Narrowing the Gap Between *In Vitro* and *In Vivo* Genetic Profiles by Deconvoluting Toxicogenomic Data *In Silico*

Yuan Liu[1], Runyu Jing[2], Zhining Wen[1]* and Menglong Li[1]*

[1] College of Chemistry, Sichuan University, Chengdu, China, [2] College of Cybersecurity, Sichuan University, Chengdu, China

Toxicogenomics (TGx) is a powerful method to evaluate toxicity and is widely used in both *in vivo* and *in vitro* assays. For *in vivo* TGx, reduction, refinement, and replacement represent the unremitting pursuit of live-animal tests, but *in vitro* assays, as alternatives, usually demonstrate poor correlation with real *in vivo* assays. In living subjects, in addition to drug effects, inner-environmental reactions also affect genetic variation, and these two factors are further jointly reflected in gene abundance. Thus, finding a strategy to factorize inner-environmental factor from *in vivo* assays based on gene expression levels and to further utilize *in vitro* data to better simulate *in vivo* data is needed. We proposed a strategy based on post-modified non-negative matrix factorization, which can estimate the gene expression profiles and contents of major factors in samples. The applicability of the strategy was first verified, and the strategy was then utilized to simulate *in vivo* data by correcting *in vitro* data. The similarities between real *in vivo* data and simulated data (single-dose 0.72, repeat-doses 0.75) were higher than those observed when directly comparing real *in vivo* data with *in vitro* data (single-dose 0.56, repeat-doses 0.70). Moreover, by keeping environment-related factor, a simulation can always be generated by using *in vitro* data to provide potential substitutions for *in vivo* TGx and to reduce the launch of live-animal tests.

Keywords: toxicogenomics, nonnegative matrix factorization, *in vitro* to *in vivo* extrapolation, *in vivo* and *in vitro* strategies, gene expression, deconvolution, liver, bioinformatics

INTRODUCTION

In vivo and *in vitro* experimental systems are two essential ways to discover functional performance during drug-discovery, vital processes occurring in living organisms and toxicological research (Nuwaysir et al., 1999). For animal trials, *in vivo* experiments still can be launched properly for toxicological research (Hussain et al., 1987; Liu et al., 2017). The only problem is that *in vivo* costs are higher than those for *in vitro* assays, so it is more difficult to utilize *in vivo* systems in large-scale projects. However, for human trials, *in vivo* studies can only be applied in the field if they cause no damage to the human body, such as the studies in the brain science (Plate et al., 1992; Glasser and Van Essen, 2011; Zhang et al., 2012), neuroscience (Fagan et al., 2006; Swain et al., 2007), and cell behavior observation fields (Gronthos et al., 2000; Pillay et al., 2010; Eskildsen et al., 2011). More

specifically, in toxicology studies, *in vivo* assays are currently difficult to apply in human trials because of the high risk to human volunteers or patients.

With the recent rapid development of genomics, toxicogenomics (TGx), which is a combination of toxicology and genomics technologies, has become a powerful method to study the underlying molecular mechanisms of toxicity (Aardema and MacGregor, 2003). TGx gives a novel perspective to investigate and predict toxicology (Nuwaysir et al., 1999; Waters and Fostel, 2004), risk assessment (Pennie et al., 2004; Boverhof and Zacharewski, 2005), and mechanistic studies (Suter et al., 2004; Suter et al., 2011; Chen et al., 2012a). Moreover, with the development of emerging technologies, novel genomic TGx features, such as microRNAs (Wang et al., 2009; Yang et al., 2011) and lncRNAs (Aigner et al., 2016; Dempsey and Cui, 2017), may provide a new way to achieve more resolution and better understand the mechanisms of toxicology. However, the fact that *in vivo* assays have high costs and are difficult to apply in humans still persists with TGx and has hindered the development of toxicology.

The ideal situation is that an *in vivo* system can be supported or even reflected by a low-cost approach or experiment involving animal reduction (Iwatsubo et al., 1997; Rostami-Hodjegan and Tucker, 2007). To reach this goal, decades of studies have tried approaches using different data and models. Among all the approaches, obtaining models from *in vitro* cells or tissue cultures (IVIVE—*in vitro* to *in vivo* extrapolation) has become a main alternative (Knight et al., 2006). For utilizing cells and tissue cultures, the concept of "the 3Rs of alternatives" (reduction, refinement, and replacement) (Russell et al., 1959) was first described. Based on the 3Rs, worldwide organizations are trying to develop methods to achieve IVIVE by using TGx. These methods include the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) program launched in Europe (Abbott, 2005), "Advancing Regulatory Science" initiated by the Food and Drug Administration (FDA) (Hamburg, 2011) of the United States, and a guideline for the use of the single-cell gel (SCG)/Comet assay developed by the International Workshop on Genotoxicity Test Procedures (IWGTP) to standardize genetic toxicology procedures (Tice et al., 2000). Additionally, other relevant programs have been launched by organizations, such as ICH (International Conference on Harmonization of technical requirements for the registration of pharmaceuticals for human use), CPCSEA (Committee for the Purpose of Control and Supervision of Experiments on Animals), NIH (National Institutes of Health), and OECD (Organization for Economic Co-operation and Development) (Rollin, 2003; Abbott, 2005; Dix et al., 2007; Hamburg, 2011; Tice et al., 2013).

