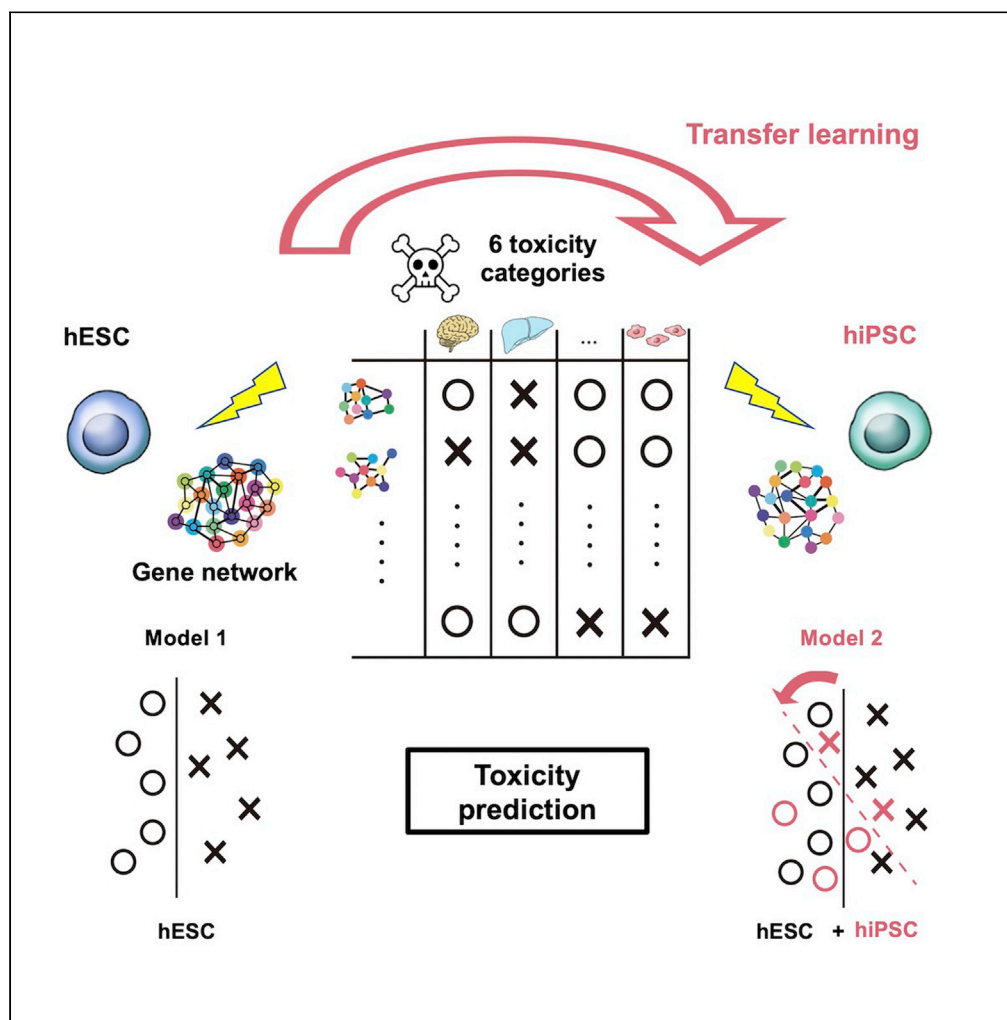**Article**

StemPanTox: A fast and wide-target drug assessment system for tailor-made safety evaluations using personalized iPS cells

Junko Yamane,
Takumi Wada,
Hironori Otsuki, ...,
Mitsujiro Osawa,
Megumu K. Saito,
Wataru Fujibuchi

wfujibuchi@gmail.com

Highlights

Undifferentiated human embryonic stem cells can detect adult chemical toxicities

Toxicity predictions using gene expression networks achieve high predictability

Transfer learning from ES to iPS cells ameliorates the ethical issues

StemPanTox may contribute to toxicity assessment using personalized iPS cells

Article

StemPanTox: A fast and wide-target drug assessment system for tailor-made safety evaluations using personalized iPS cells

Junko Yamane,[1] Takumi Wada,[1] Hironori Otsuki,[2] Koji Inomata,[2] Mutsumi Suzuki,[2] Tomoka Hisaki,[3] Shuichi Sekine,[3] Hirokazu Kouzuki,[3] Kenta Kobayashi,[1] Hideko Sone,[4] Jun K. Yamashita,[1] Mitsujiro Osawa,[1] Megumu K. Saito,[1] and Wataru Fujibuchi[1,5,6,*]

SUMMARY

An alternative model that reliably predicts human-specific toxicity is necessary because the translatability of effects on animal models for human disease is limited to context. Previously, we developed a method that accurately predicts developmental toxicity based on the gene networks of undifferentiated human embryonic stem (ES) cells. Here, we advanced this method to predict *adult* toxicities of 24 chemicals in six categories (neurotoxins, cardiotoxins, hepatotoxins, two types of nephrotoxins, and non-genotoxic carcinogens) and achieved high predictability (AUC = 0.90–1.00) in all categories. Moreover, we screened for an induced pluripotent stem (iPS) cell line to predict the toxicities based on the gene networks of iPS cells using transfer learning of the gene networks of ES cells, and predicted toxicities in four categories (neurotoxins, hepatotoxins, glomerular nephrotoxins, and non-genotoxic carcinogens) with high performance (AUC = 0.82–0.99). This method holds promise for tailor-made safety evaluations using personalized iPS cells.

INTRODUCTION

To date, chemical toxicity studies have been primarily conducted by *in vitro* testing in cultured human cancer cell lines or in animals such as mouse, rat, and rabbit. However, because these tests are not on "normal" human cells, their applications are limited (Perel et al., 2007). In addition, the use of animals has become a major issue from the standpoint of animal welfare; in 2019, the U.S. Environmental Protection Agency (EPA) announced that research studies using mammals as well as funding for mammal studies will be cut by 30% by 2025 and abolished by 2035 (Grimm, 2019). To reduce animal studies, new approach methodologies (NAMs) (Topical Scientific Workshop - New Approach Methodologies in Regulatory Science, 2016), which are any technology, methodology, approach, or combination thereof that can be used to provide information on chemical hazard and risk assessment that avoids the use of intact animals (New Approach Methods Work Plan: Reducing Use of Animals in Chemical Testing, 2016), are widely investigated and adopted for the development of new tools for testing pharmaceuticals and other chemicals for potential adverse impacts on human health and ecological endpoints, under the direction of the U.S. EPA and the amended Toxic Substances Control Act (TSCA) for improving regulatory toxicology guidelines to consider the use of toxicity information, computational toxicology and bioinformatics, and high-throughput screening methods and prediction models, prior to animal studies (Becker, 2019).

The embryonic stem cell test (EST) reported by Scholz et al. was the first to examine embryotoxicity *in vitro* using mouse fibroblasts, embryonic stem (ES) cells, and cardiomyocytes differentiated from ES cells; these developmental toxicity tests were previously performed only in animals (Scholz et al., 1999; Seiler and Spielmann, 2011). Later, this method was approved as a scientifically valid alternative by the European Centre for the Validation of Alternative Methods (ECVAM) (https://tsar.jrc.ec.europa.eu/test-method/tm1999-01). However, the EST uses mouse cells, and species-specific differences must be clarified in order to extrapolate its results to human. Subsequently, another research group reported that in an EST based on a human cell system (hEST), homologous human and mouse neurodevelopmental gene expressions are similar (Schulpen et al., 2015). Moreover, there is an extensive collection of literature on the research and

[1]Center for IPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

[2]Toxicological Research Laboratories, Translational Research Unit, R&D Division, Kyowa Kirin Co., Ltd., 1188 Shimotogari, Nagaizumi-cho, Sunto-gun, Shizuoka 411-8731, Japan

[3]MIRAI Technology Institute, Shiseido Co., Ltd., 1-2-11, Takashima, Nishi-ku, Yokohama-shi, Kanagawa 220-0011, Japan

[4]Environmental Health and Prevention Research Unit, Yokohama University of Pharmacy, 601 Matano-cho, Totsuka-ku, Yokohama-shi, Kanagawa 245-0066, Japan

[5]Present address: Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

[6]Lead contact

*Correspondence: wfujibuchi@gmail.com

https://doi.org/10.1016/j.isci.2022.104538

development of pluripotent stem cell models for predictive toxicology, including assays based on specific biomarkers such as *Hand-1* and *Sox17* (Kameoka et al., 2014; Suzuki et al., 2011), assays customized to specific toxicities such as teratogenicity (West et al., 2010), models for relevant pathways such as *Wnt* signaling or *TGF-b* (Kugler et al., 2015; Uibel et al., 2010), detected by various modalities such as luciferase reporter, metabolomics, high-throughput imaging (Kameoka et al., 2014; Kleinstreuer et al., 2011; Uibel et al., 2010), and others, which have been summarized in the review papers of Luz and Tokar (2018) and Kim et al. (Kim et al., 2019).

In 2012, the U.S. Defense Advanced Research Projects Agency (DARPA) and the NIH invested a huge amount of budget on a national project to promote the development of biomimetic systems, leading to rapid progress in the field (Low et al., 2021). These systems, which mimic the (adult) human body and are constructed by filling tissues created in individual compartments with culture fluid and connecting them together, are expected to be used in human toxicity testing systems as an alternative to animals. Currently, however, very little progress has been made in adapting these systems for developmental toxicity testing, and no evaluation method has been established to determine how accurately these systems mimic the function of the normal human body (Allwardt et al., 2020). Realizing the practical application of these systems as high-throughput toxicity screening tools is likely to take several years.

For many years, the prediction of chemical toxicity has been carried out using a method based on the physicochemical parameters of the chemicals, referred to as the quantitative structure–activity relationship (QSAR) (Schultz, 1999). Computational QSAR models have been widely used in various fields of predictive toxicology such as DILI severity in hepatotoxicity or hERG ligands in cardiotoxicity (Ekins, 2014). However, there is a limit to the predictive ability of QSAR. One reason is that the mechanism that induces toxic responses resides within the cell, so information about the chemical alone cannot predict the response. In this regard, it should be possible to detect toxicity more accurately by obtaining information about changes in gene expressions in the cells. In addition, as stem cells differentiate, only the genes essential for that lineage are expressed and conserved through DNA methylation (Reik et al., 2001), whereas in pluripotent stem cells, a very large number of genes are expressed, including transporters and transcription factors; consequently, pluripotent stem cells are superior to differentiated cells as a tool for comprehensively detecting toxic chemicals. In light of these considerations, we developed hEST-GN (human embryonic stem cell test with gene networks), a prediction method that uses information on feature gene networks based on massive gene expression datasets obtained by exposing human ES cells to toxic chemicals as input data for machine learning. Using this method, we achieved highly accurate predictions of developmental toxicity categories (Yamane et al., 2016).
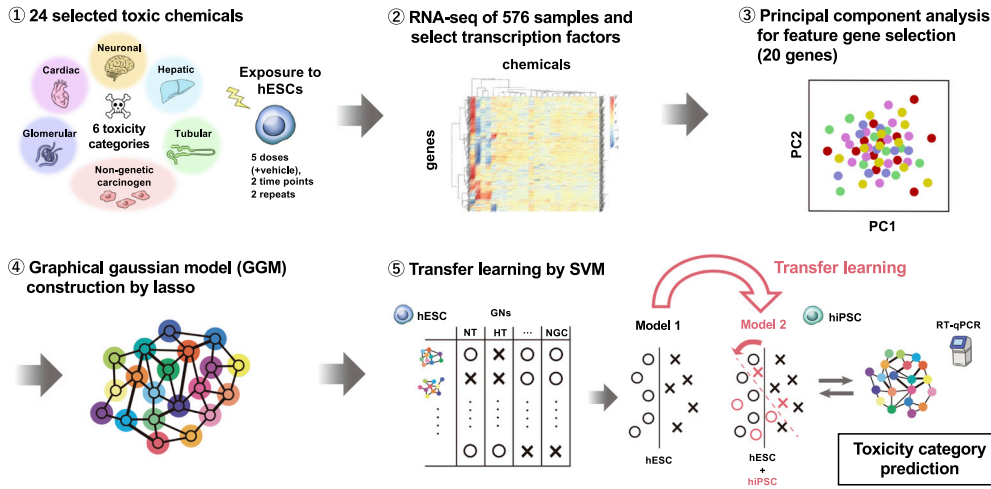
In this study, we advanced the hEST-GN and achieved high performance predictions of 24 toxic chemicals in broad toxicity categories, including adult toxicity. Furthermore, by selecting induced pluripotent stem (iPS) cells, which can be used as an alternative to ES cells in toxicity testing, and using a gene expression database created from ES cells, we were able to develop a system called "StemPanTox" that can predict chemical toxicity using iPS cell data via transfer learning and ameliorate the ethical issues related to ES cells. With further improvement, StemPanTox will contribute to the development of tailor-made, individualized toxicity assessment/prevention using personalized iPS cells, which would have enormous clinical value.
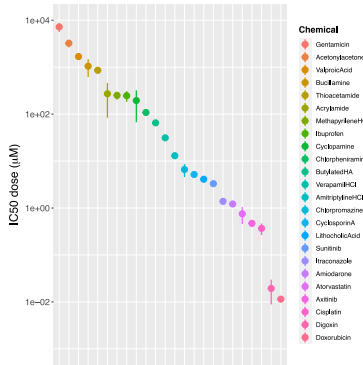
RESULTS

