Chemical structure-based read-across represents a promising method for chemical toxicity evaluation without the need for animal testing; however, a chemical structure is not necessarily related to toxicity. Therefore, in vitro studies were often used for read-across reliability refinement; however, their external validity has been hindered by the gap between in vitro and in vivo conditions. Thus, we developed a virtual DNA microarray, regression analysis–based inductive DNA microarray (RAID), which quantitatively predicts in vivo gene expression profiles based on the chemical structure and/or in vitro transcriptome data. For each gene, elastic-net models were constructed using chemical descriptors and in vitro transcriptome data to predict in vivo data from in vitro data (in vitro to in vivo extrapolation; IVIVE). In feature selection, useful genes for assessing the quantitative structure–activity relationship (QSAR) and IVIVE were identified. Predicted transcriptome data derived from the RAID system reflected the in vivo gene expression profiles of characteristic hepatotoxic substances. Moreover, gene ontology and pathway analysis indicated that nuclear receptor-mediated xenobiotic response and metabolic activation are related to these gene expressions. The identified IVIVE-related genes were associated with fatty acid, xenobiotic, and drug metabolisms, indicating that in vitro studies were effective in evaluating these key events. Furthermore, validation studies revealed that chemical substances associated with these key events could be detected as hepatotoxic biosimilar substances. These results indicated that the RAID system could represent an alternative screening test for a repeated-dose toxicity test and toxicogenomics analyses. Our technology provides a critical solution for IVIVE-based read-across by considering the mode of action and chemical structures.

Keywords: oligonucleotide array, gene expression analysis, hepatotoxicity, alternative method, new approach methodology

INTRODUCTION

Non-animal testing to assess the efficacy and safety of chemical substances is one of the key concepts in balancing animal welfare and efficient development. Since the marketing ban in the EU in March 2013 [(EC) No. 1223/2009] (EU, 2009) of cosmetic products and ingredients tested on animal models, safety assessment methodologies independent of animal testing have attracted much attention. Simultaneously, the utilization of non-animal high-throughput technology for
optimizing drug discovery processes is becoming highly important in pharmaceuticals (Loiodice et al., 2017; Rognan, 2017; Amano et al., 2020).

Read-across, a process that estimates substance toxicity based on the concept that substances with similar chemical structures have similar biological activities, represents a promising approach and has already been conceptually accepted as a reliable safety risk assessment by some regulatory authorities (ECHA, 2017; European Commission, 2018). Likewise, quantitative structure–activity relationship (QSAR) has been widely used, and impurity characterization received regulatory acceptance (ICH M7). However, since subtle structural differences may elicit different biological responses, supporting the read-across robustness by using biological similarities has been considered important (Ball et al., 2016, 2020; Zhu et al., 2016). Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) mentions that the read-across performed by registrants often fails to comply with the legal requirements due to defects in the hypothesis and justification of toxicological prediction (ECHA, 2020).

There are two approaches to enhance the reliability of read-across: 1) employment of in vitro data relevant to specific toxicity. Methodologies to incorporate in vitro data within read-across (Ball et al., 2016, 2020; ECHA, 2017; Guo et al., 2019) and some case studies (OECD, 2016a, 2016b, 2018; Nakagawa et al., 2020, 2021) have been reported. However, these approaches can be applied only to specific toxicity endpoints and substances with a known toxicity and mode of action. Such conditions were previously termed as “local validity” (Patlewicz et al., 2014). 2) The use of biologically similar substances based on their profiles obtained from a large number of bioassays. The United States Environmental Protection Agency’s (US EPA’s) research project, ToxCast and Tox21, provided hundreds of high-throughput screening assays, and several groups employed such biological activity data for toxicological evaluation (Sipes et al., 2013; Berggren et al., 2015; Richard et al., 2021). Although this concept could be applied to substances with little information to elucidate their entire toxicological profiles and find their key mode of action, it is time-consuming and expensive to conduct numerous bioassays for a new candidate substance. By contrast, transcriptome data containing approximately 30,000 gene expression values can be used to estimate perturbed mechanisms through enrichment analysis. Wang et al. (2016) tried to predict drug-induced adverse effects by employing LINCS L1000 data (Subramanian et al., 2017), whereas Iwata et al. (2019) developed a computational method to predict missing values from the LINCS L1000 transcriptomic profiles of various human cell lines and provided new drug therapeutic indications. Genomic data have been considered to be usable in read-across by Health Canada and a research group from the US FDA (Health Canada, 2019; Liu et al., 2019). However, several researchers have shown that in vitro gene expression values are not always highly correlated with in vivo data (Sutherland et al., 2016; Grinberg et al., 2018; Liu et al., 2018). Thus, interpreting toxicological meaning from the in vitro–in vivo relationship and in vitro to in vivo extrapolation (IVIVE) in omics data represents a big challenge for chemical risk assessment. IVIVE was originally researched in toxicokinetics, such as in hepatic clearance and metabolites using hepatocytes (Soars et al., 2007; Umehara and Camenisch, 2012); most recent studies on non-animal testing have focused on predicting plasma concentrations, which is relevant for identification of a margin of exposure in risk assessment (Thomas et al., 2013a; Bell et al., 2018; Li et al., 2021). However, IVIVE should be considered for both toxicokinetics and toxicodynamics. Understanding of the in vitro to in vivo relationship of bioactivity data is also essential for non-animal testing. As an IVIVE study in omics data, Liu et al. (2020) developed a useful in silico strategy to narrow the data gap between in vitro and in vivo conditions. They modified in vitro data using non-generative matrix factorization methods to improve the correlation with in vivo data, which overcame the shortcomings of previous large-scale genomic data predictions regarding the in vitro–in vivo data gap (Liu et al., 2020). Although non-generative matrix factorization enables macroscopic estimation based on a pattern recognition classifying chemical and biological responses, it does not focus on estimation of each gene. As an alternative solution, microscopic estimation of each gene expression was performed based on tensor-train weighted optimization using machine learning (Iwata et al., 2019); however, such comprehensive estimations have not been integrated within an IVIVE study. Therefore, predicting in vivo transcriptomic profiles from in vitro data for IVIVE might not only enhance the robustness of read-across but could also be utilized in other non-animal testing strategies as weight of evidence, such as in Integrated Approaches to Testing and Assessment (IATA) and new approach methods (NAMs) for safety and drug repositioning research.

