these three criteria, we narrowed down the initial 48 206 genes to 167 non-redundant candidate genes. However, the number was still large and thus might include spurious genes. Therefore, we added a fourth criterion that we considered to be the most important: (iv) abundant expression in hESCs, as determined by qRT-PCR analysis. Ultimately, we selected the following 10 genes: *NANOG*, *SOX2*, *DMTF1*, *ZNF208*, *ADRM1*, *TRIB1*, *CRY1*, *SMAD7*, *SMAD6* and *VHL1*. The top 20 genes with the highest (or lowest) weights in each principal component are shown in Supplementary Table S2. The microarray data are available on our website (http://stemcellinformatics.org/toxicology/) or the Gene Expression Omnibus (GSE60154).

qRT-PCR analysis of 22 chemicals

Using the 10 marker genes listed above, we performed qRT-PCR analysis of mRNAs in hESCs exposed to the 22 toxic chemicals. Details of the exposure doses and the toxicity categories are provided in Table 1. qRT-PCR experiments were performed according to a previously described protocol (8), and the primer sequences used in our analysis are shown in Supplementary Table S3. To calibrate gene expression levels from qRT-PCR measurements, we also measured the levels of β-actin (*ACTB*) mRNA as the internal standard; thus, a total of 11 genes were used in the qRT-PCR analysis, yielding 9768 measurements. Raw data from the qRT-PCR experiments and the sample preparation information are available at our website (http://stemcellinformatics.org/toxicology/).

Preprocessing and data normalization

The raw qRT-PCR data were processed using the limma (23) library of the Bioconductor package. Normalization

was performed in two steps. First, we estimated batch effects in standard 96-well plates using the empirical Bayesian linear model provided in the limma library based on the following equation:

$$g_i = \sum_j a_{i,j} G_{i,j} + \sum_k b_{i,k} B_{i,k} + e_i, \tag{1}$$

where i, g_i, $G_{i,j}$, $B_{i,k}$ and $e_i(\sim N(0, \sigma_i^2))$ indicate gene, expression value, chemical effect, batch effect and noise, respectively. For each of the 11 genes at four different time points (thus, i ranges from 1 to 44), the function g_i consists of j (22 chemicals + 1 control) and k (10 plates) terms, whereas only one each of $a_{i,j}$ and $b_{i,k}$ has non-zero values. Next, we estimated the linear function g_i by empirical Bayesian linear fitting (the *lmfit* function in limma) over the 222 data points, which consisted of two biological replicates of 22 chemicals multiplied by 5 doses and 1 control (2 × (22 × 5 + 1) = 222). Once $a_{i,j}$ and $b_{i,k}$ were estimated, we subtracted the batch effect terms to leave only the chemical effects. Finally, the batch-effect-adjusted mRNA levels were calculated by dividing them by the level of the internal standard (*ACTB*), and the adjusted values were used for further analysis.

Bayesian gene network analysis

Many gene network reconstruction methods are available. We tested structure equation modelings (SEMs), graphical Gaussian modelings (GGMs), cross correlation coefficients (CCCs) and Bayesian networks (BNs) in our previous study (24), and found no large differences. Thus, we simply used BNs as they are one of the sophisticated network inference methods used for analyzing biological data (25).

We reconstructed BNs from the qRT-PCR data of each of the 22 chemical exposure experiments. By 'reconstruction of gene networks,' we mean inference of the directed edge weights for a complete graph of $q \times q$ edges connecting q nodes, or genes; in this study, $q = 10$. Due to statistical limitations, BNs are characterized by the well-known constraint that they generate only non-cyclic (acyclic) graphs. For BN reconstruction, we used the previously described inference algorithm TAO-Gen (26,27), which was developed using the Gibbs sampling method. As the details of the original TAO-Gen algorithm can be downloaded from our website (http://stemcellinformatics.org/toxicology/), we describe only the additional improved algorithm here. In order to achieve more efficient and reliable sampling, we further improved this algorithm by creating a version optimized for parallel computation, based on standard ensemble Markov chain Monte Carlo (MCMC) sampling or replica-exchange MCMC sampling (28,29). The improved algorithm, called 'RX-TAOgen,' can run multiple Gibbs sampling series and exchange sampling parameters between neighbor series as follows: It exchanges the l and $l + 1$ series of different parameters C_l, only if the ratio of the posterior probabilities of BNs, which are calculated by function *f* in the original TAO-Gen, is greater than some threshold. The ratio is given by the following equation:

$$r = \frac{f(T_l, B_l, S_l|g; C_{l+1})}{f(T_l, B_l, S_l|g; C_l)} \times \frac{f(T_{l+1}, B_{l+1}, S_{l+1}|g; C_l)}{f(T_{l+1}, B_{l+1}, S_{l+1}|g; C_{l+1})}, \tag{2}$$

Table 1. Complete list of the 22 toxic chemicals and doses used in the exposure experiments

Compound Name	Dose (nM)						Toxicity Category
Valproic acid	1000	500	250	125	62.5	0	Neurotoxin
Cyclopamine	1000	500	250	125	62.5	0	(NT)
Phenytoin	1000	500	250	125	62.5	0	
Methylmercury	50	25	12.5	6.25	3.125	0	
Acrylamide	5000	2500	1250	625	312.5	0	
4-OH-2′,3,3′,4′,5′-PCB107	10	5	2.5	1.25	0.625	0	
2,5-Hexanedione	1000	500	250	125	62.5	0	
Warfarin	1000	500	250	125	62.5	0	
Thalidomide	500	250	125	62.5	31.25	0	
Benzo[a]anthrene	50	25	12.5	6.25	3.125	0	Genotoxic carcinogen
3-Methylcholanthrene	500	250	125	62.5	31.25	0	(GC)
Benzo[a]pyrene	500	250	125	62.5	31.25	0	
Diethylnitrosamine	500	250	125	62.5	31.25	0	
Diethylstilbestrol	500	250	125	62.5	31.25	0	
2,3,7,8-TCDD	10	5	2.5	1.25	0.625	0	Non-genotoxic carcinogen
Lithocholic acid	1000	500	250	125	62.5	0	(NGC)
Thioacetamide	1000	500	250	125	62.5	0	
Butylated hydroxyanisole	1000	500	250	125	62.5	0	
Methapyrilene hydrochloride	1000	500	250	125	62.5	0	
Phenobarbital	1000	500	250	125	62.5	0	
Bisphenol-A	1000	500	250	125	62.5	0	Uncategorized
Permethrin	1000	500	250	125	62.5	0	

where T_l and B_l are $q \times q$ transition and weight matrices, respectively, S_l is a vector of the standard deviations of gene expression values for q genes, and g is the observed gene expression data. For the threshold, we used random values in [0, 1]. A total of $10 \times 10 = 100$ gene-to-gene weights ($w_{i,j}$ in Figure 1) on network edges were obtained and used as additional features for the support vector machines (SVM). Once the 10×10 edge weights for each chemical were obtained, they were added to the qRT-PCR data for the corresponding chemicals; thus, the input data were increased from 10 genes $\times$ 5 doses $\times$ 4 time points = 200 values to $200 + (10 \times 10)$ genes = 300 values in the toxicity prediction system. The binary program and a detailed description of RX-TAOgen can be downloaded from our website (http://stemcellinformatics.org/toxicology/).

Toxicity prediction by SVM

There exist a variety of statistical or machine learning methods, such as neural networks, partial least squares, random forests, etc., which can be applied to supervised toxicity prediction. In this study, we did not focus on the comparison of existing methods; rather, we focused on whether or not the pseudo-human system based on hESCs has sufficient potential in general to predict late-onset chemical toxicities. Thus, we simply used SVMs (30) as one of the state-of-the-art learning methods (31). SVMs are widely used for the classification of samples on the basis of their input (gene expression) values (32), and it has been reported that their classification performance is not significantly different from random forests (33). In its basic form, an SVM is a binary classifier based on a hyper-plane that distinguishes two distributions of M-dimensional vectors or samples from different classes with a linear or non-linear similarity function called kernel matrix. As the principles of SVMs are found in many papers and books, we omit the details in this paper.

Kernel types and leave-one-chemical-out-prediction (LOCOP). According to our previous study (34), here we used six (linear, polynomial, RBF, EKM, Saigo and ME) kernels. Three of these (linear, polynomial and RBF) constitute the widely used *vectorial* kernel family, and the three others (EKM, Saigo and ME) constitute the *structured* kernel family (35). In the prediction of complex data, the structured kernel family often outperforms the vectorial kernel family. For each of the 22 chemicals, a pair of repeat qRT-PCR measurements was performed; consequently, a total of 44 data points were generated. To carefully evaluate the prediction accuracy for new chemicals, we performed not conventional leave-one-out-prediction (LOOP) but the leave-one-chemical-out-prediction (LOCOP) test, in which a pair of repeat data for each of the 22 chemicals was eliminated during the learning process, and then used to test the prediction.