Development of an hEST-GN library for 24 chemicals and prediction using iPS cells

A schematic of the chemical assay is shown in Figure 1A. The human ES cell line KhES-3 was exposed to a total of 24 chemicals in six toxicity categories [neurotoxins, hepatotoxins, cardiotoxins, two types of nephrotoxins (glomerular nephrotoxins and tubular nephrotoxins), and non-genotoxic carcinogens] at six concentrations, including vehicle (solvent) alone. The chemicals were carefully assigned to the toxicity categories by referring to previous reports (Table 1). Gene expression data were obtained by RNA-seq at two time points, 24 and 48 h, after exposure. At each time point, a principal component analysis (PCA) was performed using transcription factor genes, and a total of 20 genes from the top five PCs were extracted as feature genes. Using these genes, gene network libraries were created for each of the 24 chemicals using the graphical Gaussian model (GGM) (Goodal, 1991). Similarly, the screened iPS cells were subjected to RT-qPCR to obtain gene expression data for the same 20 genes and create gene network libraries. Next, using ES cell library labels, a chemical toxicity prediction system trained by both libraries from ES and
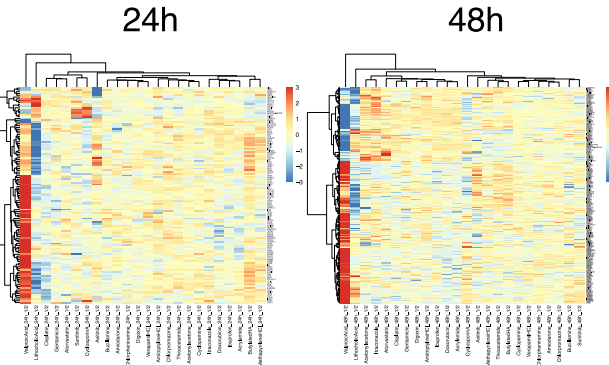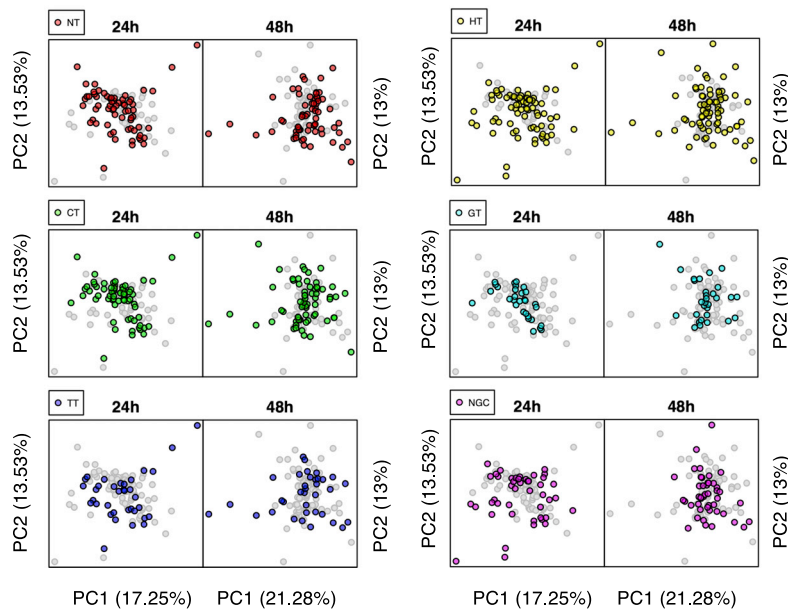
Figure 1. Construction of a gene expression database for 24 chemicals

(A) Schema of the chemical assay. hESC, human ES cells; hiPSC, human iPS cells. (B) IC50 for 24 chemicals. Data are represented as estimated value with 95% confidence intervals. (C) Transcription factor genes differentially expressed following exposure to 24 chemicals at ½ dose. (D) PCA of 24 chemicals in six toxicity categories at two time points.

iPS cells was developed via transfer learning (Vapnik, 2006) using support vector machines (SVMs) (Cortes and Vapnik, 1995).

Gene expression response database of ES cells for 24 chemicals

To obtain the largest amount of data about the expression of genes that were perturbed by the 24 chemicals, ES cells need to be exposed to chemicals at the maximum concentration that does not cause an excessive degree of cell death. To this end, we first performed ATP assays and then plotted regression curves to calculate inhibitory concentrations (ICs) by carrying out serial dilutions of stock solutions containing the maximum soluble concentrations of the 24 chemicals. ICs in the range of 0.1%–50%, at which cell death begins to be observed, were set as the maximum exposure concentration (Table S1, Figure S1). The estimated IC50 and 95% confidence interval (CI) for each chemical are shown in Figure 1B (Table S2). Serial dilutions were carried out, with the maximum exposure concentration set as 1/1 to obtain ½, 1/4, $^{1}/_{8}$, and 1/16 dilutions, and a six-step exposure including vehicle alone was performed and repeated twice, yielding a total of 6 × 2 = 12 samples for each chemical. We collected RNA 24 and 48 h after exposure to the 24 chemicals, performed transcriptome analysis, and generated gene expression datasets for a total of 12 × 24 × 2 = 576 samples.

To examine the characteristics of the 24 chemicals at the level of differentially expressed genes (DEGs), we selected transcription factor-related genes (GO: 0,006,351) from the 576 datasets (4,032 genes). After log-normalization, batch effect elimination, and repeat merging, we generated DEG sets for which differences between each exposure data and vehicle values were significant (FDR <0.01 and $\log_2|FC| > 1$) and presented them in a heatmap (Figures 1C and S2). According to this analysis, the number of DEGs was higher at 48 h than at 24 h for all concentrations, and over time, more genes were up- or downregulated due to exposure to the chemicals. At both 24 and 48 h, valproic acid, a strong neurodevelopmental toxicant, elicited gene expression patterns that were clearly distinct from those of the other chemicals. Similarly, lithocholic acid, a mammalian bile acid and well-known carcinogen, yielded distinct expression patterns.

Construction of gene networks of the 24 chemicals by the graphical Gaussian model (GGM)

To obtain feature genes used in the prediction, we performed PCA on the basis of the exposure data of each transcription factor gene, expressed as a log fold change (LFC) in expression relative to vehicle after log-normalization and batch effect elimination. There were 3,200 genes for the 24-h samples and 3,255 genes for the 48-h samples. For both exposure times, it was difficult to clearly separate the chemicals by the toxicity categories using two-dimensional PCA (Figures 1D and S3). Accordingly, we selected two genes with maximum positive and negative loading values, which were considered to contribute the most to the first to fifth PCs; at each time point, 20 genes were selected as feature genes (Table S3). Among the 20 selected genes in each group, only *ACTR3* (Welch et al., 1997), which has been implicated in cell shape and motility, was common.

Using the 20 selected genes, we estimated sparse gene networks based on GGMs using an L1 graphical lasso for each chemical at each time point (i.e., 24 and 48 h) (Figures 2A, S4, and S5). The figures illustrate the estimated 190 partial correlation coefficients incorporated into the gene networks; edges with positive partial correlation values between two genes are shown in green, and edges with negative values are shown in red; the thickness, distance, and arrangement of the edges correspond to the degree of correlation between the two genes. Because these genes were obtained from the top five PCs that maximize the dispersion of the 24 chemicals using PCA, the estimated GGMs that describe the networks of all 20 genes differed considerably among chemicals, and it was difficult to classify the chemicals simply based on the network patterns as a whole. Therefore, for actual predictions, we decomposed the networks into their constituent edges rather than using them as a whole and used those with higher discriminative potential for training data as features. In other words, the partial correlation coefficients of the edges that are characteristic of the respective toxicity categories contributed to the SVM discrimination.

Prediction of six toxicity categories using KhES-3 cells and the GGM network

Using the 190 partial correlation coefficients in the GGM as input data, we predicted the toxicities of the 24 toxic chemicals in six categories using SVMs with leave-one-out-cross-validation (LOOCV). For the

Table 1. List of 24 chemicals

Chemical name[a]	CAS RN	M.W.	NT	HT	CT	GT	TT	NGC
Acetonylacetone	110-13-4	114.14	Nervous System Diseases [LoPachin et al., 2002]	NA	NA	NA	NA	NA
Acrylamide	79-06-1	71.08	Nervous System Diseases [LoPachin et al., 2002]	Liver Diseases [Allam et al., 2010]	NA	NA	NA	Group B2 https://www.epa.gov/iris
Amiodarone	1951-25-3	645.3	Cerebellar Diseases [Abarbanel et al., 1987]	DILIrank: 8 (Most-DILI-Concern) LiverTox: Likelihood score: A	Cardiotoxicity [Clements et al., 2015]	Glomerulonephritis, Membranous [Kimura et al., 2008]	NA	NA
AmitriptylineHCl	549-18-8	313.9	Nervous System Diseases [Sudoh et al., 2004]	DILIrank: 5 (Less-DILI-Concern) LiverTox: Likelihood score: B	Arrhythmias, Cardiac [Srinivasa et al., 2003]	NA	NA	NA
Atorvastatin	134523-00-5	558.6	Hemorrhagic Stroke [https://www.drugs.com/sfx/atorvastatin-side-effects.html]	DILIrank: 5 (Most-DILI-Concern) LiverTox: Likelihood score: A	NA	NA	NA	NA
Axitinib	319460-85-0	386.5	NA	LiverTox: Likelihood score: E	Cardiotoxicity [Clements et al., 2015] Heart disease [Lin and Will, 2012]	Proteinuria [https://www.drugs.com/sfx/axitinib-side-effects.html]	NA	NA
Bucillamine	65002-17-7	223.3	NA	NA	NA	Nephrotic Syndrome [Ohno, 2004]	NA	NA
Butylated HA	25013-16-5	180.24	NA	NA	NA	NA	NA	Group 2B Precancerous Conditions [Hirose et al., 1990]
Chlorpheniramine	132-22-9	274.79	NA	DILIrank: 0 (No-DILI-Concern) LiverTox: Likelihood score: E	NA	NA	NA	NA
Chlorpromazine	50-53-3	318.9	Parkinsonian Disorders [Ferguson and Paule, 1992]	DILIrank: 2 (Less-DILI-Concern) LiverTox: Likelihood score: A	Cardiomyopathies [Saito et al., 1985]	NA	NA	NA