In this study, we developed a virtual DNA microarray that quantitatively predicts the in vivo gene expression profiles based on the chemical structure and/or in vitro transcriptome data. For each gene, elastic-net models, a regression analysis method that has been used in toxicity prediction with visualization of feature importance (e.g., Fujita et al., 2020), were constructed using chemical descriptors and in vitro transcriptome data. We named the set of prediction models “regression analysis–based inductive DNA microarray (RAID),” which inductively analyzes the mode of action and the key event in adverse effects with reference to the redundant arrays of inexpensive disks (also represented as RAID), a data storage virtualization technology that combines multiple physical disk drive components with the purpose of data redundancy. As RAID (storage technology) complements data based on information of multiple components, we hope that RAID (our microarray) will complement the relationships between multiple media (in vivo gene expression, in vitro gene expression, and chemical structure). Our RAID system achieved a quantitative in vitro to in vivo extrapolation (QIVIVE) by the integration of a structure-based approach (QSAR) with transcriptomic data. Whereas general “Q2IVIVE studies predict dose (or concentration) quantitatively in toxicological or toxicokinetic effects, our “Q2IVIVE predicts in vivo gene expression values quantitatively. Finally, the substance similarities were analyzed by principal component analysis (PCA), which proved useful in understanding the features of toxic substances based on their gene expression profile (Watanabe et al., 2012), using...
TABLE 1 | List of chemical substances used in the present study and their toxicological classes.

| Toxicological class* | Name                                                                 |
|----------------------|----------------------------------------------------------------------|
| Toxic                | Allyl alcohol (AA), 2-acetamidofluorene (AAF), α-naphthyl isothiocyanate (ANIT), Acetaminophen (APAP), Aspirin (ASA), Benz bromarone (BBR), Bromobenzene (BBZ), Butecin (BTC), Bendazac (BDZ), Benziodarone (BDI), Carboplatin (CBP), Coumarin (CM), Chlorimidine (CML), Chloramphenicol (CMP), Colchicine (COL), Cyclophosphamide monohydrate (CPA), Clomipramine hydrochloride (CPM), Chlorpropamide (CPP), Cyclosporine A (CPA), Diltiazem hydrochloride (DL), Disopyramide (DS), Disulfiram (DSP), Dantrolene sodium hemihydrate (DTL), Diazepam (DZP), Ethambutol dihydrochloride (EBL), 17-α-Ethynylestradiol (EE), DL-Ethinyl (ET), Fenofibrate (FFB), Flumidazole (FT), Gemfibrozil (GFZ), Hexachlorobenzene (HCB), Lomustine (LS), Mexiletine hydrochloride (MEX), Methapyrilene hydrochloride (MP), Metyltesterosterone (MTS), Methimazole (MTZ), Nimesulide (NIM), Phenacetin (PCT), Promethazine hydrochloride (PMZ), Propyliothiouracil (PTU), Sulphasalazine (SS), Simvastatin (SST), Thioacetamide (TAA), Terbutaline hydrochloride (TBP), Ticlopidine hydrochloride (TCP), Trimethadione (TMD), Vitamin A (VA), WY-14643 (WY) |
| Non-toxic            | Acarbose (ACA), Acetazolamide (ACZ), Aderap (ADP), Amlodipine (AM), Amiodarone hydrochloride (AM), Amitriptyline hydrochloride (AMT), Allopurinol (APL), 2-Bromomethylamino hydrobromide (BEA), Caffeine (CAF), Captopril (CAP), Carbamazepine (CBZ), Ceftriaxone (CFX), Chlorpheniramine maleate (CHL), Omeprazole (OMZ), Chloramphenicol acetate (CMA), Cefotaxime sodium (CTX), Ciprofloxacin hydrochloride (CPX), Chloropromazine hydrochloride (CPZ), Diclofenac sodium (DFN), Danazol (DNZ), Erythromycin ethylsuccinate (EME), Ethacrynic acid (EA), Ethanol (EHT), Etoposide (ETP), Famotidine (FAM), Fluphenazine hydrochloride (FP), Fluoxetine (FUR), Glibenclamide (GBC), Gentamicin sulfate (GMS), Haloperidol (HPL), Hydroxyurea (HUR), Isoniazid (INAH), Iproniazid phosphate (IPA), Ketoconazole (KZ), Methylene blue (MB), Mefenamic acid (MFA), Mefotrime hydrochloride (MMF), Minoxidil hydrochloride (MIX), Nifurtimox (NFT), Nitrofurantoin (NFT), Nitrozinc (NZC), Nifedipine (NIF), Omeprazole (OPZ), Papaverine hydrochloride (PAP), Penicillin G (PG), Penicillin V (PV), Phenacetin (PA), Phenformin hydrochloride (PHN), Phenothiazine (PHZ), Phenprocoumon (PPC), Phenprocoumon (PCP), Phenylbutazone (PB), Phenylalanine (PHE), Phenytoin (PHE), Pemoline (PML), Propionolactone (PRL), Propylthiouracil (PTU), Progesterone (PRG), Protamine (PRT), Propylthiouracil (PTU), Simvastatin (SST), Thiacetamide (TAA), Thiocarbamide (TCA), Thiourea (TU), Thioacetamide (TAA), Thioacetamide (TAA), Triamterene (TAA), Triazolam (TZA), Sodium valproate (VPA) |

*The toxicological classes of chemical substances were referred to in a previous report [Low et al., 2011]. The authors classified these substances into histopathological and serum chemistry classes. Substances with hepatotoxic histopathological findings and other histopathological findings with biochemical marker changes in serum chemistry were defined as toxic substances in this study.

RAID (the virtual microarray) data, in vivo data, in vitro data, and chemical structure data to validate the usefulness of read-across.

MATERIALS AND METHODS

Gene Expression and Chemical Structure Data

No animal experiment was performed in this study. The transcriptome data from DNA microarrays (Affymetrix Rat Genome 230 2.0 chips; Santa Clara, CA, United States) were extracted from the Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (TG-GATEs). TG-GATEs contains in vivo and in vitro transcriptome data for rat single- and repeated-dose toxicity tests of 170 compounds [Igarashi et al., 2015]. The transcriptome data obtained from the livers of rats treated with high doses for 28 days and primary rat hepatocytes treated with high doses for 24 h were downloaded and preprocessed using MAS5 (Gautier et al., 2004). In this study, chemical substances tested in vitro and in vivo those fulfilled a maximum sample number (n = 2 for in vitro and n = 3 for in vivo) and had no incalculable chemical descriptors (described below) were analyzed. Thus, 115 compounds were examined in this study (Table 1).

For the chemical structure data, the alvaDesc chemical descriptors (Mauri, 2020) were calculated using alvaDesc v1.0 software (Alvascience Srl, Lecco, Italy). AlvaDesc can calculate 3,885 2D-descriptors and 1,420 3D-descriptors. However, only 2D-descriptors were used, excluding those with a high pair correlation (>0.95), constant for all substances, and at least one missing value. Consequently, 854 descriptors were calculated. Each descriptor was normalized using the bestNormalize package (ver. 1.8.0) in R (ver. 4.1.1) (https://cran.r-project.org/). This package estimates the optimal normalization transformation from the Yeo-Johnson transformation, the Box Cox transformation, the log2 transformation, the square root transformation, and the arcsine transformation.

Construction of the Regression Analysis–Based Inductive DNA Microarray System (Virtual Microarray)

To extrapolate in vitro transcriptome data to in vivo conditions, we developed predictive models for each gene. The predictive models predicting in vivo transcriptome data from chemical descriptors and in vitro data were developed using the elastic net regression method. The value of each cell in the matrix was the fold change on a base 2 logarithmic scale. The set of those predictive models was named a virtual microarray “RAID” (as mentioned in the Introduction section) (Figure 1). To suppress overlearning, the hyperparameters (α and λ) of each model were optimized with a 5-fold cross-validation. We removed the genes that were associated with less than 10 chemical substances inducing differential expression (<1.5-fold change) since it would be difficult to run machine learning scripts on such rare genes. Consequently, RAID was composed of 1,601 prediction models for each gene.