Feature selection and addition of gene network edge weights. As shown in Figure 1, we used two different types of features for SVM predictions: feature 1, which denotes the raw qRT-PCR data consisting of 200 measurements, and feature 2, which denotes BN edge data consisting of 100 weight values. Note that feature 1 is raw gene expression data, whereas feature 2 is computer-inferred gene-to-gene interaction data derived from the raw gene expression data. To optimize the number of features (qRT-PCR data, BN edge weights or both) for input into the SVM, we used standard two-sample t-statistics in each iteration of the LOCOP cross-validation tests in order to rank the features according to their discriminative potential or P-values. Thus, the six kernels were tested with SVMs in order to analyze their classification power with various numbers of features, and their prediction accuracies were recorded for each number of features. More specifically, we initially ranked 200, 100 and 300 features for qRT-PCRs, BNs and qRT-PCRs + BNs, respectively and then trained the SVM with 21 chemicals and tested the 22nd chemical (two repeats for each) with various numbers of the top n features with n ranging from 1 to
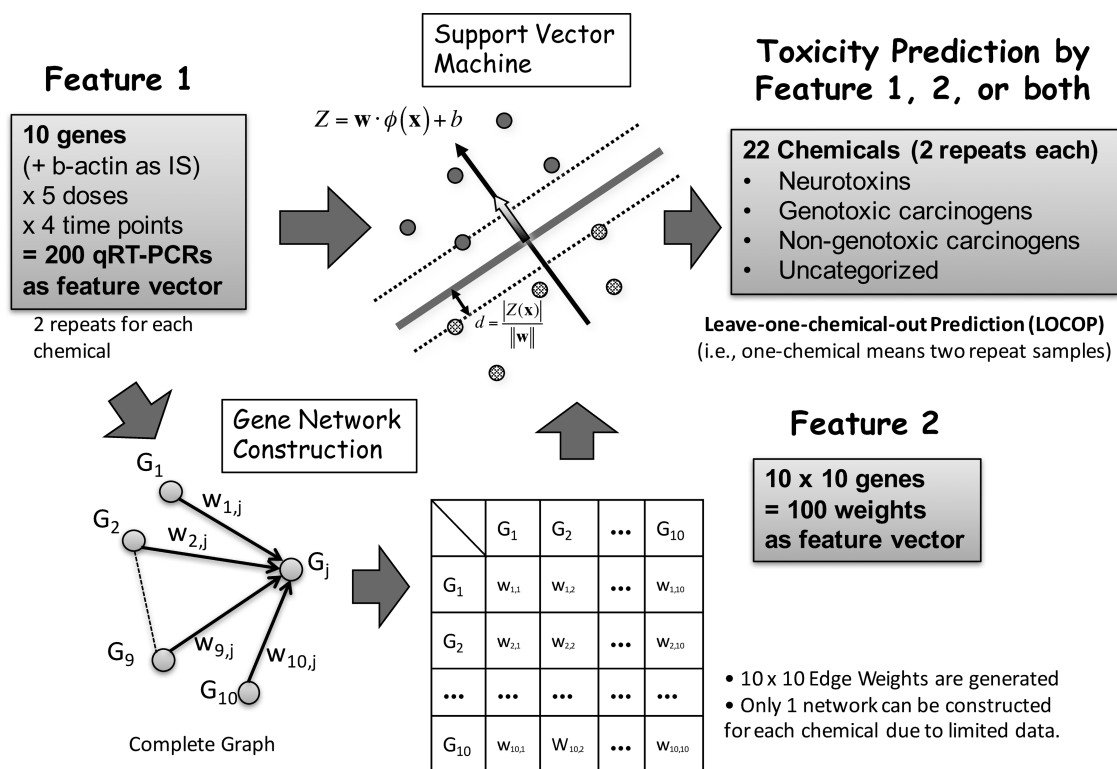
Figure 1. Schematic view of processes from data generation to SVM prediction. For each of the 22 toxic chemicals, 200 qRT-PCR measurements of 10 genes were conducted on cells treated with five doses at four time points. In addition, 100 BN data from 10×10 genes were generated. qRT-PCR data, BN data or both data types were used to train SVMs and predict each of the three toxicity categories.

the maximum number: 200 (qRT-PCRs), 100 (BNs) and 300 (qRT-PCRs + BNs).

Performance measure. For all the six kernels tested, seven SVM parameter values ($C = 10^{-3}$, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3) were tested. Additionally, for the polynomial kernel, 10 values ($D = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$) were tested, and for the RBF, EKM and Saigo kernels, 11 values ($\sigma = 10^{-10}$, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1) were tested. For the ME kernel, we tested parameters in the range $G = 2^{-5}$, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1, 2, 2^2, 2^3, 2^4, 2^5. Note that the number of parameter combinations in the ME kernel is equal to those in the RBF, EKM and Saigo kernels used in this study. For convenience, all accuracies reported in this paper were calculated using the following formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (3)$$

where *TP*, *FP*, *TN* and *FN* are true positive, false positive, true negative and false negative frequencies, respectively. We also performed area under the curve (AUC) calculations using the pROC library (36) in R. No smoothing method was applied to the receiver operator characteristic (ROC) curve.

RESULTS

Previously, we used neural cells differentiated from hESCs (24). In that study, we showed that when qRT-PCR data and BN edge weights were given as the input data of SVMs,

we were able to classify 15 chemicals as NTs or carcinogens (including GCs and NGCs) with prediction accuracies of 93.3% and 100.0%, respectively. In that case, the marker genes used as features for SVM predictions were pre-selected by an expert.

Here, we investigated difficult-to-predict late-onset chemical toxicities using only hESCs not subjected to further differentiation. In the experiments, the number of chemicals was increased from 15 to 22. In addition, the marker genes were carefully selected from microarray data obtained from cells treated with 10 toxic chemicals and one control, using classical unsupervised multivariate statistics (correspondence analysis, CA) and a filter based on four criteria. Ultimately, we selected 10 genes as qRT-PCR targets: *NANOG*, *SOX2*, *DMTF1*, *ZNF208*, *ADRM1*, *TRIB1*, *CRY1*, *SMAD7*, *SMAD6* and *VHL1*. The CA weights and the gene descriptions are shown in Table 2. The full list of chemicals and the doses used in the exposure experiments are shown in Table 1. The prediction scheme is shown in Figure 1.

qRT-PCR results revealed little dose dependence

The 9768 qRT-PCR measurements gave important general features of genes with regard to dose responses. When we checked the general tendencies of mRNA abundances, among the selected 10 marker genes, two genes, *DMTF1* and *SMAD6*, were undetectable in a large number of experiments, probably due to their low expressions. *SMAD7* was also undetectable in several experiments. Thus, most of

Table 2. Ten genes selected by correspondence analysis (CA) for 9768 qRT-PCR data

Gene Symbol	CA Component (Weight)	Definition*	Synonym
ADRM1	1 (1.70)	*Homo sapiens* adhesion regulating molecule 1 (ADRM1), transcript variant 2, mRNA.	Rpn13; GP110; MGC29536
CRY1	4 (2.47)	*Homo sapiens* cryptochrome 1 (photolyase-like) (CRY1), mRNA.	PHLL1
DMTF1	6 (−2.44)	*Homo sapiens* cyclin D binding myb-like transcription factor 1 (DMTF1), mRNA.	DMP1; DMTF; hDMP1; FLJ41265
NANOG	5 (3.03)	*Homo sapiens* Nanog homeobox (NANOG), mRNA.	–
SMAD6	6 (3.68), 7 (−2.88), 8 (2.72), 9 (−2.49)	*Homo sapiens* SMAD family member 6 (SMAD6), transcript variant 1, mRNA.	MADH6; HsT17432; MADH7
SMAD7	1 (1.66)	*Homo sapiens* SMAD family member 7 (SMAD7), mRNA.	FLJ16482; MADH8; MADH7
SOX2	1 (1.64), 4 (2.19)	*Homo sapiens* SRY (sex determining region Y)-box 2 (SOX2), mRNA.	ANOP3; MGC2413; MCOPS3
TRIB1	1 (−3.25), 2 (−2.77)	*Homo sapiens* tribbles homolog 1 (*Drosophila*) (TRIB1), mRNA.	GIG2; SKIP1; C8FW
VHL1	1(1.77), 2 (−2.72), 4 (2.49)	*Homo sapiens* von Hippel-Lindau tumor suppressor (VHL), transcript variant 2, mRNA.	HRCA1; RCA1; VHL1
ZNF208	6 (−2.54), 10 (−3.71)	*Homo sapiens* zinc finger protein 208 (ZNF208), mRNA.	ZNF95; PMIDP

*The information in the 'Definition" column is an exact copy of data in the microarray descriptions published by Illumina.

the results of our statistical analysis could be, in principle, a reflection of the remaining seven genes. To investigate the global trends in gene expression in response to the 22 chemicals, we constructed a heat map of the Pearson correlation coefficients of the qRT-PCR data (Figure 2). For each of the 22 chemicals, the Pearson correlations between two chemicals administered at five doses were distinctively plotted on a color scale. Time-dependent values from two repeat experiments were averaged within the repeats and plotted as a single correlation coefficient. This plot revealed two important observations: first, there were no remarkable changes among the five doses used for each chemical, although the 1/16 dose was sometimes weakly correlated with the other doses; and second, the Pearson correlations (ranging from r = 0.778 to 0.878) did not differ significantly between chemicals of the same category and those of different categories. It should be noted, however, that for each of the three categories, the average correlations within the same category were always stronger than those among different categories.