Table 1. *Continued*

Chemical name[a]	CAS RN	M.W.	NT	HT	CT	GT	TT	NGC
Cisplatin	15663-27-1	300.05	Peripheral Neuropathy [https://www.drugs.com/sfx/cisplatin-side-effects.html]	DILIrank: 3 (Less-DILI-Concern) LiverTox: Likelihood score: C	NA	NA	Acute Renal Failure [Ozkok and Edelstein, 2014]	Group 2A http://ntp.niehs.nih.gov/ntp/roc/eleventh/profiles/s053cycl.pdf
Cyclopamine	4449-51-8	411.6	NA	NA	NA	NA	NA	NA
Cyclosporin A	59865-13-3	1202.6	Abducens Nerve Diseases [Pilling et al., 2005]	DILIrank: 7 (Most-DILI-Concern) LiverTox: Likelihood score: C	Heart Diseases [Miller, 2002]	NA	Acute Kidney Injury [Humes et al., 1985]	Group 1 http://ntp.niehs.nih.gov/ntp/roc/eleventh/profiles/s053cycl.pdf
Digoxin	20830-75-5	780.9	Cognition Disorders [Maheut-Bosser et al., 2006]	DILIrank: 0 (No-DILI-Concern) LiverTox: Likelihood score: E	Heart arrest [Rea et al., 2003]	NA	NA	Group 2B https://monographs.iarc.who.int/list-of-classifications
Doxorubicin	23214-92-8	543.5	Spinal Cord Diseases [Jordan et al., 2004]	DILIrank: 3 (Less-DILI-Concern) LiverTox: Likelihood score: B	Arrhythmias, Cardiac [Mladosievicova et al., 2001]	Albuminuria [Baroni et al., 1999]	NA	Group 2A https://monographs.iarc.who.int/agents-classified-by-the-iarc/
Gentamicin	1403-66-3	477.6	Hearing Loss [Sha and Schacht, 1999]	LiverTox: Likelihood score: E	Heart Diseases [Hendry et al., 1988]	NA	Kidney Tubular Necrosis, Acute [Xie et al., 2001]	NA
Ibuprofen	15687-27-1	206.28	Meningoencephalitis [Lee et al., 2002]	DILIrank: 3 (Less-DILI-Concern) LiverTox: Likelihood score: A	Heart Septal Defects, Ventricular [Cappon et al., 2005]	NA	Kidney Tubular Necrosis, Acute [Marasco et al., 1987]	NA
Itraconazole	84625-61-6	705.6	NA	DILIrank: 8 (Most-DILI-Concern) LiverTox: Likelihood score: B	Heart Failure [Okuyan and Altin, 2013]	NA	Kidney Diseases [Boelaert et al., 1988]	NA
Lithocholic Acid	434-13-9	376.6	NA	Liver Cirrhosis [Fickert et al., 2006]	NA	NA	NA	Precancerous Conditions [Kitazawa et al., 1990] [Kitazawa, 1993] Cystic fibrosis [Colombo et al., 2016]

(Continued on next page)

Table 1. *Continued*

Chemical name[a]	CAS RN	M.W.	NT	HT	CT	GT	TT	NGC
Methapyrilene HCl	135-23-9	297.8	NA	Liver Neoplasms [Ellinger-Ziegelbauer et al., 2005]	NA	NA	NA	Precancerous Conditions [Mizukami et al., 2017]
Sunitinib	557795-19-4	398.5	NA	DILIrank: 8 (Most-DILI-Concern) LiverTox: Likelihood score: B	Arrhythmias, Cardiac [Doherty et al., 2013]	Proteinuria [Sorich et al., 2016]	NA	NA
Thioacetamide	62-55-5	75.14	NA	Hepatic Insufficiency [El-Tanbouly et al., 2017] Liver Cirrhosis [Seong et al., 2003]	NA	NA	NA	Group 2B http://ntp.niehs.nih.gov/ntp/roc/eleventh/profiles/s172thio.pdf
Valproic Acid	99-66-1	144.21	CNS Diseases [Ruuskanen et al., 1979]	DILIrank: 8 (Most-DILI-Concern) LiverTox: Likelihood score: A	Heart Septal Defects, Atrial [Carter and Stewart, 1989]	NA	Kidney Diseases [Raza et al., 2000]	NA
Verapamil HCl	152-11-4	491.1	NA	DILIrank: 3 (Less-DILI-Concern) LiverTox: Likelihood score: B	Ventricular Fibrillation [McGovern et al., 1986]	Kidney Diseases [Ali et al., 2002]	Kidney Tubular Necrosis, Acute [Kreft et al., 1991]	NA

Bold characters indicate positive toxicity.

NT, Neurotoxin; HT, Hepatotoxin; CT, Cardiotoxin; GT, Glomerular toxin (Nephrotoxin); TT, Tubular toxin (Nephrotoxin); NGC, Non-genotoxic carcinogen.

[a]Chemical names are given by referring to PubChem (as of April 20, 2021).
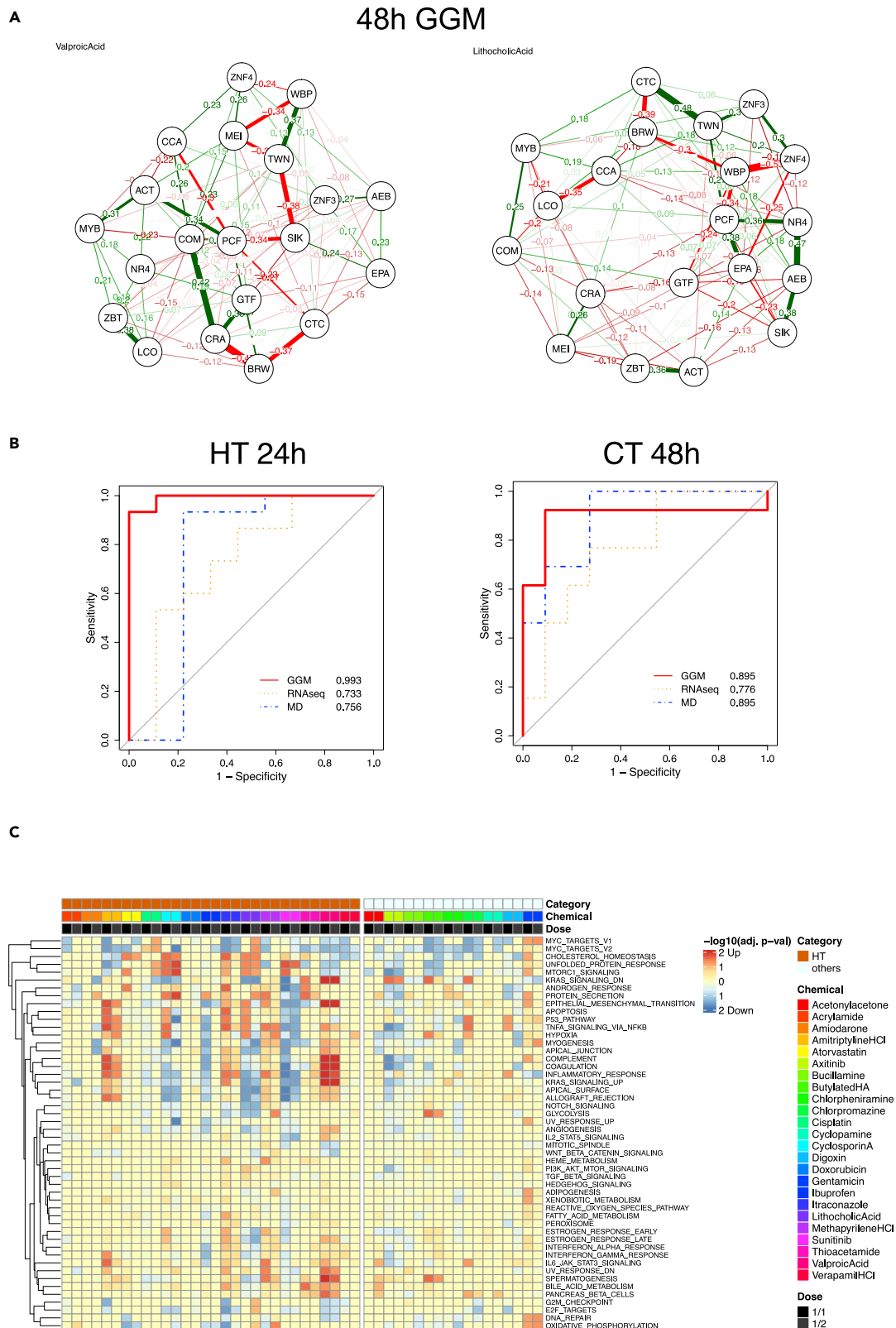
Figure 2. Prediction of six toxicity categories using KhES-3 cells

(A) Gene network representation of GGMs from KhES-3 cells. (B) ROC curves for the prediction of chemicals in two toxicity categories (C) Pathway analysis for hepatotoxins at 24 h and high-dose (1/1, ½ doses) samples.

Table 2. Summary of prediction performance for KhES-3

			NT (13)	HT (15)	CT (13)	GT (6)	TT (7)	NGC (9)
Random	24 h	Mean Accuracy (%)	85.8	83.7	83.3	87.5	87.0	87.0
		SD	8.13	6.92	6.81	6.21	6.65	5.38
		Mean AUC	0.86	0.82	0.85	0.88	0.86	0.87
		SD	0.09	0.11	0.09	0.07	0.11	0.06
	48 h	Mean Accuracy (%)	85.8	83.7	83.3	87.5	87.0	87.0
		SD	8.13	6.92	6.81	6.21	6.65	5.38
		Mean AUC	0.86	0.82	0.85	0.88	0.86	0.87
		SD	0.09	0.11	0.09	0.07	0.11	0.06
Molecular descriptors		Accuracy (%)	83.3	70.8	83.3	83.3	**91.7**	83.3
		AUC	0.83	0.76	0.90	0.82	0.86	0.60
RNA-seq log fold change	24 h	Accuracy (%)	75.0	75.0	79.2	79.2	75.0	66.7
		AUC	0.73	0.73	0.83	0.83	0.74	0.76
	48 h	Accuracy (%)	62.5	75.0	75.0	83.3	79.2	66.7
		AUC	0.71	0.73	0.78	0.66	0.61	0.79
GGM network coefficient	24 h	Accuracy (%)	83.3	**91.7***	**87.5**	87.5	79.2	**91.7**
		AUC	0.73	**0.99***	**0.91**	0.83	0.76	**0.97***
	48 h	Accuracy (%)	87.5	83.3	**91.7***	91.7	91.7	83.3
		AUC	**0.93**	0.79	0.9	**1.00***	**1.00***	0.80

p < 0.05 and *p < 0.01 in one-sample t-test (one-sided).

NT, Neurotoxin; HT, Hepatotoxin; CT, Cardiotoxin; GT, Glomerular toxin (Nephrotoxin); TT, Tubular toxin (Nephrotoxin); NGC, Non-genotoxic carcinogen.

SD, Sample SD; AUC, Area Under the ROC Curve; GGM, Graphical Gaussian Model.

predictions, we followed the procedures described in a previous report on hEST-GN (Yamane et al., 2016) using four kernels (linear, polynomial, RBF, and maximum entropy) and increased the number of top features ranked by a t-test from 1 to 190. We also performed the prediction with the raw LFC values of the 3,200 (24 h) and 3,255 (48 h) transcription factor genes at each of the five concentrations relative to the vehicle-only expression. To compare the predictive accuracy, the same number of input data used for the GGM (i.e., up to 190 genes) was used as features. Predictions with the raw LFC values did not achieve a significantly higher prediction performance than the mean predictions using 10 uniform random numbers. On the other hand, in the prediction based on the GGM, the area under the receiver operating characteristic curve (AUC) values were ≥0.90 for chemicals in all toxicity categories, and because the prediction accuracy or AUC values was significantly high (p < 0.05), we concluded that prediction with high performance is possible. Predictions were performed separately at 24 and 48 h, but depending on the chemical, the time point at which higher prediction accuracy could be obtained differed; thus, neither time point was considered particularly superior in terms of yielding a better prediction. Overall, these results demonstrated that hEST-GN based on ES cell gene networks allows for the prediction of not only developmental toxicity but also broad toxicity categories including adult toxicity.