To construct RAID that correctly predicts the bioactivities of chemical substances, the quality of training data sets was extremely
important, and differentially expressed genes should be determined strictly considering data noise. Hence, we addressed this issue by data processing (feature engineering) and model justification. First, after calculating the fold change values (sample treated groups/solvent control group), the gene differentiation values with low reliability were adjusted. Briefly, the fold change value increments were changed to half (e.g., 1.5 decreased to 1.25) in the sample with the number of flags A (low reliability) ≥2 out of 3 for in vivo and the number of flags A ≥1 out of 2 for in vitro, or in the sample with p-values ranging between 0.05 and 0.1. The fold change values were changed to one-fourth (e.g., 1.4 decreased to 1.1) in the sample with p-value over 0.1 and were treated as 1 (no differentiation) in the sample with flags A in both in vivo and in vitro. Second, the weight parameters were used in model building. The weight of samples with ≥1.5-fold change was set to 1.5 and ≥4-fold change was set to 2.

Interpretation of Biological Meaning of Regression Analysis–Based Inductive DNA Microarray Analysis

Considering the application of RAID to read-across, the gene expression data were visualized by PCA using prcomp function from stats package (ver. 4.1.1), and the probability ellipse frames of toxic and nontoxic substances were drawn using the ggfortify package (ver. 0.4.12) in R to compare in vivo, in vitro, and chemical descriptor data. The toxic class of chemical substances was determined based on previously reported histopathological and serum chemistry findings (Table 1) (Low et al., 2011). Since PCA did not use the toxicity label for classification, partial least squares discriminant analysis (PLS-DA) using the hepatotoxicity label was also conducted to confirm predictive performance (see Supplementary Material). As a reference data point, the biological meaning of genes that contributed to the PCA plot of in vivo data was analyzed using pathway analysis. The loading value of genes in the PCA was defined as length of loadings calculated using the Pythagorean theorem:

\[
\text{length} = \sqrt{(\text{loading of PC}1)^2 + (\text{loading of PC}2)^2},
\]

and genes with the top 30 loading values in the first and fourth quadrant were analyzed.

To analyze the biological consistency with in vivo data, commonality of principal component–related genes (top and bottom 30 rotations in each PC1 and PC2 of PCA) were visualized using the VennDiagram package (ver. 1.6.20) in R, and enrichment analyses of each categorized gene were conducted using Gene Ontology—biological process and Reactome pathway by Metascape (Zhou et al., 2019). Four categorized genes related to in vivo data (in vivo only, in vivo and RAID, in vivo and in vitro, and all three data) were analyzed.
to characterize which biological process could be covered by RAID and in vitro data. Furthermore, to characterize genes whose predictive models in RAID used in vitro data, enrichment analysis of the top 20 genes with the highest importance (contribution) for in vitro data in the model was conducted. In the analysis, the Affymetrix probe ID was converted to gene symbol using the biomaRt package (ver. 2.50.2) in R.

**Quantitative In Vitro to In Vivo Extrapolation Effects in Regression Analysis–Based Inductive DNA Microarray System**

For performance evaluation against the quantitative IVIVE, root-mean-square errors (RMSEs) of RAID predicted values to in vivo data were calculated and compared to those of in vitro data. To exclude the differences in gene expression value distribution of each data source, the fold change values were normalized before the RMSEs were calculated. The RMSEs were calculated both for all genes and genes for which in vitro data had importance in the model.

**Read-Across Application Using External Data**

To validate the usefulness of RAID for functional read-across–based analysis of both predicted gene expression profiles and chemical structures, substances that did not contain training data sets for model building (Table 1) were further explored using Ingenuity Pathway Analysis (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis). Specifically, substances that may promote the expression of genes (having a known relationship with the gene) that were identified by the PCA and pathway analysis of in vivo data (see the Interpretation of Biological Meaning of Regression Analysis–Based Inductive DNA Microarray Analysis section) were explored using IPA. Chemical descriptors of each substance were analyzed using the alvaDesc v1.0 software (Alvascience Srl, Lecco, Italy), and the gene expression profiles were fulfilled using median values of training data sets. Finally, RAID analyses using constructed predictive models for those substances and reanalyzed PCA data were used to evaluate similarities based on the predicted biological responses.

**RESULTS**

**Biological Analysis of Regression Analysis–Based Inductive DNA Microarray Compared to That of In Vivo and In Vitro Microarray Data**

RAID (predicted transcriptome) data were visualized using PCA (Figure 2). From a higher perspective, two directions mainly composed of toxic substances were identified, and many toxic substances were separated from non-toxic substances via RAID and in vivo data, whereas they could not be separated based on in vitro and chemical descriptor data. Moreover, two common toxic substances groups (e.g., first group (TAA, MP, and HCB) and second group (WY, FFB, BBr, and GFZ) placed in the first and fourth quadrants) were distanced from non-toxic substances along PC1 and PC2 in both RAID and in vivo data, nonetheless the PC1 and PC2 were replaced. The loading plot showed that Cyp1a1 (cytochrome P450, family 1, subfamily A, polypeptide 1), Gpx2 (glutathione peroxidase 2), and Gsta3 (glutathione S-transferase A3) gene expressions were commonly observed in RAID and in vivo data and enabled the discrimination of TAA, MP, and HCB. Furthermore, Acot1 (acyl-CoA thioesterase 1), Vmt1 (vimentin 1), and Cyp4a11 (cytochrome P450, family 4, subfamily A, polypeptide 11) contributed to discriminating WY, FFB, BBr, and GFZ.

Pathway analysis indicated that the first group–related genes would be associated with a xenobiotic response, such as Cyp1a induction via aryl hydrocarbon receptor (AHR) and carcinogenesis (Figure 3A), and the second group–related genes would be associated with peroxisome proliferative activity characterized by Cyp4a induction via peroxisome proliferator–activated receptor-alpha (PPARα) activation (Figure 3B). To clarify the biological functions that RAID covers, the commonalities between the related genes and principal components were explored (Figure 4 and Table 2). As expected from Figure 2, RAID shared more genes (36; Table 2) with the in vivo data than with the in vitro data (9). Enrichment analysis revealed that the biological processes related to metabolism and detoxification and pathways associated with peroxisomal protein transport were enriched in both in vivo and RAID data, indicating that RAID could cover these functions, and ultimately indicate key functions through pathway analysis (Figure 3). Conversely, although several metabolic processes were enriched within the in vitro data, those biological functions were covered by RAID as well (Figure 4). These results suggest that RAID data allow the detection of more in vivo key toxic events than in vitro transcriptome data.