Analysis of gene network patterns inferred from chemical exposure experiments

We inferred BNs consisting of 10 genes from the qRT-PCR data for all 20 chemicals. Each chemical exposure experiment generated only 40 data points (4 time points × 5 doses × 2 repeats = 40). Due to the limited amount of data, RX-TAOgen could infer only one BN for two repeat experiments. For each chemical, we ran RX-TAOgen in eight parallel settings with 20 000 exchanges, each executed at 10 Gibbs samplings, and took only the best network result (i.e. the one that yielded the highest likelihood) for further SVM predictions. For each chemical, this calculation took 10–15 min on a 2.93 GHz Intel X5570 CPU. The reconstructed BN structures for the 20 chemicals are shown in Supplementary Figure S1. To characterize those networks, we first performed topological analysis of the network structures. Figure 3 shows the summarized circular network diagrams displaying all the network edges that are high-weighted by any of the toxic chemicals in each of the three toxicity categories. Red edges with triangular tips indicate positive influ-

ences, whereas blue edges with diamond tips indicate negative ones. The number of colored edges between the same two nodes corresponds to the number of supported chemicals. Whereas the diagram of NTs contained a large number of colored edges representing high weights, that of NGCs exhibited a small number of edges, since NTs had higher variations than NGCs in the gene network patterns among the chemicals. It is noteworthy that GCs contained a smaller number of chemicals but showed more variations in the gene network patterns than NGCs.

BN information differs from raw qRT-PCR data

The gene network variations were revealed more clearly in the Pearson correlation analysis. Figure 4 shows heat maps of the Pearson correlation coefficients of (i) dose-averaged qRT-PCR data (Figure 4A) and (ii) the 10 × 10 = 100 edge weights in the reconstructed BNs for the 20 chemicals (Figure 4B). Comparing the heat maps of the qRT-PCRs and the BNs, we observed differences between the two Pearson correlations, indicating that the BNs provide different information from qRT-PCR data. To clarify which data contributed more, we generated a subtractive heat map, in which Pearson correlations of BNs subtracted by those of qRT-PCR for each pair of chemicals after normalization are shown in red-blue colors. In Figure 4C, Pearson correlations where BNs are more positive than qRT-PCRs are indicated by red, and those where qRT-PCRs are more positive than BNs are indicated by blue. For example, whereas the Pearson correlations for methylmercury, warfarin and thalidomide against other chemicals are more positive in qRT-PCRs, those for valproic acid and lithocholic acid are more positive in BNs. This analysis indicates the importance of using both BN and qRT-PCR data to boost the toxicity predictions, because together they could compensate for the lack of information by either method alone.

Predictions of chemical toxicity using embryonic stem cells

Using the above 10 genes, we trained SVMs for the 3 toxicity categories by supervised learning except for the ME
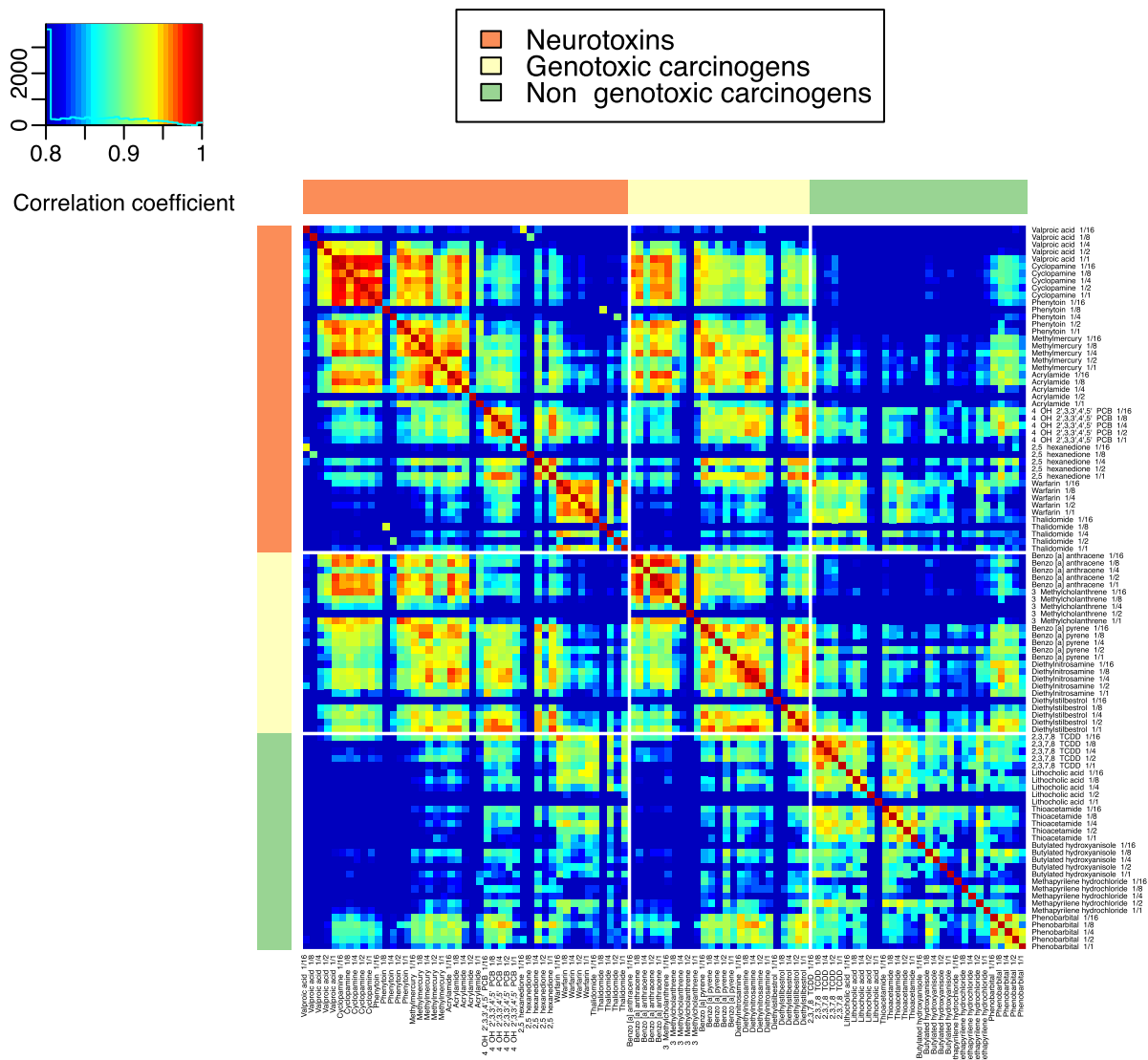
Figure 2. Pearson correlation coefficients in 100 toxic chemical experiments, constructed based on batch-adjusted qRT-PCR data. Pairwise Pearson correlation coefficients were calculated for 20 toxic chemicals administered at five doses (100 experiments).

kernel, using all the chemicals except for a set of test chemicals. The ME kernel requires the generation of kernel matrix by only a semi-supervised setting (37), i.e. learning all the chemicals including test chemicals that have no class labels, due to its intrinsic mechanisms. The two uncategorized chemicals, bisphenol-A and permethrin, were also included during the training iterations in the ME kernel. All data were assigned a class label of +1 or −1 for positive and negative data, respectively or 0 for uncategorized and test data. The uncategorized chemicals were not included in the accuracy and AUC calculations in the LOCOP tests. A total of 200, 100 and 300 feature values (see Materials and Methods) for qRT-PCR, BN and qRT-PCR+BN data, respectively, were re-ranked by *P*-values derived from the standard two-sample *t*-tests performed on the training data in each iteration of the LOCOP cross-validation tests, and the number of features from the top of the list was gradually increased. Table 3 summarizes the highest SVM predic-

tion accuracies and their AUCs for various combinations of parameters and numbers of features for qRT-PCR, BN and qRT-PCR+BN input data in each category. The upper table is for non-adjusted data, and the lower table is for batch-adjusted qRT-PCR data. Prediction accuracies by vectorial and structured kernels are separately shown in two lines. ROC curves with AUCs corresponding to the Table are shown in Supplementary Figure S2.