In addition, to compare with predictions based on the QSAR theory, we generated 5,666 molecular descriptors including 3D descriptors (Table S4) and performed predictions using top 1 to 190 feature genes according to the aforementioned method. None of the six categories, except for tubular nephrotoxin (accuracy of 91.7%), gave a significantly high prediction result. The results of all predictions are presented together in a table and as ROC curves (Table 2, Figures 2B and S6). These results suggest that chemical toxicity predictions that use the partial correlation coefficients of the GGM as features can achieve significantly higher accuracy than predictions based on gene expression values or QSAR. In the GGM-based prediction, the prediction accuracy for each of the 24 chemicals was examined from the SVM results. This analysis revealed that with respect to 16 chemicals (acetonylacetone, acrylamide, amitriptyline HCl, atorvastatin, bucillamine, chlorpheniramine, chlorpromazine, digoxin, doxorubicin, gentamicin, itraconazole, lithocholic acid, methapyrilene HCl, sunitinib, thioacetamide, and valproic acid), the prediction accuracy was 100% for all six categories at 24 and 48 h. On the other hand, for axitinib, cisplatin, and cyclosporin

A, the prediction accuracy was 66.6%, suggesting that the prediction of these chemicals is difficult (Table S5).

Pathway analysis of KhES-3 genes following exposure to 24 chemicals

To determine the effects of exposure to chemicals on biological pathways, we performed a Hallmark pathway analysis by Gene Set Enrichment Analysis (GSEA) using all genes. When performing this analysis, we divided the samples into high dose (1/1 and ½ doses) and low dose ($1/8$ and 1/16 doses). For hepatotoxins, cardiotoxins, and glomerular nephrotoxins, differences were observed in the types of pathways that were induced or suppressed in comparison with other toxic chemicals in the high-dose samples at 24 h (Figures 2C and S7). In addition, the responses of ES cell genes to the toxic chemicals were diverse and dependent on the type of chemical to which they were exposed and not limited to specific pathways such as apoptosis. This observation suggests that it is possible to predict toxicity categories on the basis of perturbed pathways that can be detected by transcription factors. On the other hand, differences in concentrations or among categories that may have been present at 48 h were not as pronounced as those at 24 h. However, analyses using available pathways based on human knowledge accumulated in the past provide limited information. Instead, the computational extraction of feature genes from the PCA of all genes without bias and predictions based on their GGM networks is likely to be more effective.

Selection of iPS cells as an alternative to human ES cells

The results of the present and previous studies suggest that hEST-GN can predict not only developmental toxicity but also broad toxicity categories with high performance. However, there are still hurdles to overcome, including ethical issues, before this system can be generally and widely accepted as a toxicity test. Accordingly, to make iPS cells a possible alternative to hEST-GN, we performed pre-screening by comparing ATP assays with ES cells exposed to 20 toxic chemicals across a wide range of categories. As candidates, we used the top 20 cell lines selected from among Japanese male cell lines (Matsuda et al., 2020) derived from healthy individuals, which had been examined and ranked in terms of their differentiation potential into the three germ layers. For exposure concentration, we adopted the IC50 that was determined using the KhES-3 cell line and examined the toxicity response of human iPS cells. Among the candidate cell lines, we selected the top three cell lines with well-correlated growth rates at IC50 (HPS4138, HPS4234, and HPS4046) and confirmed the correlation coefficients of the growth rates at IC50 with KhES-3 using 20 of the 24 toxic chemicals investigated in this study. HPS4138 had the highest value of 0.94; accordingly, this cell line was used for the predictions as an alternative to ES cells (Table S6).

Prediction of chemicals in six toxicity categories using HPS4138 iPS cells

For HPS4138, which was selected by screening, we performed ATP assays with the 24 chemicals (Figure S8; Tables S7 and S8). As in the case of ES cells, the cells were exposed to vehicle alone or five concentrations obtained by serial dilutions of stock solutions containing the maximum exposure concentration (i.e., the maximum value between IC0.1 and IC50 that did not cause an excessive degree of cell death); this experiment was repeated twice. Gene expression data for 20 genes selected from KhES-3 cells at 24 and 48 h were obtained by qRT-PCR, and GGMs were created for each of the 24 chemicals based on LFC values relative to the vehicle. Partial correlation coefficients were used for the prediction, as in the case of ES cells. In the prediction, data were created by integrating iPS cell data with ES cell data as well as by transductive transfer learning, in which toxicity category labels in the ES cell data were used for the learning to allow for category prediction using iPS cells. Assessment was performed by LOOCV, similarly to the predictions made using ES cells only. Chemicals in all categories except cardiotoxins and tubular nephrotoxins yielded AUC values from 0.82 to 0.99, and the accuracy or AUC was significantly higher than for results obtained with uniform random numbers. Thus, although this approach was not perfect, the results of the prediction using HPS4138 were very accurate for most toxicity categories (Table 3). The summary of toxicity category predictions for the 24 chemicals using HPS4138 is shown in Figure 3. Prediction was difficult for butylated HA, but satisfactory for the other chemicals (Table S9). These results demonstrate that if a gene expression database for toxicity responses could be created with ES cells, chemical toxicity prediction using iPS cells would also be possible by means of transductive transfer learning. Our findings also raise the possibility of achieving practical applications of toxicity testing using standardized or individualized iPS cells in the future. A description of transductive transfer learning is available on our web site at https://stempantox.stemcellinformatics.org.

Table 3. Prediction for HPS4138 cells

			NT (13)	HT (15)	CT (13)	GT (6)	TT (7)	NGC (9)
Random	24 h	Mean Accuracy (%)	81.6	82.5	80.8	89.5	85.0	81.2
		SD	3.49	3.82	4.89	2.95	4.49	5.65
		Mean AUC	0.81	0.77	0.78	0.79	0.80	0.74
		SD	0.06	0.08	0.07	0.08	0.10	0.11
	48 h	Mean Accuracy (%)	81.6	82.0	80.4	88.7	86.2	81.2
		SD	5.27	4.82	5.21	2.03	4.43	4.04
		Mean AUC	0.77	0.81	0.8	0.84	0.77	0.77
		SD	0.09	0.05	0.07	0.10	0.13	0.08
GGM network coefficient	24 h	Accuracy (%)	87.5*	87.5*	79.2	87.5	79.2	91.7*
		AUC	0.90*	0.85*	0.81	0.84	0.71	0.82
	48 h	Accuracy (%)	91.7*	83.3	79.2	95.8*	83.3	83.3
		AUC	0.99*	0.85	0.75	0.94*	0.81	0.77

p < 0.05 and *p < 0.01 in one-sample t-test (one-sided).
NT, Neurotoxin; HT, Hepatotoxin; CT, Cardiotoxin; GT, Glomerular toxin (Nephrotoxin); TT, Tubular toxin (Nephrotoxin); NGC, Non-genotoxic carcinogen.
SD, Sample SD; AUC, Area Under the ROC Curve; GGM, Graphical Gaussian Model.

DISCUSSION

Here, we presented a proof-of-concept study that enables a toxicity hazard assessment using human iPS cells and transfer learning based on the transcription factor gene network libraries made from the gene expression data of human ES cells exposed to 24 chemicals for 6 categories.

In this paper, we clarified that i) human ES cells are sufficient to detect not only developmental toxicities during embryogenesis but also broad toxicity categories such as adult toxins (neurotoxin, cardiotoxin, hepatotoxin, and glomerular and tubular nephrotoxins) and non-genotoxic carcinogens; ii) the chemical toxicity prediction using transcription factor gene networks of human ES cells shows an AUC = 0.90–1.00, which is significantly more accurate than predictions based on the QSAR theory or from raw gene expression data; iii) there exist differences in the biological pathways affected by the toxicity categories, suggesting the mechanisms that underlie the transcription factor networks that control the pathways may be used to predict toxicity categories; and iv) the gene network data from properly screened human iPS cells can successfully, although not perfectly, predict the toxicity categories at significant accuracies once the toxicities are learned by transfer learning using the models based on human ES cell data only.

Various alternative methods using pseudo-human systems, such as differentiated human cell lines (HepG2, MCF-7, HeLa, etc.), have been reported (Qu et al., 2021) since animal protection has become a higher priority in research. However, these lines are often derived from cancer or immortalized cells and thus have limited use (Kim et al., 2019). On the other hand, primary cells, which are assumed to resemble natural states in the human body, show batch-to-batch variability (Kim et al., 2019) and are difficult to collect at sufficient amounts. Furthermore, it is difficult to extrapolate toxicity tests on some cell types to other target cell types due to differences in cytotoxicity tolerance (Laschinski et al., 1991). Performing a multi-target toxicity prediction system based on stem cells, as we propose here, provides a more valid prediction of toxicity to a larger range of cell types.

Toxicological assessment using the transcriptome is frequently used by the U.S. EPA, Tox21 project and in Europe. Particularly, new approach methodologies (NAMs), which are any technology, methodology, approach, or combination thereof that can be used to provide information on chemical hazards and risk assessments that avoids the use of intact animals (New Approach Methods Work Plan: Reducing Use of Animals in Chemical Testing 2016), are often directional concepts using transcriptomics with other omics or traditional toxicology methods. Our study indicated that transcription factor gene networks exist in a master layer of biological pathways to activate molecular initiation events (MIEs), in which the initial chemical trigger starts an adverse outcome pathway (AOP) via DNA binding, receptor activation, or a disturbance of cellular/organelle systems (Allen et al., 2016), thus revealing toxicity reactions. Recent studies have
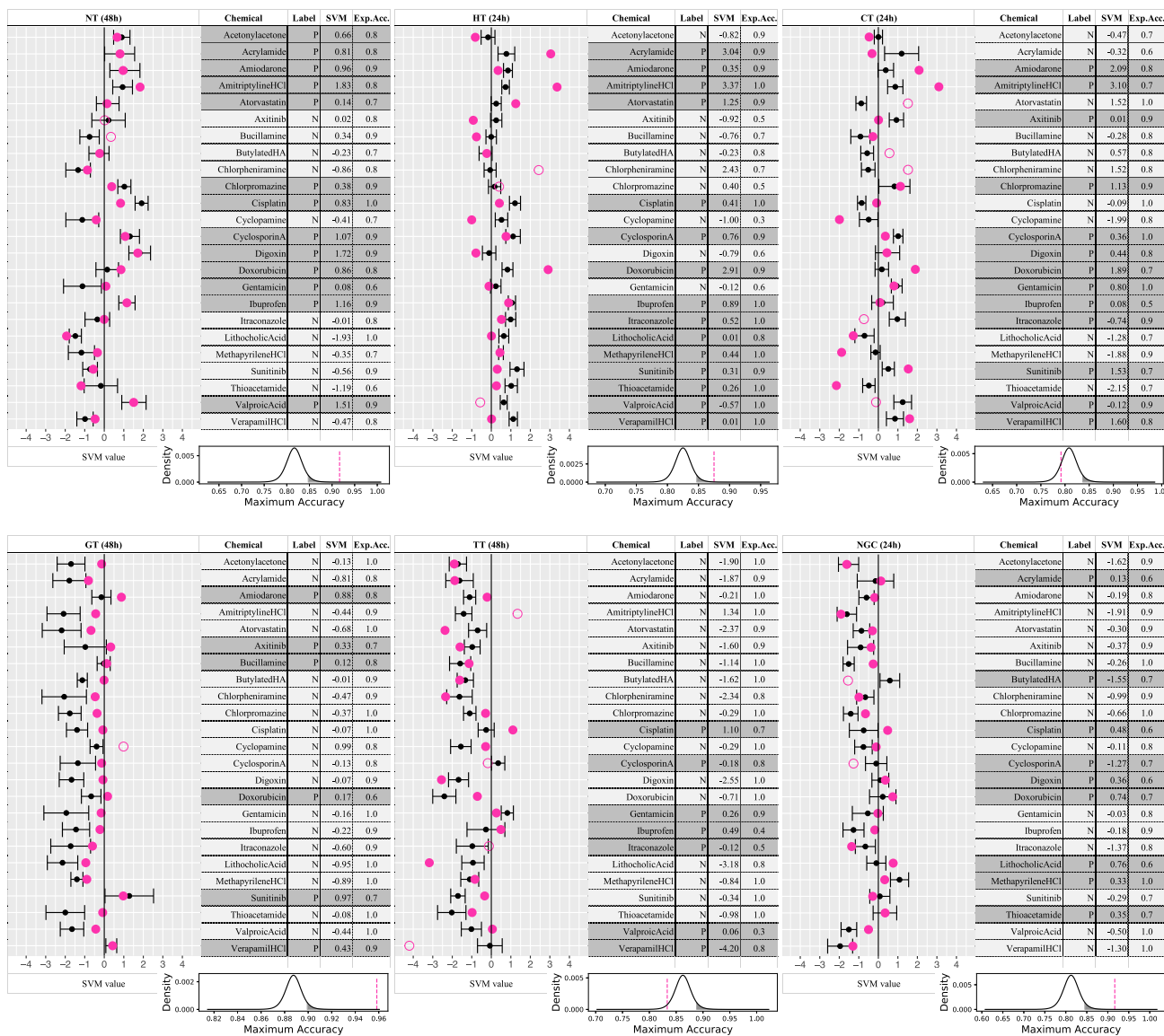
Figure 3. Summary of toxicity category prediction for 24 chemicals using HPS4138 cells