With worldwide contributions, the capability of *in vitro* data for providing assistance and references for *in vivo* data has approved substantially. For instance, IVIVE can be achieved by using physiologically-based pharmacokinetic (PBPK) modeling (Chen et al., 2012b; Sager et al., 2015) on TGx (Alam et al., 2013; Martin et al., 2014). Additionally, human health risk can be assessed based on TGx analysis (Johnson et al., 2015). A

demonstration showed that liver sections exhibit the strongest similarity to liver tissue in terms of mRNA expression (Boess et al., 2003). Reviews confirmed that TGx is able to improve comprehension of the mechanisms underlying the responses of *in vitro* and *in vivo* systems (Poma and Di Giorgio, 2008). Moreover, *in vitro* data predicted carcinogenesis in rats based on short-term TGx data (Watanabe et al., 2007; Ellinger-Ziegelbauer et al., 2008). To summarize, *in vitro* assays as alternatives for TGx play a vital role in the next-generation risk assessment paradigm and have tremendous potential to promote non-animal testing in TGx systems (Liu et al., 2018c).

However, the goal of using *in vitro* data to substitute for *in vivo* data is greatly impeded by the inconsistency between *in vivo* and *in vitro* data (Niki, 2010). Many researchers are challenged by the fact that *in vitro* data demonstrate poor correlations with *in vivo* data and have questioned the validity of IVIVE models (Tice et al., 2000; Chen et al., 2014; Sutherland et al., 2016). Previous research has stressed that many available methods result in inconsistent results regarding antioxidant capacity between *in vivo* and *in vitro* data (Niki, 2010). Laboratory technicians have even mentioned that the *in vivo* dynamics of antigen-specific regulatory T cells cannot be predicted from *in vitro* behavior (Klein et al., 2003). Some publications showed that the data obtained from *in vivo* livers without the use of chloroquine was inconsistent with *in vitro* gene expression results when using cultured HepG2 cells (Harada-Shiba et al., 2002). After assessing the capacity of *in vitro* screening studies to predict the *in vivo* pulmonary toxicity of several fine or nanoscale particles in rats, a poor correlation between *in vivo* and *in vitro* studies was observed (Sayes et al., 2007). The pair ranking (PRank) method (Liu et al., 2017) was proposed by the National Center for Toxicological Research of the U.S. FDA to assess IVIVE and to quantitatively measure similarity (Chen et al., 2012a; Chen et al., 2014). The PRank method has made it possible to properly compare gene expression data between *in vivo* and *in vitro* assays. However, the similarities between *in vitro* and *in vivo* data are unsatisfactory. The similarities between *in vivo* and *in vitro* data still can be further improved, especially for *in vivo* single-dose studies.

According to current investigations, in addition to the response signals from drug effects, many variables need to be considered inner-environmental factors that impact genetic variations in *in vivo* assays, such as cell types, culture conditions, time course of exposure, and measured end points (Sayes et al., 2007). Differences are found when demonstrating the immune response components (effectors) of *in vivo* and *in vitro* hepatocytes (Bumgardner et al., 1990). It has been indicated that the inner environment of a living subject has different patterns of physiological function and mechanism. Additionally, therapy–pharmacokinetics (Harada-Shiba et al., 2002), pharmacokinetics and pharmacodynamics (PK/PD) (Liu et al., 2017) are usually not considered within the proposed IVIVE assessment of toxicogenomic data. Eventually, with the comprehensive effects and variables that are determined by the complicated inner-environment of *in vivo* data, an inconsistent gap has emerged between *in vivo* data and *in vitro* data, and this

inconsistency is further reflected at the level of gene expression. Thus, based on TGx data, the properties of a molecule can influence a drug effect that ultimately reaches the cell in different assays (Groothuis et al., 2015; Kramer et al., 2015). Furthermore, the differences in the response components of a living system are reflected differently between gene expression profiles from *in vivo* and *in vitro* data.

To achieve better utilization of IVIVE, it is urgently necessary to find a strategy that can extract the inner-environment factor from *in vivo* TGx data, and based on that, to further develop a valid strategy that is able to more accurately simulate *in vivo* data from *in vitro* data. Non-negative matrix factorization (NMF) (Lee and Seung, 1999; Lee and Seung, 2001), consisting of a series of unsupervised learning methods, is a classical method to factorize a matrix to nonnegative matrixes. NMF is commonly utilized to reduce the computational consumption or dimensions of data, and to filter specific markers or gene selection for genetic data (Dai et al., 2006). As for *in vivo* and *in vitro* data, NMF has been utilized to image analysis, feature selection, and cancer-type classify (Maruyama et al., 2014; Shourijeh et al., 2016). In a previous study, we proposed a post-modified NMF approach to make NMF more suitable for the utilization of biological systems and further applied this approach for deconvoluting gene expression profiles of cancer samples (Liu et al., 2018a). Post-modified NMF is capable of estimating the gene expression profiles and contents of the major factors in samples without any prior reference knowledge. Therefore, post-modified NMF has the potential to factorize major factors based on *in vivo* gene expression profiles.