Even with qRT-PCR data alone, the prediction was generally successful due to the high performance of the kernel SVM method itself; the accuracies ranged from 77.5 to 100.0%, and were significantly higher than those of random predictions, i.e. 55.0–75.0%. Surprisingly, the use of BN data as SVM input tended to increase the prediction accuracies to as high as 95.0–100.0%. For NT prediction using batch-adjusted data, the accuracies were improved even more when qRT-PCR data and BN data were combined for use as SVM input: the resultant accuracies
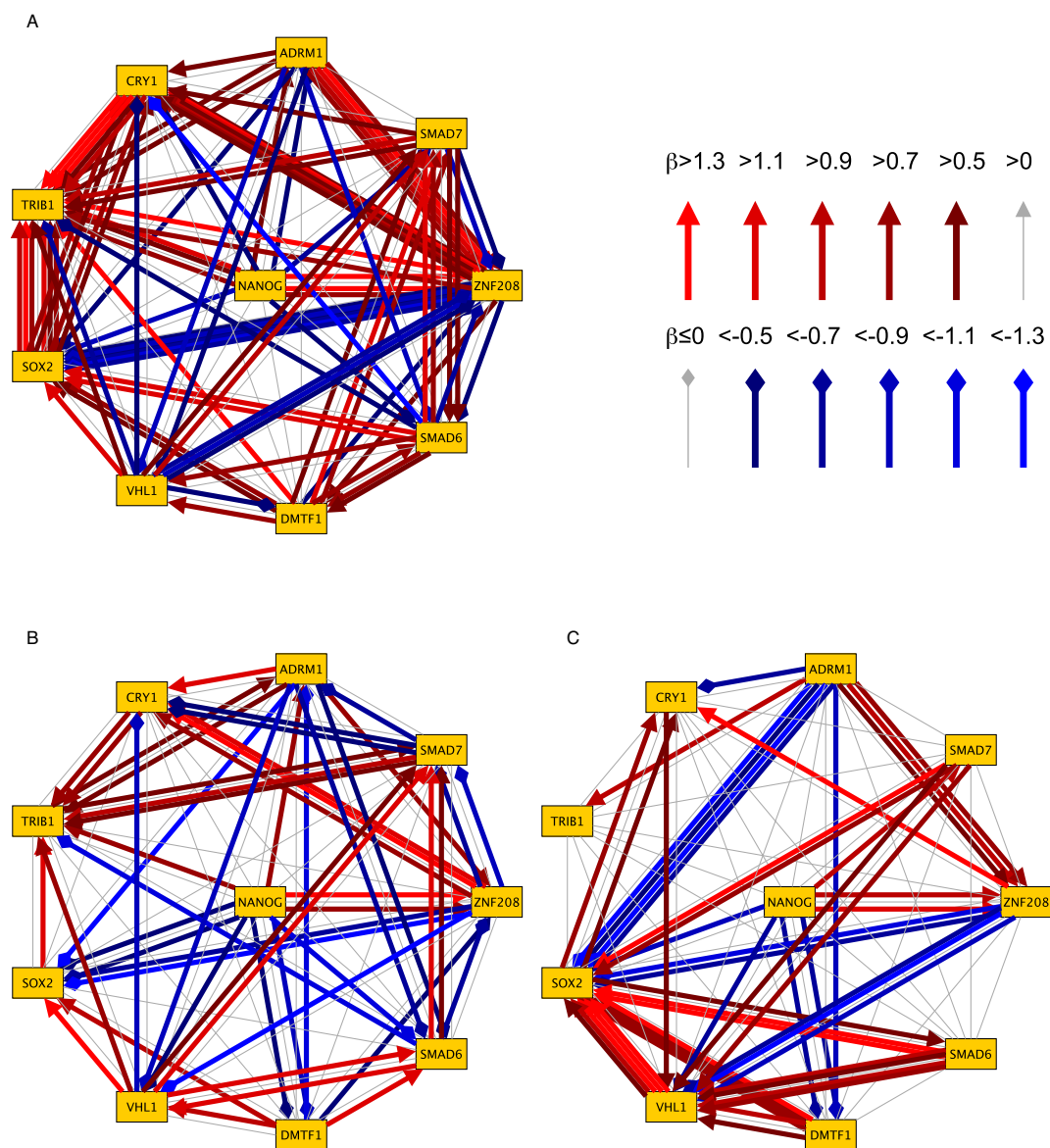
Figure 3. Category-dependent high-weighted edges of BN structures in circular network diagrams. (**A**) Neurotoxins (9 data), (**B**) genotoxic carcinogens (5 data) and (**C**) non-genotoxic carcinogens (6 data). The network for non-genotoxic carcinogens represents limited edges, indicating that the network may be specific and similar among different chemicals and thus easy to predict, whereas that for neurotoxins contains almost all edges, indicating that the network is not easy to characterize and thus difficult to predict.

Table 3. Summary of toxicity category predictions

Toxicity Category	#Data*	Random**	qRT-PCR***	BN***	qRT-PCR+BN***
			Upper: Vectorial Kernel (linear, polynomial and RBF)/Lower: Structured Kernel (EKM, Saigo and ME)		
Non-adjusted (Raw) qRT-PCR					
Neurotoxin	18/40	55.00	70.0 (0.664)	80.0 (0.758)	70.0 (0.697)
			77.5 (0.773)	**95.0 (0.889)**	82.5 (0.881)
Genotoxic carcinogen	10/40	75.00	82.5 (0.553)	90.0 (0.893)	82.5 (0.863)
			82.5 (0.777)	**100.0 (1.000)**	97.5 (0.980)
Non-genotoxic carcinogen	12/40	70.00	**100.0 (1.000)**	90.0 (0.929)	**100.0 (1.000)**
			100.0 (1.000)	95.0 (0.964)	**100.0 (1.000)**
Batch-adjusted qRT-PCR					
Neurotoxin	18/40	55.00	70.0 (0.705)	90.0 (0.859)	87.5 (0.919)
			80.0 (0.874)	95.0 (0.980)	**97.5 (0.998)**
Genotoxic carcinogen	10/40	75.00	85.0 (0.800)	85.0 (0.640)	82.5 (0.793)
			85.0 (0.830)	**100.0 (1.000)**	92.5 (0.940)
Non-genotoxic carcinogen	12/40	70.00	**100.0 (1.000)**	90.0 (0.964)	97.5 **(1.000)**
			100.0 (1.000)	95.0 (0.929)	97.5 **(1.000)**

*#Data: X/Y, where X = number of samples in the indicated category, and Y = total number of samples