Red dots indicate predicted SVM values of iPS cell data, and filled and open markers indicate true and false predictions, respectively. Black dots and bars indicate the means ± SEM of SVM values for random data. In the tables, the label columns contain prior knowledge regarding whether the chemical shows toxicity (P: positive) or not (N: negative). The SVM column contains the SVM values of the iPS cell data. The expected accuracy (Exp.Acc.) columns contain the prediction accuracy using random data. The graphs below the tables show the probability distribution of the prediction accuracy using random data. Black lines indicate the probability density estimated using the t distribution (degrees of freedom = 9). Black shaded areas represent the upper 5%. Red dashed lines indicate the prediction accuracy using iPS cell data.

endorsed the idea that transcription factors of signal receptors might play a role in interfacing outside signals, such as aryl hydrocarbon or androgen, to activate toxicological AOPs in HepaRG™ cells (Franzosa et al., 2021). In our system, we used 20 transcription factor genes from 5 PCs due to limited resources and cost, but it should be possible to customize the set of genes and PCs to reflect more accurately the specific AOPs in the endpoint organ. To pursue a full coverage of endpoint organs and AOPs, RNA-seq analysis and a library of all 4,033 transcription factor genes for the test chemicals at low cost are needed.

Previous systems using QSAR theory depend on the information of chemicals only and are thus inapplicable to mixtures such as food, Chinese or herbal medicines, and other compounds to assess toxicity as a whole. Our StemPanTox system detects the cellular toxic events of these mixtures, providing the

prediction of holistic cellular reactions, making it a resource for industries that mix independent chemical ingredients, including cosmetic, air-conditioner, and automotive companies who need to assess the toxicity of mixtures in their final or intermediate products, media, emissions, detergents, etc. In fact, more than 100 members from a wide variety of industry-government-academia fields are involved in our non-profit consortium (scChemRISC). By developing products from candidate substances that are predicted to have little toxicity, our system will contribute to industry not only for efficiency but also for human health.

Recently, toxicity reaction differences due to ethnicity, or genome haplotypes, have been widely reported due to the globalization of foods and products among countries. For example, catechin, which is contained in green tea and is widely consumed in Asian countries, is reported to induce severe liver injury in the United States (Oketch-Rabah et al., 2020). The CiRA Foundation (Kyoto, Japan) has announced myiPS cells, a project in which individuals can have their own iPS cells generated and banked. Our StemPanTox system could allow a tailor-made chemical toxicity assessment for these cells to detect individual differences in toxic tolerance for different substances. Ideally, it also has the potential to reduce medical accidents if myiPS cells could be used to diagnosis whether a medication is toxic before receiving the treatment (Easley, 2019; Kim et al., 2019). In general, by performing a battery of toxicity assessments using multiple iPS cell lines from individuals with various haplotypes, our system may contribute to reducing toxicity accidents often caused by a small number of test samples of limited genomic variances.

In conclusion, the largest advantage of our StemPanTox is the ability to perform toxicity hazard assessments for multiple endpoints with high accuracy in a short amount of time and a low cost. We believe that our system will greatly benefit research aimed at amending gaps between current animal studies, which are resource-intensive and show uncertainty for human extrapolation, and NAMs, which are expected to use the latest science, technology, and computational models for improving regulatory toxicology guidelines.

Limitations of the study

Although we showed that human pluripotent stem cells are useful devices capable of qualitatively predicting adult toxicities of 24 chemicals in six categories, it is yet to be confirmed whether other chemicals and categories can be predicted and whether it is extendable to quantitative prediction.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture experiments
- METHOD DETAILS
  - Selection of toxic chemicals
  - Chemical exposures and determination of IC logistic model equations
  - RNA-seq analysis
  - Differentially expressed gene analysis and principal component analysis
  - Gene network construction by Graphical Gaussian Model (GGM)
  - Prediction by support vector machine (SVM)
  - Gene set enrichment analysis
  - Selection of HPS4138 iPS cells
  - Gene expression data from HPS4138 by RT-qPCR
  - Transductive transfer learning by SVM
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

AUTHOR CONTRIBUTIONS

WF conceptualized the research and led the project. JY drafted the manuscript. JY and KK provided the stem cell data. JY, TW, WF, and JKY analyzed the data. HO, KI, MS, TH, SS, HK, and HS selected the toxic chemicals. MO and MKS provided the iPS cell lines.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Abarbanel, J.M., Osiman, A., Frisher, S., Herishanu, Y., and OsimAn, A. (1987). Peripheral neuropathy and cerebellar syndrome associated with amiodarone therapy. Isr. J. Med. Sci. 23, 893–895.

Ali, B.H., Al-Qarawi, A.A., and Mousa, H.M. (2002). The effect of calcium load and the calcium channel blocker verapamil on gentamicin nephrotoxicity in rats. Food. Chem. Toxicol. 40, 1843–1847. https://doi.org/10.1016/s0278-6915(02)00167-9.

Allam, A.A., El-Ghareeb, A.W., Abdul-Hamid, M., Bakery, A.E., Gad, M., and Sabri, M. (2010). Effect of prenatal and perinatal acrylamide on the biochemical and morphological changes in liver of developing albino rat. Arch. Toxicol. 84, 129–141. https://doi.org/10.1007/s00204-009-0475-2.

Allen, T.E.H., Goodman, J.M., Gutsell, S., and Russell, P.J. (2016). A history of the molecular initiating event. Chem. Res. Toxicol. 29, 2060–2070. https://doi.org/10.1021/acs.chemrestox.6b00341.

Allwardt, V., Ainscough, A.J., Viswanathan, P., Sherrod, S.D., McLean, J.A., Haddrick, M., and Pensabene, V. (2020). Translational roadmap for the organs-on-a-chip industry toward broad adoption. Bioengineering 7, 112. https://doi.org/10.3390/bioengineering7030112.

Baroni, E.A., Costa, R.S., Volpini, R., and Coimbra, T.M. (1999). Sodium bicarbonate treatment reduces renal injury, renal production of transforming growth factor-beta, and urinary transforming growth factor-beta excretion in rats with doxorubicin-induced nephropathy. Am. J.

Kidney. Dis. 34, 328–337. https://doi.org/10.1016/s0272-6386(99)70363-x.

Becker, R.A. (2019). New approach methodologies: the brave new world in toxicology & risk assessment. https://www.toxicology.org/groups/rc/ncac/docs/06-Becker.pdf.

Boelaert, J., Schurgers, M., Matthys, E., Daneels, R., van Peer, A., De Beule, K., Woestenborghs, R., Heykants, J., and vAn Peer, A. (1988). Itraconazole pharmacokinetics in patients with renal dysfunction. Antimicrobial agents and chemotherapy 32, 1595–1597. https://doi.org/10.1128/aac.32.10.1595.

Cappon, G.D., Fleeman, T.L., Cook, J.C., and Hurtt, M.E. (2005). Combined treatment potentiates the developmental toxicity of ibuprofen and acetazolamide in rats. Drug. Chem. Toxicol. 28, 409–421. https://doi.org/10.1080/01480540500262805.

Carter, B.S., and Stewart, J.M. (1989). Valproic acid prenatal exposure. Association with lipomyelomeningocele. Clinic. 28, 81–85. https://doi.org/10.1177/000992288902800205.

Clements, M., Millar, V., Williams, A.S., and Kalinka, S. (2015). Bridging functional and structural cardiotoxicity assays using human embryonic stem cell-derived cardiomyocytes for a more comprehensive risk assessment. Toxicol. Sci. 148, 241–260. https://doi.org/10.1093/toxsci/kfv180.

Colombo, C., Crosignani, A., Alicandro, G., Zhang, W., Biffi, A., Motta, V., Corti, F., and Setchell, K.D.R. (2016). Long-Term

ursodeoxycholic acid therapy does not alter lithocholic acid levels in patients with cystic fibrosis with associated liver disease. J. Pediatr. 177, 59–65.e1. https://doi.org/10.1016/j.jpeds.2016.05.008.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/bf00994018.

Doherty, K.R., Wappel, R.L., Talbert, D.R., Trusk, P.B., Moran, D.M., Kramer, J.W., Brown, A.M., Shell, S.A., and Bacus, S. (2013). Multi-parameter in vitro toxicity testing of crizotinib, sunitinib, erlotinib, and nilotinib in human cardiomyocytes. Toxicol. Appl. Pharmacol. 272, 245–255. https://doi.org/10.1016/j.taap.2013.04.027.

Easley, C.A. (2019). Induced pluripotent stem cells (iPSCs) in developmental toxicology. Methods. Mol. Biol. 1965, 19–34. https://doi.org/10.1007/978-1-4939-9182-2_3.

Ekins, S. (2014). Progress in computational toxicology. J. Pharmacol. Toxicol. Methods. 69, 115–140. https://doi.org/10.1016/j.vascn.2013.12.003.

El-Tanbouly, D.M., Wadie, W., and Sayed, R.H. (2017). Modulation of TGF-β/Smad and ERK signaling pathways mediates the anti-fibrotic effect of mirtazapine in mice. Toxicol. Appl. Pharmacol. 329, 224–230. https://doi.org/10.1016/j.taap.2017.06.012.

Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W., and Ahr, H.J. (2005). Comparison of the expression profiles induced by genotoxic and nongenotoxic carcinogens in rat liver. Mutat. Res.

Fund Mol. Mech. Mutagen 575, 61–84. https://doi.org/10.1016/j.mrfmmm.2005.02.004.

Epskamp, S., and Fried, E.I. (2018). A tutorial on regularized partial correlation networks. Psychol. Methods. 23, 617–634. https://doi.org/10.1037/met0000167.

Ferguson, S.A., and Paule, M.G. (1992). Acute effects of chlorpromazine in a monkey operant behavioral test battery. Pharmacol. Biochem. Behav. 42, 333–341. https://doi.org/10.1016/0091-3057(92)90536-o.

Fickert, P., Fuchsbichler, A., Marschall, H.U., Wagner, M., Zollner, G., Krause, R., Zatloukal, K., Jaeschke, H., Denk, H., and Trauner, M. (2006). Lithocholic acid feeding induces segmental bile duct obstruction and destructive cholangitis in mice. Am. J. Pathol. 168, 410–422. https://doi.org/10.2353/ajpath.2006.050404.

Franzosa, J.A., Bonzo, J.A., Jack, J., Baker, N.C., Kothiya, P., Witek, R.P., Hurban, P., Siferd, S., Hester, S., Shah, I., et al. (2021). High-throughput toxicogenomic screening of chemicals in the environment using metabolically competent hepatic cell cultures. NPJ Syst Biol Appl 7, 7. https://doi.org/10.1038/s41540-020-00166-2.

Goodall, C.R. (1991). Graphical models in applied multivariate statistics. Technometrics 33, 476–478. https://doi.org/10.1080/00401706.1991.10484880.

Grimm, D. (2019). EPA plan to end animal testing splits scientists. Science (American Association for the Advancement of Science) 365, 1231. https://doi.org/10.1126/science.365.6459.1231.

Hendry, P.J., Taichman, G.C., Taichman, S.J., and Keon, W.J. (1988). Human myocardial responses to antibiotics: gentamicin, tobramycin and cephalothin. Can. J. Cardiol. 4, 219–222.