For performance confirmation of discriminative analysis for hepatotoxicity, PLS-DA using RAID data allowed us to separate toxic chemicals with high accuracy (Supplementary Table S1). The accuracy using RAID data was better than that without RAID, when calibration and test data set were prepared.

**Importance of In Vitro Data in the Regression Analysis–Based Inductive DNA Microarray System**

Enrichment analysis of genes whose predictive model used highly relevant in vitro data (top 20 genes for which in vitro data had high importance in all predictive models; Table 3) indicated that in vitro data contributed to estimating the gene expression values associated with metabolic processes of fatty acids, xenobiotics, and drugs and peroxisome proliferative activity (pathway on peroxisome protein import and biological processes associated with the regulation of peroxisome size; Figure 5).

**Quantitative In Vitro to In Vivo Extrapolation Performance in the Regression Analysis–Based Inductive DNA Microarray System**

To evaluate RAID performance in terms of gene expression values, the RMSEs were calculated for all genes and the genes for which in vitro data had importance in predictive models. Considering RAID would be used in read-across, we compared...
FIGURE 2 | PCA score plots for chemical substances and the gene loading in the transcriptome data of (A) in vivo, (B) virtual microarray (RAID), and (C) in vitro data. PCA score plot with (D) chemical descriptor data. Uppercase letters in PCA score plots: abbreviations of chemical substances are described in Table 1. Blue: nontoxic substances. Red: hepatotoxic substances. Gene symbols are presented on the arrowhead (loading).
the RMSEs of RAID data with those of in vitro data, from conventional non-animal test approaches (Figure 6). The RMSEs were lower in RAID, indicating a better performance than what could be obtained using in vitro data.

**Validation of Prediction Models Using External Data**

In PCA with in vivo and RAID data, as well as the pathway analysis of PC-related genes (Figures 2, 3), the genes related to peroxisome proliferative activity and xenobiotic metabolism activity possibly leading to liver cancer, which were respectively characterized by Cyp1a induction via AHR and Cyp4 induction via PPARα, were identified as key features. Thus, potential Cyp1a and Cyp4a inducers were explored using the knowledge-based approach using the IPA software. Moreover, using the top 30 genes identified using PCA (described in the Interpretation of Biological Meaning of Regression Analysis-Based Inductive DNA Microarray Analysis section), upstream regulator analysis focusing on chemical substances...
was performed, and 20 chemicals were identified. Finally, a total of 21 chemicals (potential Cyp1a inducers: 10 chemicals; potential Cyp4a inducers: 11 chemicals) were selected as candidates for external validation and were subjected to RAID analyses (Table 4). Substances already present in the TG-GATE (training sets) or had uncalculated chemical descriptors data were excluded.

For the PCA, approximately half of the substances were plotted with positive PC scores, which was consistent with the direction expected from the training data set for both potential Cyp1a and Cyp4a inducers (Figure 7). Lastly, pentachlorobiphenyl, polychlorinated biphenyls, and pentachlorodibenzofuran were isolated as Cyp1a inducers, whereas nafenopin, ciprofibrate, and di(2-ethylhexyl)phthalate were isolated as Cyp4a inducers.

**DISCUSSION**

The transcriptome data signatures derived from the RAID (the virtual microarray) system were in good agreement with those of in vivo data, and the technology provided an understanding of the features of hepatotoxic substances based on the toxicological mechanism interpretation. The mechanism of action of the two characteristic toxic substances separated using PCA analysis was shown to be achieved through Cyp1a induction via AHR and Cyp4a induction via PPARa (pathway and gene ontology analysis). The AHR-induced drugs raise safety concerns during developmental periods (Qin et al., 2019), and PPARa-induced drug toxicity requires species differentiation considerations (Ito et al., 2006). Therefore, predicting the involvement of these nuclear receptors and induction of metabolic enzymes is critical for understanding the molecular initiating events and the key events associated with adverse outcome pathway. RAID enables the prediction of gene expression levels, thus exhibiting properties required for the next-generation risk assessment methods.

The first substance group (TAA, MP, and HCB), representing toxic substances commonly differentiated from non-toxic substances using PCA on in vivo and RAID data, has been reported to have carcinogenicity with metabolic activation (Uehara et al., 2008; Hajovsky et al., 2012; US HSS, 2015). Furthermore, these substances have been shown to activate xenobiotic-related receptors, such as AHR inducing Cyp1a (Ushel et al., 2002; Yamashita et al., 2014; Clara et al., 2015). Moreover, in vivo transcriptome data in this study showed that TAA, MP, and HCB induce Cyp1a activation. AHR is known for mediating the toxicity and tumor promoting properties despite the mechanism through which AHR activates carcinogenesis needing to be elucidated (Safe et al., 2013; Murray et al., 2014).

The second substance group (WY, FFB, BBr, and GFZ) includes fibrates which are recognized as PPARa agonists (Schoonjans et al., 1996), implying that induction of Cyp4a via PPARa and perturbation of lipid-related genes are involved as a series of key events. Although another fibrate included in training data—clofibrate (CFB)—was classified as a non-toxic substance.
### TABLE 2 | Principal components relating common genes in a virtual microarray (RAID) and in vivo data.

| Probe ID    | Symbol | Description                     |
|-------------|--------|---------------------------------|
| 1398250_at  | Acot1  | Acyl-CoA thioesterase 1         |
| 1370269_at  | Cyp1a1 | Cytochrome P450, family 1, subfamily a, polypeptide 1 |
| 1387022_at  | Aldh1a1| Aldehyde dehydrogenase 1, family member A1 |
| 1386534_at  | Cyp4a1 | Cytochrome P450, family 4, subfamily a, polypeptide 1 |
| 1388211_s_at| Acot1  | Acetyl-CoA thioesterase 1       |
| 1374070_at  | Gpx2   | Glutathione peroxidase 2        |
| 1367811_at  | Pdhgd  | Phosphoglycerate dehydrogenase  |
| 1389253_at  | Vmn1   | Vanin 1                         |
| 1388210_at  | Acot2  | Acyl-CoA thioesterase 2         |
| 1371089_at  | Gsta3  | Glutathione S-transferase alpha 3 |
| 1370491_a_at| Hdc    | Histidine decarboxylase         |
| 1379275_at  | Snx10  | Sorting nexin 10                |
| 1370902_at  | Akr1b6 | Aldo-keto reductase, family 1, member B8 |
| 1367733_at  | Car2   | Carbonic anhydrase              |
| 1386889_at  | Sred2  | Stearoyl-Coenzyme A desaturase 2 |
| 1386901_at  | LOC103690020 | Platelet glycoprotein 4-like |
| 1391187_at  | Pipl   | Periplakin                       |
| 1384225_at  | Dab1   | DAB adaptor protein 1           |
| 1384274_at  | AABR07037307 | similar to Spindlin-like protein 2 |
| 1395403_at  | Stac3  | SH3 and cysteine-rich domain 3   |
| 1375845_at  | Alg1   | Androgen induced 1              |
| 1368283_at  | Ehhadh | Enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase |
| 1387740_at  | Pex11a | Peroxisomal biogenesis factor 11 alpha |
| 1370067_at  | Met1   | Malic enzyme 1                  |
| 1370670_at  | Met1   | Malic enzyme 1                  |
| 1371886_at  | Crat   | Carotin O-acetyltransferase      |
| 1379361_at  | Pex11a | Peroxisomal biogenesis factor 11 alpha |
| 1386885_at  | Ech1   | Enoyl-CoA hydratase 1           |
| 1367659_s_at| Eci1   | Enoyl-CoA delta isomerase 1     |
| 1378169_at  | Acot3  | Acyl-CoA thioesterase 3         |
| 1374475_at  | Abhd1  | Abhydrolase domain containing 1 |
| 1387783_a_at| Aca1a  | Acetyl-Coenzyme A acetyltransferase 1A |
| 1386807_at  | Cyp4a9 | Cytochrome P450, family 4, subfamily a, polypeptide 8 |
| 1370698_at  | Ugt2b10| UDP-glucuronosyltransferase, family 2, member B10 |
| 1370387_at  | Cyp3a9 | Cytochrome P450, family 3, subfamily a, polypeptide 9 |