**Random (prior) = max $\{Y/X, (40 - Y)/X\} \times 100$, with X and Y defined as in "#Data"

***Values for qRT-PCR, BN and qRT-PCR+BN represent maximum accuracy and AUC (in parentheses) in each toxicity category.
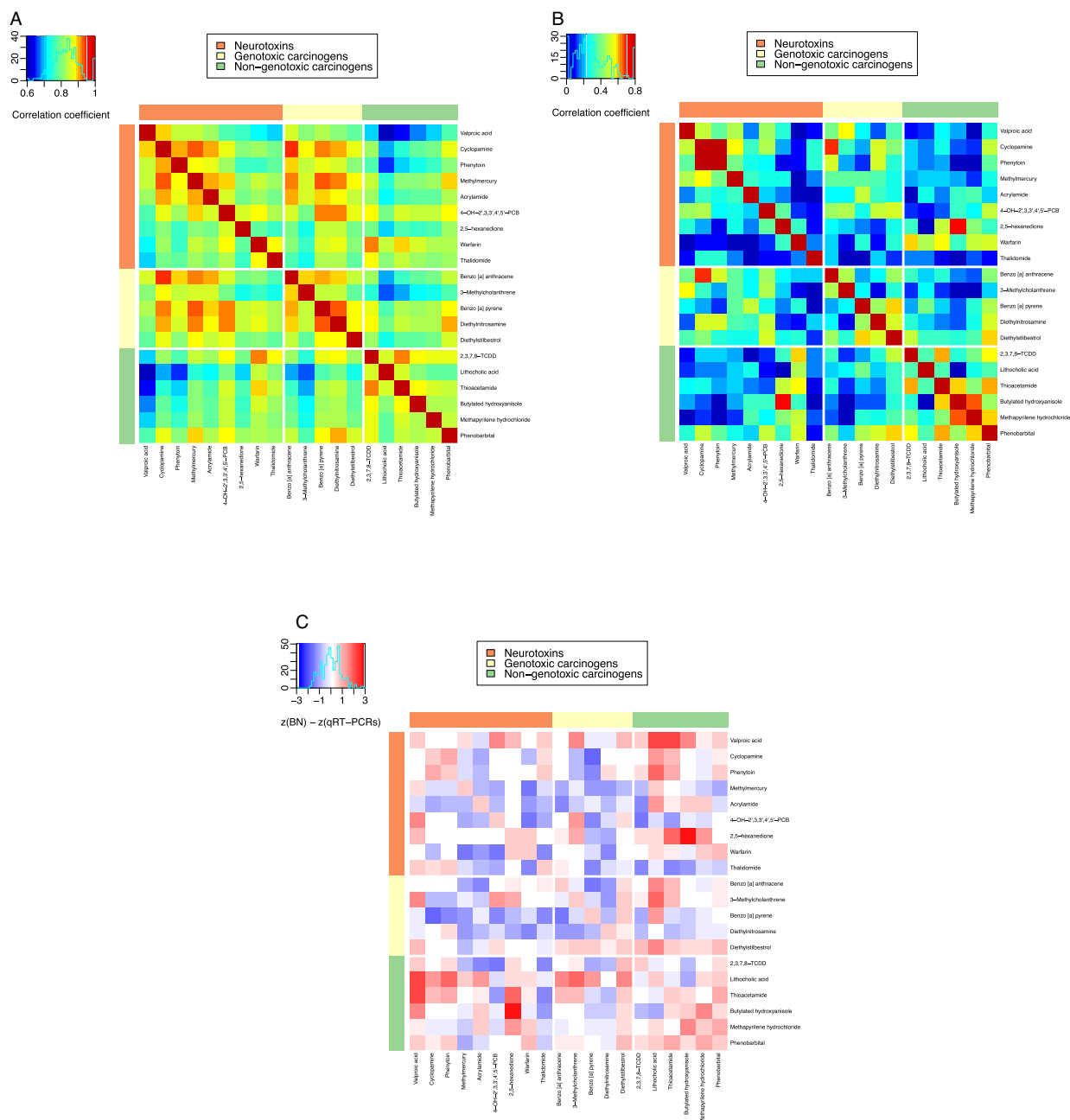
Figure 4. Analysis of BN information by Pearson correlation coefficients. Heat maps of pairwise Pearson correlation coefficients of the 20 toxic chemicals constructed based on (**A**) five dose-averaged qRT-PCR Pearson correlations and (**B**) 10×10 BN edge weights, where two experimental repeats were averaged for use in the calculation, are shown. (**C**) The qRT-PCR and BN contributions of the Pearson correlation coefficients calculated by, $z(|BN|) - z(|qRT\text{-}PCR|)$, where z is Gaussian normalization, are indicated by colors: red for BN and blue for qRT-PCR contributions.

were 80.0% (qRT-PCR), 95.0% (BN) and 97.5% (qRT-PCR+BN). Furthermore, BN data alone achieved perfect accuracy (100.0%) in GC prediction. In NT and GC predictions using qRT-PCR data alone, accuracies were higher when the analyses were performed on batch-adjusted data; by contrast, BN predictions often yielded similar accuracies in both batch-adjusted and non-adjusted data. These results indicate that BN predictions may be more robust than qRT-PCR predictions, perhaps because the gene-to-gene interac-

tion information is not significantly affected by noise caused by batch effects.

To summarize, if qRT-PCR data yield accuracies lower than 100.0%, BN (as well as a combination of qRT-PCR+BN in some cases) can increase accuracy in all predictions; therefore, it is advantageous to add BN to qRT-PCR as input data. Conversely, if BN alone achieves high predictive power (such as 100.0% in the GC case), the prediction accuracy is sometimes decreased when BN is used in combination with qRT-PCR, probably because the latter

method provides noisy and low-level information that prevents SVMs from producing better discriminant functions.

Effective genes, hours and doses for SVM prediction

Figure 5 shows the effects of the number of features on the prediction accuracy. Peaks were observed for the combined data of NTs and GCs. For NTs, the highest accuracies were obtained at around 170 features, whereas for GCs, two small peaks were observed at around 50 and 250 features. For NGCs, no distinct peak was observed; nevertheless, the accuracy was consistently high for all numbers of features. When we examined the top 100 of 300 features (ranked by the two-sample t-test) in LOCOP iterations performed on the combined data, the choice frequencies of features at 96 h were the lowest among the four time points (50 features for each): 12.14, 13.95 and 9.82 times on average for NTs, GCs and NGCs, respectively, whereas the randomly expected choice frequencies should be $50/300 \times 100 = 16.67$. This result indicates that the 96-h data were segregated from the top candidate features, probably due to low data quality caused by sample degradation resulting from long-term toxic chemical exposure. Furthermore, no choice frequency preference was observed for the rest of the time points (24-, 48- or 72-h) in the top 100 features among the three toxicity categories.

Typing of gene networks inferred from toxic chemical experiments

Detailed analysis of the SVM predictions in each toxicity category may give us a better understanding of the characteristics of the chemicals used in this study. Therefore, we determined which chemicals were difficult to predict, and which chemicals exhibited improved predictions in combination with gene networks. In the case of NT prediction, 11 chemicals for which qRT-PCR data only prediction failed (accuracy: 80.0%) were noted: cyclopamine (NT), benzo[a]anthracene (GC), 3-methylcholanthrene (GC), methylmercury (NT), benzo[a]pyrene (GC), diethylstilbestrol (GC), 4-OH-2′,3,3′,4′,5′-PCB (NT), 2,5-hexanedione (NT), warfarin (NT), 2,3,7,8-TCDD (NGC) and phenobarbital (NGC). After BNs were incorporated, the predictions were dramaticaly improved and only one chemical, acrylamide (NT), failed to be predicted (accuracy: 97.5%), representing a new failure of the combined prediction method.

In the case of GC prediciton, qRT-PCR data only prediction (accuracy: 85.0%) failed for as many as 10 types of chemicals, including valproic acid (NT), benzo[a]anthracene (GC), 3-methylcholanthrene (GC), methylmercury (NT), acrylamide (NT), benzo[a]pyrene (GC), diethylnitrosamine (GC), diethylstilbestrol (GC), 2,5-hexanedione (NT) and methapyrilene hydrochloride (NGC). That number was reduced to two in qRT-PCR+BN prediction (accuracy: 92.5%): 3-methylcholanthrene (GC) and diethylnitrosamine (GC). Thus, we conclude that these two GC chemicals are difficult to predict as GCs, probably due to their unique gene response and BN architecture (see Supplementary Figure S1).

DISCUSSION

In this study, we attempted to predict late-onset chemical toxicities using only an undifferentiated hESC system. The results indicated that this system is sufficiently powerful to predict what will happen during the later stages of development when using SVM with gene networks in which network edge weights are given as input feature vectors.

Comparisons with QSAR

As the most popular toxic chemical prediction method, the quantitative structure-activity relationship (QSAR) model is often used, assuming that chemical structures are intrinsically related to specific biological activities (38–40). However, the QSAR model is plagued by several problems, including: (i) the inability to directly measure cellular activities, (ii) difficulty in predicting racemic chemicals, such as thalidomide, (iii) dependence of prediction accuracies on how molecular descriptors are set, and so on (41). Recent QSAR-based approaches that use such molecular descriptors have been improved with the introduction of machine learning techniques (42–45). However, even the most recent approaches can handle only analogous chemicals or their derivatives due to intrinsic difficulties in the prediction of active chemicals that have no or little commonality of the structural framework or functional groups to the learning chemicals. This is exactly the case in our selected chemicals, as shown in Supplementary Table S1.

Alternatively, our gene expression-based approach directly measures biological activities and is thus expected to have great predictive power for various chemicals, even if learning data for similar transcriptional (network) reactions are available. Moreover, the QSAR model basically does not consider conditional cellular activities, such as developmental-stage-specific reactions. Thus, our qRT-PCR measurements based on the hESC system should be more accurate because they reflect initial as well as complex molecular reactions in actual developmental processes.