Hirose, M., Masuda, A., Hasegawa, R., Wada, S., and Ito, N. (1990). Regression of butylated hydroxyanisole (BHA)-induced hyperplasia but not dysplasia in the forestomach of hamsters. Carcinogenesis 11, 239–244. https://doi.org/10.1093/carcin/11.2.239.

Humes, H.D., Jackson, N.M., O'Connor, R.P., Hunt, D.A., and White, M.D. (1985). Pathogenetic mechanisms of nephrotoxicity: insights into cyclosporine nephrotoxicity. Transplant. Proc. 17, 51–62.

Jordan, B., Pasquier, Y., Schnider, A., and Schnider, A. (2004). Neurological improvement and rehabilitation potential following toxic myelopathy due to intrathecal injection of doxorubicin. Spinal. Cord. 42, 371–373. https://doi.org/10.1038/sj.sc.3101592.

Kameoka, S., Babiarz, J., Kolaja, K., and Chiao, E. (2014). A high-throughput screen for teratogens using human pluripotent stem cells. Toxicol. Sci. 137, 76–90. https://doi.org/10.1093/toxsci/kft239.

Kim, T.W., Che, J.H., and Yun, J.W. (2019). Use of stem cells as alternative methods to animal experimentation in predictive toxicology. Regul. Toxicol. Pharmacol. 105, 15–29. https://doi.org/10.1016/j.yrtph.2019.03.016.

Kimura, T., Kuramochi, S., Katayama, T., Yoshikawa, T., Yamada, T., Ueda, Y., and Okada, Y. (2008). Amiodarone-related pulmonary mass

and unique membranous glomerulonephritis in a patient with valvular heart disease: diagnostic pitfall and new findings. Pathol. Int. 58, 657–663. https://doi.org/10.1111/j.1440-1827.2008.02286.x.

Kitazawa, S. (1993). Studies on initiating activity of secondary bile acids for rat hepatocarcinogenesis. Hokkaido. Igaku. Zasshi. 68, 110–120.

Kitazawa, S., Denda, A., Tsutsumi, M., Tsujiuchi, T., Hasegawa, K., Tamura, K., Maruyama, H., and Konishi, Y. (1990). Enhanced preneoplastic liver lesion development under 'selection pressure' conditions after administration of deoxycholic or lithocholic acid in the initiation phase in rats. Carcinogenesis 11, 1323–1328. https://doi.org/10.1093/carcin/11.8.1323.

Kleinstreuer, N.C., Smith, A.M., West, P.R., Conard, K.R., Fontaine, B.R., Weir-Hauptman, A.M., Palmer, J.A., Knudsen, T.B., Dix, D.J., Donley, E.L., and Cezar, G.G. (2011). Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. Toxicol. Appl. Pharmacol. 257, 111–121. https://doi.org/10.1016/j.taap.2011.08.025.

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2016). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.

Kreft, B., de Wit, C., Marre, R., and Sack, K. (1991). Experimental studies on the nephrotoxicity of amphotericin B in rats. Journal 28, 271–281. https://doi.org/10.1093/jac/28.2.271.

Kugler, J., Tharmann, J., Chuva de Sousa Lopes, S.M., Kemler, R., Luch, A., and Oelgeschläger, M. (2015). A Bmp reporter transgene mouse embryonic stem cell model as a tool to identify and characterize chemical teratogens. Toxicol. Sci. 146, 374–385. https://doi.org/10.1093/toxsci/kfv103.

Laschinski, G., Vogel, R., and Spielmann, H. (1991). Cytotoxicity test using blastocyst-derived euploid embryonal stem cells: a new approach to in vitro teratogenesis screening. Reprod. Toxicol. 5, 57–64. https://doi.org/10.1016/0890-6238(91)90111-r.

Lee, R.Z., Hardiman, O., and O'Connell, P.G. (2002). Ibuprofen-induced aseptic meningoencephalitis. Rheumatology 41, 353–355. https://doi.org/10.1093/rheumatology/41.3.353.

Lin, Z., and Will, Y. (2012). Evaluation of drugs with specific organ toxicities in organ-specific cell lines. Toxicol. Sci. 126, 114–127. https://doi.org/10.1093/toxsci/kfr339.

LoPachin, R.M., Ross, J.F., Reid, M.L., Das, S., Mansukhani, S., and Lehning, E.J. (2002). Neurological evaluation of toxic axonopathies in rats: acrylamide and 2, 5-hexanedione. Neurotoxicology 23, 95–110. https://doi.org/10.1016/s0161-813x(02)00003-7.

Low, L.A., Mummery, C., Berridge, B.R., Austin, C.P., and Tagle, D.A. (2021). Organs-on-chips: into the next decade. Nat. Rev. Drug. Discov. 20, 345–361. https://doi.org/10.1038/s41573-020-0079-3.

Luz, A.L., and Tokar, E.J. (2018). Pluripotent stem cells in developmental toxicity testing: a review of methodological advances. Toxicol. Sci. 165, 31–39. https://doi.org/10.1093/toxsci/kfy174.

Maheut-Bosser, A., Brembilla-Perrot, B., Hanesse, B., Piffer, I., Paille, F., and MAheut-Bosser, A. (2006). Cognitive impairment induced by digoxin intake in patients older than 65 years. Ann. Cardiol. Angeiol. 55, 246–248. https://doi.org/10.1016/j.ancard.2006.01.001.

Marasco, W.A., Gikas, P.W., Azziz-Baumgartner, R., Hyzy, R., Eldredge, C.J., and Stross, J. (1987). Ibuprofen-associated renal dysfunction. Pathophysiologic mechanisms of acute renal failure, hyperkalemia, tubular necrosis, and proteinuria. Arch. Intern. Med. 147, 2107–2116. https://doi.org/10.1001/archinte.147.12.2107.

Matsuda, M., Yamanaka, Y., Uemura, M., Osawa, M., Saito, M.K., Nagahashi, A., Nishio, M., Guo, L., Ikegawa, S., Sakurai, S., et al. (2020). Recapitulating the human segmentation clock with pluripotent stem cells. Nature 580, 124–129. https://doi.org/10.1038/s41586-020-2144-9.

McGovern, B., Garan, H., and Ruskin, J.N. (1986). Precipitation of cardiac arrest by verapamil in patients with Wolff-ParkinsonWhite syndrome. Ann. Intern. Med. 104, 791. https://doi.org/10.7326/0003-4819-104-6-791.

Miller, L.W. (2002). Cardiovascular toxicities of immunosuppressive agents. Am. J. Transplant. 2, 807–818. https://doi.org/10.1034/j.1600-6143.2002.20902.x.

Mizukami, S., Watanabe, Y., Nakajima, K., Hasegawa-Baba, Y., Jin, M., Yoshida, T., and Shibutani, M. (2017). Downregulation of TMEM70 in rat liver cells after hepatocarcinogen treatment related to the warburg effect in hepatocarcinogenesis producing GST-P-expressing proliferative lesions. Toxicol. Sci. 159, 211–223. https://doi.org/10.1093/toxsci/kfx131.

Mladosievicova, B., Foltinová, A., Petrásová, H., Petrásová, H., Bernadic, M., Hulín, I., Bernadic, M., Hulín, I., and Foltinova, A. (2001). Signal-averaged electrocardiography in survivors of Hodgkin's disease treated with and without dexrazoxane. Neoplasma 48, 61–65.

Nakatsuji, N. (2005). Establishment and manipulation of monkey and human embryonic stem cell lines for biomedical research. Ernst. Schering. Res. Found. Workshop. 54, 15–26. https://doi.org/10.1007/3-540-37644-5_2.

New Approach Methods Work Plan: Reducing Use of Animals in Chemical Testing (U.S. Environmental Protection Agency). https://www.epa.gov/.

Ohno, I. (2004). Drug induced nephrotic syndrome. Nihon Rinsho 62, 1919–1924.

Oketch-Rabah, H.A., Roe, A.L., Rider, C.V., Bonkovsky, H.L., Giancaspro, G.I., Navarro, V., Paine, M.F., Betz, J.M., Marles, R.J., Casper, S., et al. (2020). United States Pharmacopeia (USP) comprehensive review of the hepatotoxicity of green tea extracts. Toxicol Rep 7, 386–402. https://doi.org/10.1016/j.toxrep.2020.02.008.

Okuyan, H., and Altin, C. (2013). Heart failure induced by itraconazole. Indian J. Pharmacol. 45, 524. https://doi.org/10.4103/0253-7613.117751.

Ozkok, A., and Edelstein, C.L. (2014). Pathophysiology of cisplatin-induced acute kidney injury. BioMed Res. Int. *2014*, 967826. https://doi.org/10.1155/2014/967826.

Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Sandercock, P., Macleod, M., Mignini, L.E., Jayaram, P., and Khan, K.S. (2007). Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ *334*, 197. https://doi.org/10.1136/bmj.39048.407928.be.

Pilling, R.F., Sabri, K., and Chaudhuri, P.R. (2005). Non-traumatic sixth nerve palsy in a young patient. Strabismus *13*, 11–13. https://doi.org/10.1080/09273970590890930.

Qu, W., Crizer, D.M., DeVito, M.J., Waidyanatha, S., Xia, M., Houck, K., and Ferguson, S.S. (2021). Exploration of xenobiotic metabolism within cell lines used for Tox21 chemical screening. Toxicol. Vitro *73*, 105109. https://doi.org/10.1016/j.tiv.2021.105109.

Raza, M., al-Shabanah, O.A., al-Bekairi, A.M., and Qureshi, M. (2000). Pathomorphological changes in mouse liver and kidney during prolonged valproate administration. Int. J. Tissue. React. *22*, 15–21.

Rea, T.D., Siscovick, D.S., Psaty, B.M., Pearce, R.M., Raghunathan, T.E., Whitsel, E.A., Cobb, L.A., Weinmann, S., Anderson, G.D., Arbogast, P., and Lin, D. (2003). Digoxin therapy and the risk of primary cardiac arrest in patients with congestive heart failure. J. Clin. Epidemiol. *56*, 646–650. https://doi.org/10.1016/s0895-4356(03)00075-1.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science (American Association for the Advancement of Science) *293*, 1089–1093. https://doi.org/10.1126/science.1063443.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic. Acids. Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

Ruuskanen, I., Kilpeläinen, H.O., and Riekkinen, P.J. (2009). Side effects of sodium valproate during long-term treatment in epilepsy. Acta. Neurol. Scand. *60*, 125–128. https://doi.org/10.1111/j.1600-0404.1979.tb02959.x.

Saito, K., Daitoku, K., Fukunaga, H., Matsuoka, T., Biroh, S., Kakei, M., Kashima, T., and Tanaka, H.

(1985). Chlorpromazine-induced cardiomyopathy in rats. Heart. Vessels. *1*, 283–285. https://doi.org/10.1007/bf02072410.

Scholz, G., Genschow, E., Pohl, I., Bremer, S., Paparella, M., Raabe, H., Southee, J., and Spielmann, H. (1999). Prevalidation of the embryonic stem cell test (EST) - a new in vitro embryotoxicity test. Toxicol. Vitro *13*, 675–681. https://doi.org/10.1016/s0887-2333(99)00046-6.

Schulpen, S.H.W., Theunissen, P.T., Pennings, J.L., and Piersma, A.H. (2015). Comparison of gene expression regulation in mouse- and human embryonic stem cell assays during neural differentiation and in response to valproic acid exposure. Reprod. Toxicol. *56*, 77–86. https://doi.org/10.1016/j.reprotox.2015.06.043.

Schultz, T.W. (1999). Structure-toxicity relationships for benzenes evaluated with Tetrahymena pyriformis. Chem. Res. Toxicol. *12*, 1262–1267. https://doi.org/10.1021/tx9900730.

Seiler, A.E.M., and Spielmann, H. (2011). The validated embryonic stem cell test to predict embryotoxicity in vitro. Nat. Protoc. *6*, 961–978. https://doi.org/10.1038/nprot.2011.348.