### TABLE 3 | List of top 20 genes with high importance in vitro data in the predictive models in RAID.

| Probe ID    | Symbol | Description                     | Importance of in vitro data |
|-------------|--------|---------------------------------|-----------------------------|
| 1398250_at  | Acot1  | Acyl-CoA thioesterase 1         | 0.550                       |
| 1368934_at  | Cyp1a1 | Cytochrome P450, family 4, subfamily a, polypeptide 1 | 0.412                       |
| 1367659_s_at| Eci1   | Enoyl-CoA delta isomerase 1     | 0.380                       |
| 1368283_at  | Ehhadh | Enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase | 0.348                       |
| 1387740_at  | Pex11a | Peroxisomal biogenesis factor 11 alpha | 0.314                       |
| 1370067_at  | Met1   | Malic enzyme 1                  | 0.284                       |
| 1386885_at  | Ech1   | Enoyl-CoA hydratase 1           | 0.252                       |
| 138253_at   | Vmn1   | Vanin 1                         | 0.244                       |
| 1377783_a_at| Aca1a  | Acetyl-Coenzyme A acetyltransferase 1A | 0.238                       |
| 1371076_at  | Cyp2b1 | Cytochrome P450, family 2, subfamily a, polypeptide 1 | 0.220                       |
| 1375845_at  | Alg1   | Androgen induced 1              | 0.166                       |
| 1388211_s_at| Acot1  | Acyl-CoA thioesterase 1         | 0.127                       |
| 1379361_at  | Pex11a | Peroxisomal biogenesis factor 11 alpha | 0.125                       |
| 1386901_at  | LOC103690020 | Platelet glycoprotein 4-like | 0.115                       |
| 1386807_at  | Cyp4a3 | Cytochrome P450, family 4, subfamily a, polypeptide 3 | 0.114                       |
| 1383883_at  | Acat2  | Acetyl-CoA acyltransferase 2    | 0.096                       |
| 1384244_at  | Hsd2   | Hydroxysteroid dehydrogenase like 2 | 0.074                       |
| 1370698_at  | Ugt2b10| UDP-glucuronosyltransferase, family 2, member B10 | 0.073                       |
| 1397468_at  | Hsd2   | Hydroxysteroid dehydrogenase like 2 | 0.071                       |
| 1367777_at  | Decr1  | 2,4-dienoyl-CoA reductase 1     | 0.070                       |
oxidation process of various lipids in peroxisomes (Hunt et al., 2012). Furthermore, hepatocytes and is involved in lipid and xenobiotic metabolism (Bartucci et al., 2019), whereas RAID could detect gene expressions related to major drug metabolism and hepatotoxicity. In addition to the genes described above were known to be involved in drug metabolism and xenobiotic metabolism (Luu et al., 2011), whereas peroxisomal biogenesis factor 11 alpha (Pex11a) is involved in peroxisome maintenance and proliferation associated with dyslipidemia (Chen et al., 2018). All of these genes are known as PPARα target genes (Rakhshandehroo et al., 2010; Lake et al., 2016). Thus, these features indicate that RAID can predict possible toxicity by taking into account a broader range of mechanisms than the range of in vitro data. Indeed, the in vivo changes detected using the in vitro data were limited (Figure 4), and the PCA showed that most of the differentially expressed genes were associated with irrelevant nonphysiological conditions. Thus, the IVIVE effect combining the QSAR technique and in vitro data would allow for more precise predictions through de-noising these types of in vitro specific biological responses.

In vitro data contribute to accurate gene expression predictions that could not be achieved with QSAR alone (Figure 2D). In vitro data contributed to the prediction of the mechanism shown in Figure 5. The biological mechanisms related to metabolic processes were consistent with the key mechanisms of characteristic hepatotoxic substances described above, which indicates that in vitro data contributes to the precise predictions obtained using RAID. In addition, whether in vitro responses were observed in the suggested mode of action predicted by the RAID system or not is an important point in terms of weight of evidence. This study provides valuable evidence supporting that transcriptome data should be considered in light of previous reports indicating that in vitro data does not necessarily reflect in vivo conditions (Tamura et al., 2006; Sutherland et al., 2016). Simultaneously, in vitro studies focusing on a specific mechanism should consider the external validity of their findings and whether the findings reflect in vivo situations.

Evaluating the read-across performance using external substances, such as 3,4,5,3′,4′-pentachlorobiphenyl, 2,2′,4,4′-tetrachlorobiphenyl (a type of polychlorinated biphenyl) and pentachlorodibenzofuran (dioxin-like compounds) (Figure 7A), which are known as IARC group 1 carcinogens and Cyp1a1 inducers (EPA, U.S., 1996; Walker et al., 2005; National; Toxicology Program, 2006); these were separated as toxic substances. Additionally, benzo(a)pyrene, 3-methylcholanthrene, and 9,10-dimethyl-1,2-benzanthracene plotted apart from the origin of coordinates (PC1 = 0 and PC2 = 0), and are polycyclic aromatic hydrocarbons inducing Cyp1a1 (Moorthy et al., 2007; Pushparajah et al., 2008). Non-carcinogenic chemical substances, such as food components or...
preservatives, were positioned near the origin, second quadrant or third quadrant, indicating low risk. Furthermore, substances interacting with Cyp4a (Figure 7B), such as ciprofibrate, nafenopin, clofenapate, clofibric acid, and di(2-ethylhexyl) phthalate, which plotted in the area of the 2nd substance group (PC1 > 0), are also known as PPARα agonists (Bocos et al., 1995; Roberts et al., 2002; Yadetie et al., 2003; Currie et al., 2005; Pyper et al., 2010). Chemicals that were not characterized by the PC1 component (PC1 < 0) are not hyperlipidemia drugs. These results suggest that the RAID system effectively classifies substances based on their mode of action as well as the strength of their toxicity, and ultimately contributes to precise read-across. Thus, the RAID system provides a new method for read-across in line with IATA that should be called “a virtual functional read-across”. Here, we showed that substances without high structural similarities might have similar toxicological properties, and our new approach interpreted the shared mechanism of action. This means that RAID considers the qualitative and quantitative similarities of biological responses, which was one of the major issues of QSAR-based read-across. The structural similarities of TAA, MP, and HCB observed using correlation coefficients of the chemical descriptor used for the predictive model, and the maximum common substructure (MCS) similarities with the Tanimoto coefficient, were less than 0.5; however, the homology of RAID and in vivo data was as high as a 0.8 Pearson’s correlation coefficient. Furthermore, achieving such an accurate read-across without using in vitro data will provide a new perspective on the structural information-based predictions.