As a simple proof of study, we generated 1665 molecular descriptors for our selected chemicals using E-Dragon (46) website (http://www.vcclab.org/lab/edragon/) and predicted chemical toxicity categories with our SVM using the same scheme as BNs. We exhaustively tested the number of feature descriptors ranked by two-sample t-test (see Materials and Methods) from 1 to the full range of 1665, which is more than 16 times larger than the range of BNs that have only 100 edge weight parameters. As a result, the highest accuracies for NTs, GCs, and NGCs were as low as 75.0%, 95.0% and 80.0%, respectively, whereas those were 95.0%, 100.0% and 95.0%, respectively, for BN only predictions, as shown in Table 3.

Prediction of uncategorized chemicals

One of the most intriguing issues surrounding this high-accuracy prediction method is whether it is applicable to new toxic chemicals. We were limited by cost and time due to the large number (9768) of qRT-PCR experiments performed on hESCs. Nonetheless, we incorporated two uncategorized chemicals, bisphenol-A and permethrin, into
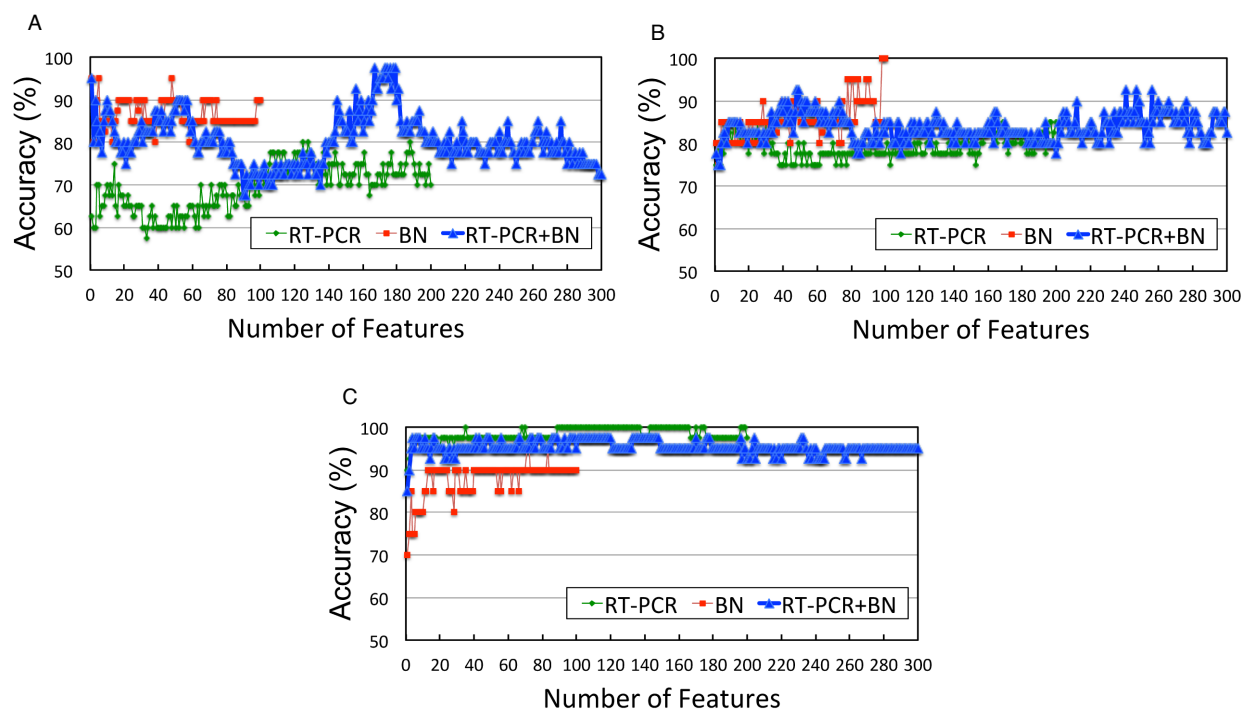
Figure 5. Maximum accuracies for various numbers of features. (**A**) Neurotoxins (9 data), (**B**) genotoxic carcinogens (5 data) and (**C**) non-genotoxic carcinogens (6 data). qRT-PCR only (200 features), BN only (100 features) and qRT-PCR + BN (300 features) are shown by different lines. In the qRT-PCR + BN lines, NTs and GCs exhibit peak accuracies at specific feature numbers, whereas NGCs constantly exhibit 90.0% or higher accuracy.

our SVM system with an eye to predicting their potential toxic effects. Bisphenol-A is a weak environmental hormone, and is suspected to be potentially toxic as an NT and/or an NGC. Several studies have shown that bisphenol-A prevents synaptogenic response to estradiol (47). On the other hand, due to its estrogenicity, bisphenol-A can also cause neural disorders or early maturation in animal experiments (48). In addition, developmental exposure to bisphenol-A increases susceptibility to prostate carcinogenesis via unknown epigenetic mechanisms (49). Permethrin is a pyrethroid that is widely used as an insecticide; it also functions as an NT in mice by decreasing Ca^{2+} influx into neurons and repressing mRNA expression of activity-dependent c-fos and brain-derived neurotrophic factor (BDNF) (50). Neonatal exposure to permethrin leads to long-lasting effects, causing changes in behaviors and striatal monoamine levels and increasing oxidative stress (51).

Figure 6 shows chemicals with the highest similarity to bisphenol-A and permethrin at the BN-structure level. Interestingly, bisphenol-A has marked similarity to 2,3,7,8-TCDD (an NGC; r = 0.713), whereas permethrin has moderate similarity to both 4-OH-2′,3,3′,4′,5′-PCB (an NT; r = 0.505) and methapyrilene hydrochloride (an NGC; r = 0.456). When we investigated the SVM prediction results using both qRT-PCR data and BNs at their maximum accuracies, bisphenol-A was predicted to have an NGC effect, and permethrin was predicted to have both NT and NGC effects. GC effects were not predicted for either chemical.

Several recent papers have described the relationship between bisphenol-A and its carcinogenicity: specifically,

fetal exposure to bisphenol-A induces preneoplastic and neoplastic lesions in adult rat mammary gland (52), and bisphenol-A induces human ovarian cancer growth (53). Other reports have described the relationships between permethrin and neural disorders. For example, prenatal exposure to permethrin influences vascular development in mouse fetal brain (54). To our knowledge, however, no previous report has provided solid evidence of a relationship between permethrin and NGC toxicity. Because the underlying mechanisms of the suspected toxicities of these two uncategorized chemicals remain unknown, additional investigations are necessary to determine their toxicity types.

Further improvement and implications for kernel methods

In this study, we used the two-sample *t*-test to rank 300 features and observed the tendency for the 96-h qRT-PCR data to be segregated from the top 100 features. Our method of selecting features is often called the 'filters method.' There also exist other selection methods, such as wrappers, embedded selection and others (55). Although not the primary focus of this study, the use of recursive feature elimination (RFE) might have enabled us to effectively eliminate non-informative features (56).