Seong, J., Han, K.H., Park, Y.N., Nam, S.H., Kim, S.H., Keum, W.S., and Kim, K.S. (2003). Lethal hepatic injury by combined treatment of radiation plus chemotherapy in rats with thioacetamide-induced liver cirrhosis. Int. J. Radiat. Oncol. Biol. Phys. *57*, 282–288. https://doi.org/10.1016/s0360-3016(03)00540-6.

Sha, S.H., and Schacht, J. (1999). Salicylate attenuates gentamicin-induced ototoxicity. Lab. Invest. *79*, 807–813.

Sorich, M.J., Rowland, A., Kichenadasse, G., Woodman, R.J., and Mangoni, A.A. (2016). Risk factors of proteinuria in renal cell carcinoma patients treated with VEGF inhibitors: a secondary analysis of pooled clinical trial data. British journal of cancer *114*, 1313–1317. https://doi.org/10.1038/bjc.2016.147.

Srinivasa, V., Gerner, P., Haderer, A., Abdi, S., Jarolim, P., and Wang, G.K. (2003). The relative toxicity of amitriptyline, bupivacaine, and levobupivacaine administered as rapid infusions in rats. Anesthesia *97*, 91–95. https://doi.org/10.1213/01.ane.0000065600.15574.ab.

Sudoh, Y., Desai, S.P., Haderer, A.E., Sudoh, S., Gerner, P., Anthony, D.C., Girolami, U.D., Wang, G.K., HAderer, A., and Degirolami, U. (2004). Neurologic and histopathologic evaluation after high-volume intrathecal amitriptyline. Reg. Anesth. Pain. Med. *29*, 434–440. https://doi.org/10.1016/j.rapm.2004.06.008.

Suzuki, N., Ando, S., Yamashita, N., Horie, N., and Saito, K. (2011). Evaluation of novel high-

throughput embryonic stem cell tests with new molecular markers for screening embryotoxic chemicals in vitro. Toxicol. Sci. *124*, 460–471. https://doi.org/10.1093/toxsci/kfr250.

Takahashi, H., Qin, X.Y., Sone, H., and Fujibuchi, W. (2018). Stem cell-based methods to predict developmental chemical toxicity. Methods Mol. Biol. *1800*, 475–483. https://doi.org/10.1007/978-1-4939-7899-1_21.

Topical Scientific Workshop - New Approach Methodologies in Regulatory Science (2016), 19-20 April 2016 | Helsinki, Finland. https://echa.europa.eu/view-article/-/journal_content/title/topical-scientific-workshop-new-approach-methodologies-in-regulatory-science.

Uibel, F., Mühleisen, A., Köhle, C., Weimer, M., Stummann, T.C., Bremer, S., and Schwarz, M. (2010). ReProGlo: a new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. Reprod. Toxicol. *30*, 103–112. https://doi.org/10.1016/j.reprotox.2009.12.002.

Vapnik, V. (2006). Estimation of Dependences Based on Empirical Data (Springer Science).

Welch, M.D., DePace, A.H., Verma, S., Iwamatsu, A., and Mitchison, T.J. (1997). The human arp2/3 complex is composed of evolutionarily conserved subunits and is localized to cellular regions of dynamic actin filament assembly. J. Cell Biol. *138*, 375–384. https://doi.org/10.1083/jcb.138.2.375.

West, P.R., Weir, A.M., Smith, A.M., Donley, E.L.R., and Cezar, G.G. (2010). Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. Toxicol. Appl. Pharmacol. *247*, 18–27. https://doi.org/10.1016/j.taap.2010.05.007.

Xie, Y., Nishi, S., Iguchi, S., Imai, N., Sakatsume, M., Saito, A., Ikegame, M., Iino, N., Shimada, H., Ueno, M., et al. (2001). Expression of osteopontin in gentamicin-induced acute tubular necrosis and its recovery process. Kidney Int. *59*, 959–974. https://doi.org/10.1046/j.1523-1755.2001.059003959.x.

Yamane, J., Aburatani, S., Imanishi, S., Akanuma, H., Nagano, R., Kato, T., Sone, H., Ohsako, S., and Fujibuchi, W. (2016). Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells. Nucleic acids research *44*, 5515–5528. https://doi.org/10.1093/nar/gkw450.

Zhang, Y., Xu, Y.Y., Sun, W.J., Zhang, M.H., Zheng, Y.F., Shen, H.M., Yang, J., and Zhu, X.Q. (2016). FBS or BSA inhibits EGCG induced cell death through covalent binding and the reduction of intracellular ROS production. BioMed Res. Int. *2016*, 1–8. https://doi.org/10.1155/2016/5013409.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Vitronectin (VTN-N)	Thermo Fisher Scientific	Cat # A14700
CultureSure® Y-27632	FUJIFILM WAKO	Cat # 036-24023
Dimethyl sulfoxide, for molecular biology	Sigma-Aldrich	Cat # D8418
2,5-Hexanedione	Sigma-Aldrich	Cat # 00770-50ml
Acrylamide(monomer)	NACALAI TESQUE	Cat # 06114-24
Amiodarone Hydrochloride	TCI (Tokyo Chemical Industry)	Cat # A2530
Amitriptyline hydrochloride	Sigma-Aldrich	Cat # A8404
Atorvastatin Calcium Salt Trihydrate	TCI (Tokyo Chemical Industry)	Cat # A2476
Axitinib	Selleck Chemicals	Cat # S1005
Bucillamine	Toronto Research Chemicals	Cat # B689375
Butylated hydroxyanisole	Sigma-Aldrich	Cat # B1253-5G
(+)-Chlorpheniramine Maleate	FUJIFILM WAKO	Cat # 030-13271
Chlorpromazine Hydrochloride	FUJIFILM WAKO	Cat # 033-10581
cis-Diammineplatinum(II) dichloride crystalline	Sigma-Aldrich	Cat # P4394-250MG
Cyclopamine hydrate	SIGMA	Cat # C4116-1MG
Ciclosporin A	FUJIFILM WAKO	Cat # 031-24931
Digoxin	Sigma-Aldrich	Cat # D6003-1G
Doxorubicin (hydrochloride)	Cayman Chemical	Cat # 15007
Gentamicin sulfate	Sigma-Aldrich	Cat # G3632-1G
Ibuprofen	FUJIFILM WAKO	Cat # 098-02641
Itraconazole	Sigma-Aldrich	Cat # I6657-100MG
Lithocholic acid	Sigma-Aldrich	Cat # L6250-10G
Methapyrilene HCl	Sigma-Aldrich	Cat # 442641
Sunitinib malate	Tocris Bioscience	Cat # 510-86011
Thioacetamide	Sigma-Aldrich	Cat # 172502-25G
Valproic acid	SIGMA	Cat # PHR1061-1G
Verapamil Hydrochloride	Sigma-Aldrich	Cat # V4629-1G
RNasin® Plus RNase Inhibitor, 2500u	Promega	Cat # N2611
Dimethoate Standard	FUJIFILM WAKO	Cat # 041-15843
Arsenic (III) oxide	Sigma-Aldrich	Cat # 311383-125G
Quinidine hydrochloride monohydrate technical grade	Sigma-Aldrich	Cat # Q0750-5G
Critical commercial assays		
CellTiter-Glo® Luminescent Cell Viability Assay	Promega	Cat # G7571
RNeasy Mini Kit	QIAGEN	Cat # 74106
RNase-Free DNase Set (50)	QIAGEN	Cat # 79254
TruSeq Stranded mRNA Library Prep (48 Samples)	Illumina	Cat # 2020594
TruSeq RNA Single Indexes Set A (12 Indexes, 48 Samples)	Illumina	Cat # 20020492

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TruSeq RNA Single Indexes Set B (12 Indexes, 48 Samples)	illumina	Cat # 20020493
HiSeq 3000/4000 SR Cluster Kit	illumina	Cat # GD-410-1001
HiSeq 3000/4000 SBS Kit (50 Cycles)	illumina	Cat # FC-410-1001
PrimeScript™ RT Reagent Kit (Perfect Real Time)	TAKARA BIO	Cat # RR037B
KAPA SYBR Fast qPCR Kit (Universal qPCR kit)	KAPA BIOSYSTEMS	Cat # KK4602
Deposited data		
RNA-seq	This study	GEO: GSE188203
Experimental models: Cell lines		
KhES-3 cell line (XY)	(Nakatsuji, 2005)	NA
HPS4138 cell line (XY)	RIKEN BRC	NA
HPS4234 cell line (XY)	RIKEN BRC	NA
HPS4046 cell line (XY)	RIKEN BRC	NA
Oligonucleotides		
Primers for qRT-PCR	This study	https://www.stemcellinformatics.org/toxicology/
Software and algorithms		
alvaDesc	Affinity Science	https://www.alvascience.com/alvadesc/
CORINA Classic	mn-am.com	https://www.mn-am.com/online_demos/corina_demo/
Other		
The source code of R analysis script and the data supporting the findings of this study	This study	https://www.stemcellinformatics.org/toxicology/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Professor Wataru Fujibuchi (wfujibuchi@gmail.com).

Materials availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact.

Data and code availability

- The accession number for RNA-seq data reported in this paper is GEO: GSE188203.

- The source code of R analysis script is available from: https://www.stemcellinformatics.org/toxicology/.

- The data supporting the findings of this study are available from: https://www.stemcellinformatics.org/toxicology/.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture experiments

The KhES-3 cell line was established at and provided by Kyoto University (Nakatsuji, 2005). The protocol of this study was reviewed by the Ethics Committee of CiRA in accordance with the "Guidelines for Derivation

and Utilization of Human Embryonic Stem Cells" by the Ministry of Education, Culture, Sports, Science and Technology, Japan. The iPS cell lines were established from healthy Japanese donors at CiRA, Kyoto University, and were approved for use by the Ethics Committee of Kyoto University.

Since it has been reported that the toxicity of antioxidants such as catechin is suppressed in the presence of albumin (Zhang et al., 2016), maintenance culture was carried out for all cell lines including human ES cells using albumin-free Essential 8 Medium (Thermo Fisher Scientific) in six-well feeder-free culture dishes coated with 5 µg/mL vitronectin (VTN-N; Thermo Fisher Scientific). When seeding the cells, 10 µM CultureSure® Y-27632 (FUJIFILM WAKO) was added, and medium exchange on day 1 and thereafter was performed without Y-27632.

METHOD DETAILS

Selection of toxic chemicals

Twenty-four chemicals were selected and mainly included neurotoxins, hepatotoxins, cardiotoxins, nephrotoxins, and non-genotoxic carcinogens (Table 1). The presence or absence of toxicity was determined mainly on the basis of information regarding toxicity and assorted disorders and diseases available at PubChem (https://pubchem.ncbi.nlm.nih.gov/). With respect to neurotoxicity, among the chemicals previously reported in the literature, those having only developmental toxicity were classified as 'negative,' as the present study targeted adult toxicity. In addition, when determining the presence or absence of hepatotoxicity, chemicals with DILI rank $\geq$ 3 were considered hepatotoxic chemicals; with regard to others, those with reliable reports of liver diseases were considered hepatotoxic. As for cardiotoxicity, chemicals that have been reported to be associated with heart disease were considered cardiotoxic. With regard to the kidney, due to its diverse and complex structure, the area of damage was divided into two sites, namely, the glomeruli and renal tubules. Chlorpheniramine and cyclopamine were the chemicals judged to be completely 'negative' and belonged to none of the toxicity categories examined in the present study. On the other hand, 19 chemicals had multiple overlapping toxicities, whereas acetonylacetone, bucillamine, and butylated HA had only one toxicity (Abarbanel et al., 1987; Ali et al., 2002; Allam et al., 2010; Baroni et al., 1999; Boelaert et al., 1988; Cappon et al., 2005; Carter and Stewart, 1989; Clements et al., 2015; Colombo et al., 2016; Doherty et al., 2013; El-Tanbouly et al., 2017; Ellinger-Ziegelbauer et al., 2005; Ferguson and Paule, 1992; Fickert et al., 2006; Hendry et al., 1988; Hirose et al., 1990; Humes et al., 1985; Jordan et al., 2004; Kimura et al., 2008; Kitazawa, 1993; Kitazawa et al., 1990; Kreft et al., 1991; Lee et al., 2002; Lin and Will, 2012; LoPachin et al., 2002; Maheut-Bosser et al., 2006; Marasco et al., 1987; McGovern et al., 1986; Miller, 2002; Mizukami et al., 2017; Mladosievicova et al., 2001; Ohno, 2004; Okuyan and Altin, 2013; Ozkok and Edelstein, 2014; Pilling et al., 2005; Raza et al., 2000; Rea et al., 2003; Ruuskanen et al., 1979; Saito et al., 1985; Seong et al., 2003; Sha and Schacht, 1999; Sorich et al., 2016; Srinivasa et al., 2003; Sudoh et al., 2004; Xie et al., 2001).