PCA analysis was used to understand the features of substances to predict the modes of action and identify biologically similar substances for read-across in this study. The examples of applications of RAID for read-across described above were compared to other methods (Table 5). The RAID system could enhance read-across reliability by estimating toxicity including modes of action, while this was difficult by other methods (e.g., QSAR or read-across using chemical structure data) (Supplementary Figure S1). On the other hand, focusing on certain specific toxicities, discriminant analysis, classifier model, or biomarker analysis might improve the separation of toxic substances. Indeed, as shown in Supplementary Table S1, when the RAID system was applied to discriminating

**TABLE 4** | List of chemical substances used for external validation of the RAID system.

| Name (Potential Cyp1a inducers) | CAS no. | Name in PCA plot |
|----------------------------------|---------|-----------------|
| 2,3,4,7,8-Pentachlorodibenzofuran | 57117-31-4 | Pentachlorodibenzofuran |
| 3,4,5,3′,4′-Pentachlorobiphenyl | 57465-28-8 | Pentachlorobiphenyl |
| 3-Methylcholanthrene | 56-49-5 | Methylcholanthrene |
| 9,10-Dimethyl-1,2-benzanthracene | 57-97-6 | Dimethylbenzanthracene |
| Benzo(a)pyrene | 50-32-8 | Benzo(a)pyrene |
| Dexamethasone | 8054-59-9 | Dexamethasone |
| Genistein | 446-72-0 | Genistein |
| 2′,4′,4′-Tetrachlorobiphenyl | 1336-36-3 | Tetrachlorobiphenyl |
| Quercetin | 117-39-5 | Quercetin |
| Resveratrol | 501-36-0 | Resveratrol |
| Thiazolidodazole | 148-79-8 | Thiazolidodazole |

| Name (Potential Cyp4a inducers) | CAS no. | Name in PCA plot |
|----------------------------------|---------|-----------------|
| Streptozotocin | 18883-66-4 | Streptozotocin |
| 2-Ethylhexanol | 104-76-7 | Ethylhexanol |
| Di(2-ethylhexyl) phthalate | 117-81-7 | Di(2-ethylhexyl) phthalate |
| Clofenapate | 21340-68-1 | Clofenapate |
| Clofibric acid | 882-09-7 | Clofibric acid |
| Ciprofibrate | 52214-84-3 | Ciprofibrate |
| Nafenopin | 3711-19-5 | Nafenopin |
| TO-901317 | 293754-55-9 | TO-901317 |
| Acetaminophen | 719293-04-6 | Acetaminophen |
| Diltiazem | 332862-22-6 | Diltiazem |

**FIGURE 7** | Read-across using PCA plot of external data predicted by a virtual microarray (RAID). (A) Cyp1a and (B) Cyp4a inducing chemical substances were analyzed for validation.
TABLE 5 | The relationships between the pros and cons of RAID and other methods for read-across.

| Examples of chemical substances in the present study | QSAR | Read-across using PCA of chemical structure data | Read-across using PCA of RAID data |
|------------------------------------------------------|------|-----------------------------------------------|----------------------------------|
| Internal data                                         |      | Pros. Toxicity may be identified.             | Pros. The toxicity and modes of action of in vivo can be estimated from the PCA plot. |
| A: TAA                                               |      |                                               | Animal testing data of similar substances can be utilized for the assessment. A: HCB and MP were similar substances, and it was estimated that TAA could make Cyp1a induction via AHR, and substances plotted in the first quadrant would have similar possibilities. B: WY, FFB, BBr, and GFZ were similar substances, and it was estimated that FFB could make Cyp4a induction via PPARa activation, and substances plotted in the first quadrant would have similar possibilities. C: Its toxicological response could be similar to “TAA, MP, and HCB,” indicating that it could induce Cyp1a. Its carcinogenic potential should be confirmed using further additional testing. D: Its toxicological response could be similar to “WY, FFB, BBr, and GFZ,” indicating that it could induce Cyp4a and also affect the expression of PPARa-related genes. |
| B: FFB                                               |      |                                               |                                  |
| External data                                        |      | Cons. Mechanisms cannot be fully estimated because of the lack of biological activity data. Toxicity in organs and individuals cannot be characterized. Biologically similar substances cannot be identified. | Cons. The reliability of the estimated modes of action would depend on the accuracy of the RAID system. |
| C: 3,4,5,3′,4′-pentachlorobiphenyl                  |      | Cons. Estimation of the toxicity and modes of action from the PCA plot is complicated because toxic substances cannot be separated well from non-toxic substances. A, C: Estimation of the toxicity and modes of action was difficult since similar substances were both toxic and non-toxic. B, D: Specific similar substances were not identified since they were surrounded by many substances. |
| D: Nafenopin                                         |      |                                               |                                  |

hepatotoxicity, PLS-DA using RAID data showed a good predictive performance, indicating usefulness for the specified toxicity prediction. Thus, the use of RAID data instead of experimental transcriptome data would achieve previously reported biomarker-based classification without using animals. For example, Liu et al. (2017) indicated that certain genes were associated with hepatocellular hypertrophy and hepatocarcinogenesis, as well as markers such as Cyp1a1, Aco1, Stac3 (SH3 and cysteine rich domain 3), and Hdc (histidine decarboxylase), which were correctly evaluated in the present study to characterize hepatotoxic substances. Similarly, the constructed RAID system could be applied to previous studies to predict carcinogenicity or estimate transcriptional benchmark dose by toxicogenomics analysis of short term in vivo studies (Ellinger-ziegelbauer et al., 2008; Thomas et al., 2013b; Matsumoto et al., 2014; Kawamoto et al., 2017).

One important issue that should be considered in toxicological evaluation using the RAID system is consideration of species differences. The RAID system provides mechanistic insights on repeated-dose toxicity in animal models; however, since some species differences have been observed, the suggested mode of action and the corresponding molecules need to be confirmed by toxicologists. The interspecies extrapolations, such as rat-to-human extrapolations, could be achieved by further experiments to construct new RAID systems with these different species’ transcriptome data. In addition, RAID data of substances that were separated as toxic substances in PCA (e.g., TAA, MP, HCB, WY, FFB, BBr, and GFZ) showed high similarity to in vivo data (Supplementary Table S2). Since regression analysis requires certain levels of standard deviation of training data, the RAID accuracy for substances may be related to the number of substances with similar modes of action. Thus, database expansion for several substance groups with minor modes of action would contribute to further improving the accuracy and applicability domain. In addition, evaluation of RAID usefulness for various toxicities is required.