It is also notable that the method described here, which provides BN structured data for input as features, offers new insight into research fields that use kernel methods. The use of gene networks, or any other data reconstructed from original raw data, as the input data for an SVM may be considered an *external kernel*, which sometimes yields explicit models and thus provides more discriminant power in non-linear and/or conceptualized data space than raw
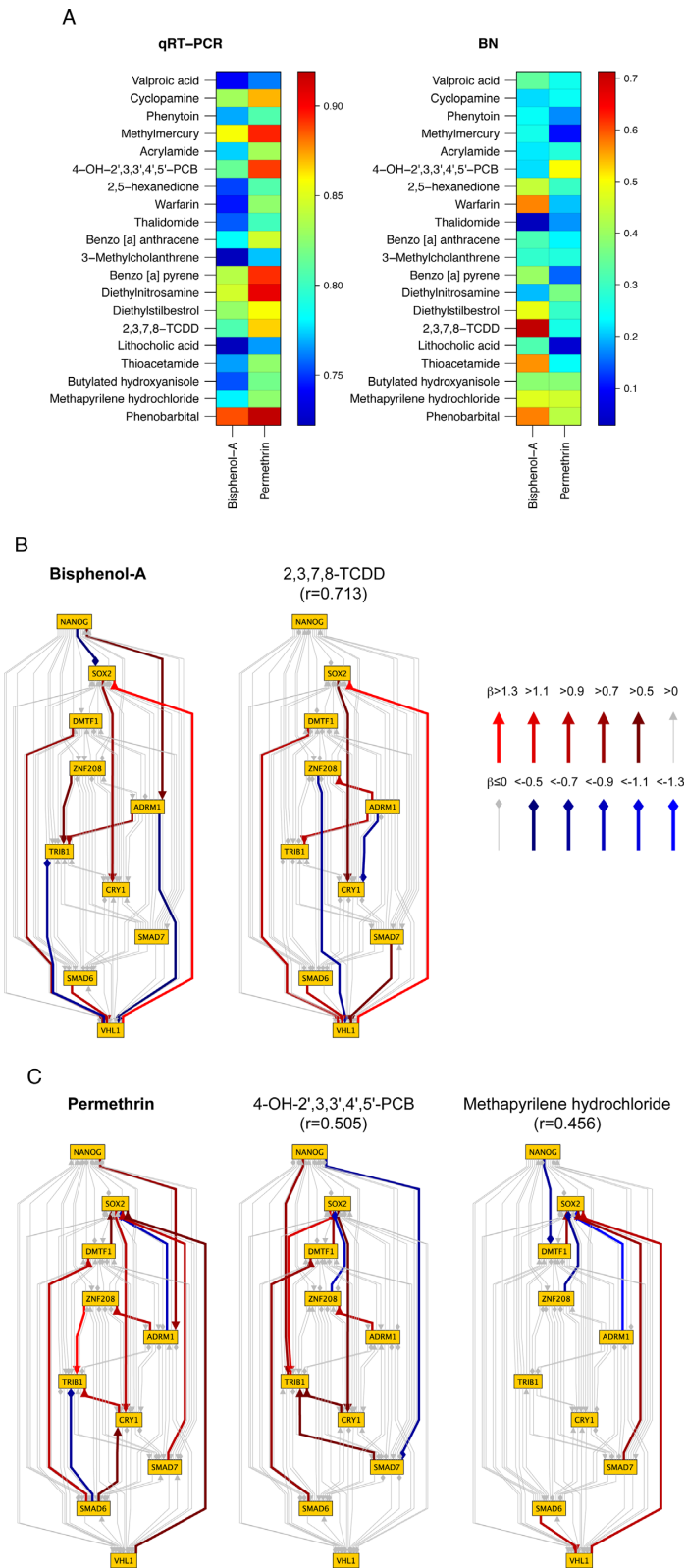
Figure 6. Pearson correlations of bisphenol-A and permethrin against toxic chemicals based on qRT-PCR and BN edge weights. (**A**) Pearson correlation coefficients calculated by qRT-PCR and BN edge weights. (**B**) Bisphenol-A BN exhibits the highest similarity to 2,3,7,8-TCDD (an NGC), with a relatively high Pearson correlation coefficient ($r = 0.713$), whereas (**C**) permethrin BN is moderately similar to both 4-OH-2′,3,3′,4′,5′-PCB (an NT; $r = 0.505$) and methapyrilene hydrochloride (an NGC; $r = 0.456$).

data in the original data space. The concept of an external kernel is expected to provide a new arena for SVMs and other machine-learning methodologies, as well as multivariate statistical analyses, which may yield better mathematical models for real problems, leading in turn to better solutions.

CONCLUSION

We demonstrated that a human stem cell-based system that uses only hESCs without further differentiation shows considerable promise for predicting late-onset chemical toxicities, such as abnormalities that occur during embryonic development. Particularly, we show that toxicity category prediction was dramatically improved by SVMs when adopting gene networks in which the network edge weights are added as feature vectors. In recent years, in order to circumvent ethical problems, methods based on human induced pluripotent stem (iPS) cells rather than hESCs have attracted a great deal of attention in the field of stem cell research. As proposed by Anson *et al.* (57), human iPS cells and their derivatives, such as tissues and organs, are a promising resource for toxicity assessment.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Waters,M., Stasiewicz,S., Merrick,B.A., Tomer,K., Bushel,P., Paules,R., Stegman,N., Nehls,G., Yost,K.J., Johnson,C.H. *et al.* (2008) CEBS–Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic Acids Res.*, **36**(Suppl 1), D892–D900.
2. Deng,Y., Johnson,D.R., Guan,X., Ang,C.Y., Ai,J. and Perkins,E.J. (2010) In vitro gene regulatory networks predict in vivo function of liver. *BMC Syst. Biol.*, **4**, 153.
3. Noriyuki,N., Igarashi,Y., Ono,A., Yamada,H., Ohno,Y. and Urushidani,T. (2011) Evaluation of DNA microarray results in the Toxicogenomics Project (TGP) consortium in Japan. *J. Toxicol. Sci.*, **37**, 791–801.
4. Low,Y., Uehara,T., Minowa,Y., Yamada,H., Ohno,Y., Urushidani,T., Sedykh,A., Muratov,E., Kuz'min,V., Fourches,D. *et al.* (2011) Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chem. Res. Toxicol.*, **24**, 1251–1262.
5. Hirai,T. and Kiyosawa,N. (2013) Developing a practical toxicogenomics data analysis system utilizing open-source software. *Computational Toxicology*, Humana Press, pp. 357–374.
6. Ito,T., Ando,H., Suzuki,T., Ogura,T., Hotta,K., Imamura,Y., Yamaguchi,Y. and Handa,H. (2010) Identification of a primary target of thalidomide teratogenicity. *Science*, **327**, 1345–1350.
7. Davidson,P.W., Myers,G.J. and Weiss,B. (2004) Mercury exposure and child development outcomes. *Pediatrics*, **113**(Suppl 3), 1023–1029.
8. He,X., Imanishi,S., Sone,H., Nagano,R., Qin,X.-Y., Yoshinaga,J., Akanuma,H., Yamane,J., Fujibuchi,W. and Ohsako,S. (2012) Effects of methylmercury exposure on neuronal differentiation of mouse and human embryonic stem cells. *Toxicol. Lett.*, **212**, 1–10.
9. Suemori,H., Yasuchika,K., Hasegawa,K., Fujioka,T., Tsuneyoshi,N. and Nakatsuji,N. (2006) Efficient establishment of human embryonic stem cell lines and long-term maintenance with stable karyotype by enzymatic bulk passage. *Biochem. Biophysic. Res. Commun.*, **345**, 926–932.
10. Judson,R., Richard,A., Dix,D.J., Houck,K., Martin,M., Kavlock,R., Dellarco,V., Henry,T., Holderman,T., Sayre,P. *et al.* (2009) The toxicity data landscape for environmental chemicals. *Environ. Health Perspect.*, **117**, 685–695.
11. Sakuratani,Y., Zhang,H., Nishikawa,S., Yamazaki,K., Yamada,T., Yamada,J., Gerova,K., Chankov,G., Mekenyan,O. and Hayashi,M. (2013) Hazard Evaluation Support System (HESS) for predicting repeated dose toxicity using toxicological categories. *SAR QSAR Environ. Res.*, **24**, 351–363.
12. Kadereit,S., Zimmer,B., van Thriel,C., Hengstler,J. and Leist,M. (2012) Compound selection for in vitro modeling of developmental neurotoxicity. *Front. Biosci.*, **17**, 2442–2460.
13. Theunissen,P., Robinson,J., Pennings,J., de Jong,E., Claessen,S., Kleinjans,J. and Piersma,A. (2012) Transcriptomic concentration-response evaluation of valproic acid, cyproconazole, and hexaconazole in the neural embryonic stem cell test (ESTn). *Toxicol. Sci.*, **125**, 430–438.
14. Magkoufopoulou,C., Claessen,S., Tsamou,M., Jennen,D., Kleinjans,J. and van Delft,J. (2012) A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo. *Carcinogenesis*, **33**, 1421–1429.
15. Gómez-Lechón,M., Tolosa,L., Castell,J. and Donato,M. (2010) Mechanism-based selection of compounds for the development of innovative in vitro approaches to hepatotoxicity studies in the LIINTOP project. *Toxicol. In Vitro*, **24**, 1879–1889.
16. Vinken,M., Doktorova,T., Ellinger-Ziegelbauer,H., Ahr,H., Lock,E., Carmichael,P., Roggen,E., van Delft,J., Kleinjans,J., Castell,J. *et al.* (2008) The carcinoGENOMICS project: critical selection of model compounds for the development of omics-based in vitro carcinogenicity screening assays. *Toxicol. In Vitro*, **659**, 202–210.
17. Do,J.H. (2014) Neurotoxin-induced pathway perturbation in human neuroblastoma SH-EP cells. *Mol. Cells*, **37**, 672–684.
18. Pisani,C., Gaillard,J.-C., Nouvel,V., Odorico,M., Armengaud,J. and Prat,O. (2015) High-throughput, quantitative assessment of the effects of low-dose silica nanoparticles on lung cells: grasping complex toxicity with a great depth of field. *BMC Genomics*, **16**, 315.
19. Lan,J., Gou,N., Rahman,S.M., Gao,C., He,M. and Gu,A.Z. (2016) A quantitative toxicogenomics assay for high-throughput and mechanistic genotoxicity assessment and screening of environmental pollutants. *Environ. Sci. Technol.*, **50**, 3202–3214.
20. Ritchie,M.E., Dunning,M.J., Smith,M.L., Shi,W. and Lynch,A.G. (2011) BeadArray expression analysis using bioconductor. *PLoS Comp. Biol.*, **7**, e1002276.
21. Benzecri,J. (1973) Analyse des Données: l'Analyse des Correspondances (Tome 2, Dunod).
22. Ripley,B. (2011) MASS: support functions and datasets for Venables and Ripley's MASS.
23. Smyth,G.K. (2005) limma: Linear Models for Microarray Data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer, NY, pp. 397–420.
24. Fujibuchi,W., Aburatani,S., Yamane,J., Imanishi,S., Akanuma,H., Sone,H. and Ohsako,S. (2011) Prediction of chemical toxicity by network-based SVM on ES-cell validation system. In: *Proc. of the 2011 Joint Conference of CBI-Society and JSBi*. Japanese Society for Bioinformatics.
25. Friedman,N., Linial,M., Nachman,I. and Pe'er,D. (2000) Using Bayesian networks to analyze expression data. *J. Comp. Biol.*, **7**, 601–620.