Chemical exposures and determination of IC logistic model equations

DMSO or water was used as solvent (vehicle) for the 24 chemicals based on known information (https://pubchem.ncbi.nlm.nih.gov/). For each of the 24 toxic chemicals, a stock solution was prepared with the highest soluble concentration. First, in order to perform the ATP assay to determine the exposure doses for testing, we performed 10 serial three-fold dilutions of the stock solution; the prepared exposure solution was added to the cells (i.e., exposure) so that the concentration in the medium was 0.1% of the exposure solution. The cells were cultured on 96-well black/clear flat bottom TC-treated plates (Falcon), 8,000 cells were seeded, and the medium was exchanged on day 1. The cells were exposed to chemicals on day 2. No medium exchange was performed after exposure, and the ATP assay was performed 48 h after exposure. For 100 µL of culture medium, 100 µL of CellTiter-Glo® Luminescent Cell Viability Assay (Promega Corporation) was added, and emitted light was measured with a 2104 EnVision Multilabel Plate Reader (PerkinElmer). From four luminescence measurements for 10 concentrations and a blank, regression analysis was performed by fitting the three-parameter log-logistic model with the R-4.0.5 drc package, and IC0.1 and IC50 values were obtained (Tables S1 and S2). IC50 values were plotted using the ggplot2 package (Figure 1B).

RNA-seq analysis

On day 2 after seeding, the cells were exposed to the chemicals. For each of the 24 chemicals, the exposure dose was set between IC0.1 and IC50 depending on the degree of cell death. Using this value as the

maximum exposure dose, five serial two-fold dilutions (1/1, 1/2, 1/4, 1/8, 1/16) were performed, and a total of six doses including a solvent-only control (vehicle) were used. After exposure, no medium exchange was performed, and samples were obtained at two time points (24 h and 48 h) with two repeats, i.e., a total of 24 x 6 x 2 x 2=576 samples. After RNA purification using an RNeasy Mini Kit (QIAGEN), sequence libraries were prepared for each sample using TruSeq Stranded mRNA Library Prep/TruSeq RNA Single Indexes Set A & Set B (Illumina, Inc.). For sequencing, high-throughput sequencing was performed using HiSeq4000 (Illumina, Inc.). We used bowtie-2.2.5 with the option "–very-sensitive-local" to map the obtained Illumina reads to Ensemble GRCh38r100 human cDNA and ncRNA sequences, added up the reads for each gene using MAPQ $\geq$ 1 transcript, and obtained average counts of 28,652,809 and 28,291,682 reads for each sample at 24 h and 48 h, respectively. From these, we selected only transcription factor-related genes included in the Gene Ontology GO:0006351 using BioMart (4,032 genes). These genes were filtered using the statistical analysis language R-4.0.5 package edgeR (Robinson et al., 2010) (https://www.r-project.org/) with the filterByExpr function min.count = 30, min.total.count = 0, and then normalized to log2 counts per million (logCPM) using the voom function. Furthermore, the removeBatchEffect function was used to eliminate batch effects.

Differentially expressed gene analysis and principal component analysis

At 24 h and 48 h, for a total of 122 groups including 120 conditions (five concentrations each for 24 chemicals) and two solvent conditions (DMSO or water), we used a linear model fitting with the lmFit function of the limma package (Ritchie et al., 2015) in R and moderated t-statistics with eBayes to analyze differentially expressed genes (DEGs) with respect to gene expression levels in terms of the log-fold–change (LFC) between 24 chemicals and their corresponding solvent, and created a heatmap of genes with an LFC > 1 and FDR (false discovery rate) < 0.01 using the pheatmap package in R (Figures 1C and S2). In addition, using the LFC values obtained for 120 conditions, PCA was performed for each of the two time points (24 h, 48 h) using the prcomp function in R (Figures 1D and S3).

Gene network construction by Graphical Gaussian Model (GGM)

Based on results obtained in the aforementioned PCA, a total of 20 genes (two genes each with the top positive and negative loading values in the first to fifth PCs) were used to construct gene networks. To estimate the GGM for each of the 24 chemicals, we used the aforementioned LFC values to calculate the sparse partial correlation coefficient network with L1 graphical lasso using EBICglasso in the R package qgraph (https://cran.r-project.org/web/packages/qgraph/qgraph.pdf) (Epskamp and Fried, 2018). For model fitting, regular BIC with gamma = 0 was used, and regularization of sparsity was tried 1000 different ways with nlambda = 1000 for estimation. In the estimation, in order to avoid the problem of covariance matrices failing to be positive definite, calculations were performed with checkPD = FALSE (Figures 2A, S4, and S5).

Prediction by support vector machine (SVM)

The SVM program and protocol used in the present study were adopted according to the report of Takahashi et al. (2018) Four kernel functions were used: linear, polynomial, RBF, and maximum entropy, and parameter types and combinations were calculated according to the above report (Takahashi et al., 2018). In the calculation, 190 values of partial correlation coefficients among 20 genes in the GGM for each of the 24 chemicals were used as input data, and using leave-one-out-cross-validation (LOOCV), the genes were ranked using the two-sample t-test (two-sided) in each iteration, and the maximum accuracy and the maximum AUC (area under the receiver operating characteristic curve) to achieve the maximum accuracy were recorded, varying the number of values from 1 to 190. For comparison, LFC data for the transcription factor genes at five concentrations before calculating the GGM (24 x 5 x 3,200 or 3,255 values in total) were used as input data, and the maximum accuracy and the corresponding AUC value were recorded in a similar manner. In addition, to compare with predictions based on QSAR, 5,666 molecular descriptors were created using alvaDesc (Affinity Science) (https://www.alvascience.com/alvadesc/). We obtained information from PubChem DB (https://pubchem.ncbi.nlm.nih.gov/) 3D Conformer data regarding 20 of the 24 chemicals; for the four other chemicals for which information could not be obtained from PubChem DB (cisplatin, cyclosporin A, digoxin, and gentamicin), the SMILES format was converted into 3D molecular descriptors using CORINA Classic (https://www.mn-am.com/online_demos/corina_demo) and entered into alvaDesc.

Gene set enrichment analysis

We created LFC data for all genes by performing the same preprocess as for the transcription factor genes (21,650 and 22,298 genes for 24 h and 48 h, respectively) and divided them into high-dose (1/1/, 1/2) and low-dose (1/8, 1/16) groups to perform GSEA for each group using the R package fgsea (Korotkevich et al., 2016). As for the gene sets used, among the MSigDB Collections provided by GSEA (https://www.gsea-msigdb.org/gsea/index.jsp), we used 50 hallmark gene sets. The heatmap was generated with FDR-adjusted p values obtained using the fgsea package.

Selection of HPS4138 iPS cells

The measurement of the differentiation potential to the three germ layers was performed according to a previous report (Matsuda et al., 2020), where we examined the expression ratio of two marker genes for each layer (*PAX6*, *SOX2*, *BRA*, *NCAM*, *SOX17*, and *FOXA2*) by means of fluorescence activated cell sorting. Among the ranked Japanese male cell lines derived from healthy individuals, we used the top 20 cell lines according to their total ratios (Matsuda et al., 2020). These cell lines were kept in maintenance culture with StemFit AK02N medium (Ajinomoto) and then cultured for two passages in maintenance culture using Essential 8 Medium (Thermo Fisher Scientific) as in the case of KhES-3. Among the 20 cell lines, one underwent cell death, and the remaining 19 were subjected to pre-screening using 20 chemicals with a wide range of toxicities (valproic acid, cyclopamine, acrylamide, acetonylacetone, chlorpromazine, chlorpheniramine, atorvastatin, amiodarone, verapamil HCl, dimethoate, arsenic trioxide, quinidine, axitinib, doxorubicin, gentamicin, ibuprofen, lithocholic acid, thioacetamide, butylated HA, and methapyrilene HCl) by comparing the ATP assay results with those for ES cells. The details of the ATP assay are described in "Chemical exposures and determination of IC logistic model equations". For exposure concentrations, the IC50 determined with KhES-3 was used, and the growth rate of human iPS cells was examined. Among the candidate cell lines, the top three cell lines (HPS4138, HPS4234, and HPS4046) whose growth rates at IC50 correlated well with that of KhES-3 were selected, and again, the growth rate at IC50 was confirmed by the ATP assay using 20 of the 24 toxic chemicals examined in the present study (Table S6) to select the cell line with the largest correlation coefficient, i.e., HPS4138.

Gene expression data from HPS4138 by RT-qPCR

The five serial exposure concentrations of the 24 chemicals for iPS cells were determined by the ATP assay, as described in "Chemical exposures and determination of IC logistic model equations" with respect to ES cells. For each of the 20 genes at 24 h and 48 h used for the construction of the GGM for ES cells, the primer sequence pair was designed using Primer 3 (version 0.4.0) based on the human cDNA sequence data obtained from Ensembl GRCh38r100, and the obtained primer sequence pair was synthesized (Hokkaido System Science). We confirmed whether the target PCR product could be obtained from the primer sequence pair according to the product size determined by electrophoresis. On day 2 after seeding, HPS4138 cells were exposed to the 24 chemicals. With regard to exposure concentrations, time points, and repeat experiments, we followed the experiments performed with ES cells. After purification using an RNeasy Mini Kit, RNA was transcribed to cDNA using a PrimeScript^{TM} RT Reagent Kit (Perfect Real Time) (TaKaRa), and the synthesized primers for qRT-PCR and KAPA SYBR Fast qPCR Kit (KAPA BIOSYSTEMS) were used to perform qRT-PCR with StepOnePlus (Applied Biosystems). ΔC_T values of the resulting genes were obtained by subtracting from the C_T values the value of the internal control (*GAPDH* gene). The average value of two repeated measurements at each concentration was determined. In addition, for solvents (DMSO and water), ΔC_T values were obtained by subtracting the value of *GAPDH* gene, and the average of all values was determined. The difference, $\Delta\Delta C_T$, between the ΔC_T value at each concentration and that of the solvent was determined and is referred to as LFC. Covariance matrices among the 20 genes were calculated and used as input data in the construction of the GGM.

Transductive transfer learning by SVM

A total of 380 edges in the GGM of ES cells and the GGM of iPS cells were used to perform the prediction of chemicals in the six toxicity categories under conditions similar to the SVM protocol described in "Prediction by support vector machine (SVM)". For learning, labels of the 24 chemicals in ES cells were provided, whereas none of the labels of the 24 chemicals in iPS cells were provided (i.e., zero); prediction was performed via transductive transfer learning. As in the case of ES cells, prediction was also performed by replacing with uniform random numbers 10 times only the values of the 24 x 190 input data for iPS cells, and

similar to ES cells, the one-sample t-test (one-sided) was used to assess the maximum accuracy and the corresponding AUC value.

QUANTIFICATION AND STATISTICAL ANALYSIS

In this paper, we used accuracy (=(TP+TN)/(TP+TN+FP+FN) where TP: true positive, TN: true negative, FP: false positive, and FN: false negative) and AUC (the area under the receiver operating characteristic curve) for the statistical measurement of prediction performance. To statistically evaluate the maximum accuracy and AUC, uniform random numbers that have the same dimensions as the actual input data were generated 10 times, and the maximum accuracy with the maximum AUC were recorded. One-sample t-test (one-sided) was performed with the average values and standard deviations obtained from these 10 attempts to test actual maximum accuracies and AUCs. Differences were considered significant for p values at two levels $p < 0.05$ and $p < 0.01$ and indicated by boldface and boldface with asterisks, respectively.

ADDITIONAL RESOURCES

None.