The present approach integrates QSAR and IVIVE and will contribute to other areas of research, such as drug repositioning, which has recently attracted attention toward pharmaceuticals that are available on the market and might be repurposed for new diseases (Jourdan et al., 2020). However, the previously proposed methodologies (Iwata et al., 2018; Lippmann et al., 2018; Zhu et al., 2020; He et al., 2021) have room for improving the IVIVE aspect of in vivo predictions. Thus, our system provides an
alternative to screening candidate drugs and exploring new biologically similar drugs at a low cost.

In conclusion, we developed a virtual DNA microarray system that quantitatively predicts in vivo gene expression profiles based on the chemical structure and/or in vitro transcriptome data. Estimated transcriptomes are considered scientifically relevant from PCA data interpretation as well as pathway and GO analysis. Based on its external validation, our system works as an alternative test for repeated dose toxicity tests with toxicogenomics analysis enabling IVIVE and mechanism estimation. Although our technology might have limited applicability domain due to the small data size of chemical substances and their characteristics (using hepatotoxic substances), the concept of the virtual microarray analysis contributes to the 3Rs (reduction, refinement, and replacement) and might benefit much future animal testing.

DATA AVAILABILITY STATEMENT

Publicly available data sets were analyzed in this study. This data can be found at https://dbarchive.biostacksdbc.jp/en/open-ttgates/download.html.

REFERENCES

Amano, Y., Honda, H., Sawada, R., Nukada, Y., Yamane, M., Ikeda, N., et al. (2020). In Silico systems for Predicting Chemical-Induced Side Effects Using Known and Potential Chemical Protein Interactions, Enabling Mechanism Estimation. J. Toxicol. Sci. 45, 137–149. doi:10.2131/jts.45.137

Ball, N., Cronin, M. T., Shen, J., Blackburn, K., Booth, E. D., Bouhid, M., et al. (2016). Toward Good Read-Across Practice (GRAP) Guidance. ALTEx 33, 149–166. doi:10.10473/alex.1601251

Ball, N., Madden, J., Paini, A., Mathea, M., Palmer, A. D., Sperber, S., et al. (2020). Key Read across Framework Components and Biology Based Improvements. Mutat. Res. 851, 503172. doi:10.1016/j.mirxen.2020.503172

Bartucci, R., Salvati, A., Olinga, P., and Boersma, Y. L. (2019). Vanin 1: Its Physiological Function and Role in Diseases. Int. J. Mol. Sci. 20, 3891. doi:10.3390/ijms20163891

Bell, S. M., Chang, X., Wambaugh, J. F., Allen, D. G., Bartels, M., Brouwer, K. L. R., Bergegren, E., Amc off, P., Benigni, R., Blackburn, K., Booth, E. D., Bouhid, M., et al. (2019). Toward Good Read-Across Practice (GRAP) Guidance. ALTEx 33, 149–166. doi:10.10473/alex.1601251

Bell, S. M., Chang, X., Wambaugh, J. F., Allen, D. G., Bartels, M., Brouwer, K. L. R., Bergegren, E., Amc off, P., Benigni, R., Blackburn, K., Booth, E. D., Bouhid, M., et al. (2019). Toward Good Read-Across Practice (GRAP) Guidance. ALTEx 33, 149–166. doi:10.10473/alex.1601251

Bell, S. M., Chang, X., Wambaugh, J. F., Allen, D. G., Bartels, M., Brouwer, K. L. R., Bergegren, E., Amc off, P., Benigni, R., Blackburn, K., Booth, E. D., Bouhid, M., et al. (2019). Toward Good Read-Across Practice (GRAP) Guidance. ALTEx 33, 149–166. doi:10.10473/alex.1601251

Ball, N., Cronin, M. T., Shen, J., Blackburn, K., Booth, E. D., Bouhid, M., et al. (2016). Toward Good Read-Across Practice (GRAP) Guidance. ALTEx 33, 149–166. doi:10.10473/alex.1601251

Ball, N., Madden, J., Paini, A., Mathea, M., Palmer, A. D., Sperber, S., et al. (2020). Key Read across Framework Components and Biology Based Improvements. Mutat. Res. 851, 503172. doi:10.1016/j.mirxen.2020.503172

Ball, N., Madden, J., Paini, A., Mathea, M., Palmer, A. D., Sperber, S., et al. (2020). Key Read across Framework Components and Biology Based Improvements. Mutat. Res. 851, 503172. doi:10.1016/j.mirxen.2020.503172

Bartucci, R., Salvati, A., Olinga, P., and Boersma, Y. L. (2019). Vanin 1: Its Physiological Function and Role in Diseases. Int. J. Mol. Sci. 20, 3891. doi:10.3390/ijms20163891

Bell, S. M., Chang, X., Wambaugh, J. F., Allen, D. G., Bartels, M., Brouwer, K. L. R., et al. (2018). In Vitro to In Vivo Extrapolation for High Throughput Prioritization and Decision Making. Toxicol Vitro 47, 213–227. doi:10.1016/j.tiv.2017.11.016

Berggren, E., Amcoff, P., Benigni, R., Blackburn, K., Carney, E., Cronin, M., et al. (2015). Chemical Safety Assessment Using Read-Across: Assessing the Use of Novel Testing Methods to Strengthen the Evidence Base for Decision Making. Environ. Health Perspect. 123, 1223–1246. doi:10.1289/ ehp.1409342

Bocois, C., Götlich, M., Gearing, K., Banner, C., Enmark, E., Teboul, M., et al. (1995). Fatty Acid Activation of Peroxisome Proliferator-Activated Receptor (PPAR). J Steroid Biochem. Mol. Biol. 53, 467–473. doi:10.1016/0960-0760(95) 00939-f

Chen, C., Wang, H., Chen, B., Chen, D., Lu, C., Li, H., et al. (2018). Pex11a Deficiency Causes Dyslipidaemia and Obesity in Mice. J. Cell. Mol. Med. 23, 2020–2031. doi:10.1111/jcmm.14108

Currie, R. A., Bombl, V., Oliver, J. D., Moore, D. J., Lim, F. L., Gwilliam, V., et al. (2005). Gene Ontology Mapping as an Unbiased Method for Identifying Molecular Pathways and Processes Affected by Toxicant Exposure: Application to Acute Effects Caused by the Rodent Non-genotoxic Carcinogen Diethylhexylphthalate. Toxicol. Sci. 86, 453–469. doi:10.1093/toxsci/kf6207

AUTHOR CONTRIBUTIONS

YA and HH contributed to the conception and design of the study, YA and HH constructed in silico models, performed enrichment analyses, interpreted the biological meanings of the models, and contributed to statistical analyses. HH collected the data sets from TG-GATE. HH and MY supervised this project. YA and HH drafted the manuscript. All authors contributed to manuscript writing, confirmed the final version of the manuscript, and agreed to the contents.