26. Yamanaka,T., Toyoshiba,H., Sone,H., Parham,F.M. and Portier,C.J. (2004) The TAO-Gen algorithm for identifying gene interaction networks with application to SOS repair in E. coli. *Environ. Health Perspect.*, **112**, 1614–1621.

27. Nagano,R., Akanuma,H., Qin,X.-Y., Imanishi,S., Toyoshiba,H., Yoshinaga,J., Ohsako,S. and Sone,H. (2011) Multi-parametric profiling network based on gene expression and phenotype data: a novel approach to developmental neurotoxicity testing. *Int. J. Mol. Sci.*, **13**, 187–207.

28. Swendsen,R.H. and Wang,J.-S. (1986) Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.*, **57**, 2607.

29. Earl,D.J. and Deem,M.W. (2005) Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, **7**, 3910–3916.

30. Vapnik,V.N. (1995) *The Nature of Statistical Learning Theory*. Springer-Verlag, NY.

31. Vapnik,V.N. (1998) *Statistical Learning Theory*. John Wiley & Sons, NY. Vol. **1**.

32. Ramaswamy,S., Tamayo,P., Rifkin,R., Mukherjee,S., Yeang,C.-H., Angelo,M., Ladd,C., Reich,M., Latulippe,E., Mesirov,J.P. *et al.* (2001) Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 15149–15154.

33. Fernández-Delgado,M., Cernadas,E., Barro,S. and Amorim,D. (2014) Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.*, **15**, 3133–3181.

34. Fujibuchi,W. and Kato,T. (2007) Classification of heterogeneous microarray data by maximum entropy kernel. *BMC Bioinformatics*, **8**, 267.

35. Kato,T. and Fujibuchi,W. (2010) Kernel Classification Methods for Cancer Microarray Data. In: Emmert-Streib,F and Dehmer,M (eds). *Medical biostatistics for complex diseases*, John Wiley & Sons, pp. 279–303.

36. Robin,X., Turck,N., Hainard,A., Tiberti,N., Lisacek,F., Sanchez,J.-C. and Müller,M. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, **12**, 77.

37. Bennett,K.P. and Demiriz,A. (1998) Semi-supervised support vector machines. In: *Advances in neural information processing systems*. MIT Press, pp. 368–374.

38. Schultz,T.W. (1999) Structure-toxicity relationships for benzenes evaluated with Tetrahymena pyriformis. *Chem. Res. Toxicol.*, **12**, 1262–1267.

39. Jalali-Heravi,M. and Mani-Varnosfaderani,A. (2011) QSAR modelling of integrin antagonists using enhanced Bayesian regularised genetic neural networks. *SAR QSAR Environ. Res.*, **22**, 293–314.

40. Cumming,J.G., Davis,A.M., Muresan,S., Haeberlein,M. and Chen,H. (2013) Chemical predictive modelling to improve compound quality. *Nat. Rev. Drug Discov.*, **12**, 948–962.

41. Koutsoukas,A., Paricharak,S., Galloway,W.R., Spring,D.R., IJzerman,A.P., Glen,R.C., Marcus,D. and Bender,A. (2013) How diverse are diversity assessment methods? A comparative analysis and benchmarking of molecular descriptor space. *J Chem. Inf. Model.*, **54**, 230–242.

42. Su,B.-H., Tu,Y.-S., Esposito,E.X. and Tseng,Y.J. (2012) Predictive toxicology modeling: protocols for exploring hERG classification and Tetrahymena pyriformis end point predictions. *J. Chem. Inf. Model.*, **52**, 1660–1673.

43. Cao,D.-S., Zhao,J.-C., Yang,Y.-N., Zhao,C.-X., Yan,J., Liu,S., Hu,Q.-N., Xu,Q.-S. and Liang,Y.-Z. (2012) In silico toxicity prediction by support vector machine and SMILES representation-based string kernel. *SAR QSAR Environ. Res.*, **23**, 141–153.

44. Jolly,R., Ahmed,K. B.R., Zwickl,C., Watson,I. and Gombar,V. (2015) An evaluation of in-house and off-the-shelf in silico models: Implications on guidance for mutagenicity assessment. *Regul. Toxicol. Pharmacol.*, **71**, 388–397.

45. Omer,A., Singh,P., Yadav,N. and Singh,R. (2014) An overview of data mining algorithms in drug induced toxicity prediction. *Mini Rev. Med. Chem.*, **14**, 345–354.

46. Tetko,I.V., Gasteiger,J., Todeschini,R., Mauri,A., Livingstone,D., Ertl,P., Palyulin,V.A., Radchenko,E.V., Zefirov,N.S., Makarenko,A.S. *et al.* (2005) Virtual computational chemistry laboratory–design and description. *J. Comput. Aid. Mol. Des.*, **19**, 453–463.

47. Leranth,C., Hajszan,T., Szigeti-Buck,K., Bober,J. and MacLusky,N.J. (2008) Bisphenol A prevents the synaptogenic response to estradiol in hippocampus and prefrontal cortex of ovariectomized nonhuman primates. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 14187–14191.

48. Beronius,A., Johansson,N., Rudén,C. and Hanberg,A. (2013) The influence of study design and sex-differences on results from developmental neurotoxicity studies of bisphenol A, implications for toxicity testing. *Toxicology*, **311**, 13–26.

49. Ho,S.-M., Tang,W.-Y., de Frausto,J.B. and Prins,G.S. (2006) Developmental exposure to estradiol and bisphenol A increases susceptibility to prostate carcinogenesis and epigenetically regulates phosphodiesterase type 4 variant 4. *Cancer Res.*, **66**, 5624–5632.

50. Imamura,L., Hasegawa,H., Kurashina,K., Hamanishi,A., Tabuchi,A. and Tsuda,M. (2000) Repression of activity-dependent c-fos and brain-derived neurotrophic factor mRNA expression by pyrethroid insecticides accompanying a decrease in Ca^{2+} influx into neurons. *J. Pharmacol. Exp. Ther.*, **295**, 1175–1182.

51. Nasuti,C., Gabbianelli,R., Falcioni,M.L., Di Stefano,A., Sozio,P. and Cantalamessa,F. (2007) Dopaminergic system modulation, behavioral changes, and oxidative stress after neonatal administration of pyrethroids. *Toxicology*, **229**, 194–205.

52. Ptak,A., Hoffman,M., Gruca,I. and Barc,J. (2014) Bisphenol A induce ovarian cancer cell migration via the MAPK and PI3K/Akt signalling pathways. *Toxicol. Lett.*, **229**, 357–369.

53. Dhimolea,E., Wadia,P.R., Murray,T.J., Settles,M.L., Treitman,J.D., Sonnenschein,C., Shioda,T. and Soto,A.M. (2014) Prenatal exposure to BPA alters the epigenome of the rat mammary gland and increases the propensity to neoplastic development. *PLoS One*, **9**, e99800.

54. Imanishi,S., Okura,M., Zaha,H., Yamamoto,T., Akanuma,H., Nagano,R., Shiraishi,H., Fujimaki,H. and Sone,H. (2013) Prenatal exposure to permethrin influences vascular development of fetal brain and adult behavior in mice offspring. *Environ. Toxicol.*, **28**, 617–629.

55. Guyon,I. and Elisseeff,A. (2003) An introduction to variable and feature selection. *J. Mach. Learn. Res.*, **3**, 1157–1182.

56. Guyon,I., Weston,J., Barnhill,S. and Vapnik,V. (2002) Gene selection for cancer classification using support vector machines. *Mach. Learn.*, **46**, 389–422.

57. Anson,B.D., Kolaja,K. and Kamp,T.J. (2011) Opportunities for human iPS cells in predictive toxicology. *Clin. Pharmacol. Ther.*, **89**, 754–758.