ACKNOWLEDGMENTS

We thank Dr. Osamu Morita, Dr. Kaede Miyata, and Mr. Yasuaki Inoue for their helpful suggestions and valuable discussions to the present study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.879907/full#supplementary-material
Richard, A. M., Huang, R., Waidyanatha, S., Shinn, P., Collins, B. J., Thillainadarajah, L., et al. (2021). The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. Chem. Res. Toxicol. 34, 189–216. doi:10.1021/acs.chemrestox.0c00264

Roberts, R. A., Chevalier, S., Hasmall, S. C., James, N. H., Cosulich, S. C., and Macdonald, N. (2002). PPAR Alpha and the Regulation of Cell Division and Apoptosis. Toxiconology 181-182, 167–170. doi:10.1007/s00383-001-0267-5

Rognan, D. (2017). The Impact of In Silico Screening in the Discovery of Novel and Safer Drug Candidates. Pharmacol. Ther. 175, 47–66. doi:10.1016/j.pharmthera.2017.02.034

Safe, S., Lee, S. O., and Jin, U. H. (2013). Role of the Aryl Hydrocarbon Receptor in Carcinogenesis and Potential as a Drug Target. Toxicol. Sci. 135, 1–16. doi:10.1093/toxsci/kft028

Schoonjans, K., Staels, B., and Auwerx, J. (1996). Role of the Peroxisome Proliferator-Activated Receptor (PPAR) in Mediating the Effects of Fibrates and Fatty Acids on Gene Expression. J. Lipid Res. 37, 907–925. doi:10.1016/s0022-2275(20)42003-6

Sipes, N. S., Martin, M. T., Kothiya, P., Reif, D. M., Judson, R. S., Richard, A. M., et al. (2013). Profiling 976 ToxCast Chemicals across 331 Enzymatic and Receptor Signaling Assays. Chem. Res. Toxicol. 26, 878–895. doi:10.1021/tx400021f

Soars, M. G., McGinnity, D. F., Grime, K., and Riley, R. J. (2007). The Pivotal Role of Hepatocytes in Drug Discovery. Chem. Biol. Interact. 168, 2–15. doi:10.1016/j.cbi.2006.11.002

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natali, T. E., Lu, X., et al. (2017). A Next Generation Connectivity Map: 11000 Platform and the First 1,000,000 Profiles. Cell. 171, 1437–e17. doi:10.1016/j.cell.2017.10.049

Sutherland, J. J., Jolly, R. A., Goldstein, K. M., and Stevens, J. L. (2016). Assessing Concordance of Drug-Induced Transcriptional Response in Rodent Liver and Cultured Hepatocytes. PLoS Comput. Biol. 12, e1004847–31. doi:10.1371/journal.pcbi.1004847

Tamura, K., Ono, A., Miyagishima, T., Nagao, T., and Urushidani, T. (2006). Profiling of Gene Expression in Rat Liver and Rat Primary Cultured Hepatocytes Treated with Peroxisome Proliferators. J. Toxicol. Sci. 31, 471-490. doi:10.2131/jts.31.471

Thomas, R. S., Philbert, M. A., Auerbach, S. S., Wetmore, B. A., Devito, M. J., Cote, L., et al. (2013a). Incorporating New Technologies into Toxicity Testing and Risk Assessment: Moving from 21st Century Vision to a Data-Driven Framework. Toxicol. Sci. 136, 4–18. doi:10.1093/toxsci/kft178

Thomas, R. S., Wesselskamper, S. C., Wang, N. C., Zhao, Q. J., Petersen, D. D., Lambert, J. C., et al. (2013b). Temporal Concordance between Apical and Transcriptional Points of Departure for Chemical Risk Assessment. Toxicol. Sci. 134, 180–194. doi:10.1093/toxsci/kft094

Toxicology Program, National. (2006). NTP Toxicology and Carcinogenesis Studies of 3,3',4,4',5-pentachlorobiphenyl (PCB 126) (CAS No. 57465-28-8) in Female Harlan Sprague-Dawley Rats (Gavage Studies). Natl. Toxicol. Program Tech. Rep. Ser. 520, 4–246.

Uchera, T., Kiyosawa, N., Hirose, M., Omura, K., Shimizu, T., Ono, A., et al. (2008). Gene Expression Profiling of Methylylrene-Induced Hepatotoxicity in Rat. J. Toxicol. Sci. 33, 37–50. doi:10.2131/jts.33.37

Umehara, K., and Camenisch, G. (2012). Novel In Vitro-In Vivo Extrapolation (IVIVE) Method to Predict Hepatic Organ Clearance in Rat. Pharm. Res. 29, 603–617. doi:10.1007/s11095-011-0607-2

Walker, N. J., Crockett, P. W., Nyska, A., Brix, A. E., Jokinen, M. P., Sells, D. M., et al. (2005). Dose-Additive Carcinogenicity of a Defined Mixture of 'Dioxin-like Compounds'. Environ. Health Perspect. 113, 43–48. doi:10.1289/ehp.7351

Wang, Z., Clark, N. R., and Ma’ayan, A. (2016). Drug-induced Adverse Events Prediction with the LINCS L1000 Data. Bioinformatics 32, 2338–2345. doi:10.1093/bioinformatics/btw168

Watanabe, T., Suzuki, T., Natsume, M., Nakajima, M., Narumi, K., Hamada, S., et al. (2012). Discrimination of Genotoxic and Non-genotoxic Hepatocarcinogens by Statistical Analysis Based on Gene Expression Profiling in the Mouse Liver as Determined by Quantitative Real-Time PCR. Mutat. Res. 747, 164–175. doi:10.1016/j.mrgentox.2012.04.011

Yadetie, F., Laegreid, A., Bakke, I., Kusnierczyk, W., Komorowski, J., Waldum, H. L., et al. (2003). Liver Gene Expression in Rats in Response to the Peroxisome Proliferator-Activated Receptor-Alphagonist Ciprofibrate. Physiol. Genomics 15, 9–19. doi:10.1115/physiogenomics.00064.2003

Yamashita, Y., Ueyama, T., Nishi, T., Yamamoto, Y., Kawakoshi, A., Sunami, S., et al. (2014). Nrf2-inducing Anti-oxidation Stress Response in the Rat Liver: New Beneficial Effect of Lansoprazole. PLoS One 9, e97419. doi:10.1371/journal.pone.0097419

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). MetaScape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. Nat. Commun. 10, 1523. doi:10.1038/s41467-019-09234-6

Zhu, H., Bouhidi, M., Donley, E., Egnash, L., Kleinstreuer, N., Kroese, E. D., et al. (2016). Supporting Read-Across Using Biological Data. ALTEX 33, 167–182. doi:10.14573/altex.1601252

Zhu, L., Roberts, R., Huang, R., Zhao, J., Xia, M., Delavan, B., et al. (2020). Drug Repositioning for Noonan and LEOPARD Syndromes by Integrating Transcriptomics with a Structure-Based Approach. Front. Pharmacol. 11, 927. doi:10.3389/fphar.2020.00927

Conflict of Interest: YA, MY, and HH were employed by the company Kao Corporation.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, or the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Amano, Yamane and Honda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.