In this study, we developed a strategy that can extract the inner-environment factor from *in vivo* TGx data at the gene expression level. We first verified the applicable of Post-modified NMF on *in vivo* and *in vitro* profiles. The strategy is able to factorize *in vivo* data into "drug-responding component" and "environmental component" by using post-modified NMF. And then, by combining *in vitro* data with the "environmental component" factorized from *in vivo* data, simulated *in vivo* data was obtained. The results indicated that the simulated *in vivo* data is more compatible with the original *in vivo* data than the use of *in vitro* data directly, and can be utilized to narrow the gap between *in vivo* and *in vitro* data on the gene expression level.

RESULTS

Study Design

We sought to verify the capability of the deconvolution method to decrease the inconsistency between *in vivo* and *in vitro* TGx data and to develop a valid simulation for *in vivo* data using *in vitro* data. All investigations involved two stages: verification and simulation. The verification stage entailed evaluating whether post-modified NMF can be utilized to factorize the inner-environmental factors of *in vivo* data from TGx data. The simulation stage narrowed the gap between two assay types, so that *in vivo* data can be simulated better by *in vitro* assays.

We collected *in vivo* data (single dose/repeat doses) and *in vitro* data (**Table 1**) on 170 compounds from The Open Japanese Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System (Open TG-GATEs) (Uehara et al., 2010; Igarashi et al., 2014). In the verification stage, two group comparisons based on three systems were carried out: *in vivo* single VS *in vitro* and *in vivo* repeat VS *in vitro*. In the simulation stage, two group comparisons based on *in vivo* systems were carried out: *in vivo* single VS original *in vivo* and *in vivo* repeat VS original *in vivo*.

TABLE 1 | Data information and usage in this study.

Species—plat-form	Assay	System	Tissues	Number of compounds	Time points	Dose*	Number of main factors	Involved data	
								Original	Deconvoluted
Sprague-Dawley rat (6 weeks old) —Affymetrix GeneChip® Rat Genome 230 2.0 Array	*In vitro*	*In vitro*	Primary hepatocytes	144	2, 8, 24 h	Control, low, middle, high (0:1:5:25)	1	*In vitro*	None
	In vivo	*In vivo* (single dose)	Liver	158	3, 6, 9, 24 h	Control, low, middle, high (mainly 0:1:3:10)	2	Original *In vivo* (single dose)	Drug-responding component
									Environmental component
		In vivo (repeat doses)	Liver	143	24 h after the last dose after repeated treatment for 3, 7, 14, 28 days	Control, low, middle, high (mainly 0:1:3:10)		Original *in vivo* (repeat doses)	Drug-responding component
									Environmental component

*For the in vitro assay, the highest concentration was defined as the dose level generating an 80–90% relative survival ratio with a ratio of 1:5:25 for the low, middle, and high concentration levels. For the in vivo assay, the highest dose was chosen to match the level that induces the minimum toxic effect in a 4-week toxicity study. Then, the ratio of the low, middle, and high dose levels was appropriately set as 1:3:10.

Figure 1 shows the overall flowchart of our study, including the data and analytical approaches for both the verification stage and application stage.

In Vivo "Drug-Responding Component" Showed Higher Similarity With *In Vitro* Data

In the verification stage, two group comparisons based on three systems were carried out: *in vivo* single dose VS *in vitro* and *in vivo* repeat doses VS *in vitro*. The deconvolution approach was launched to factorize and verify two main factors of the *in vivo* data. Two factors related to body-environment and drug-response in *in vivo* systems were named "environmental component" and "drug-responding component," respectively.

Using factor analysis, our study first confirmed that the expression data of both *in vivo* systems (single dose and repeat doses) contained two main factors (**Figure S1**), whereas the *in vitro* system only had one main factor (discussed in *Confirm the Number of Factors (k Value) for Deconvoluting by Factor Analysis*). When deconvoluting *in vivo* data by post-modified NMF, the matrix of the gene expression profiles of each compound was factorized into two matrices: W^* represented the gene expression profiles, and H^* represented the weights of the two main factors. Additionally, for the two lines in W^*, the line that had a higher correlation (Pearson's correlation coefficient) with corresponding control profiles was profiled as "environmental component," and the other line that had less correlation with corresponding control profiles was profiled as "drug-responding component."

As shown in **Figure 2**, based on the comparison of the consistency obtained between the original *in vivo* data and the *in vitro* data, the "drug-responding component" and the *in vitro*

data achieved higher consistency during the comparison. Both single-dose and repeat-doses *in vivo* data achieved more than 9% extra similarity when using "drug-responding component" instead of using the original *in vivo* data for comparison with the *in vitro* data. In more detail, we describe three indicators to show the improved consistencies achieved by using "drug-responding component." 1) We first compared the similarities [PRank score (Liu et al., 2017)] between the *in vivo* (original and "drug-responding component") and *in vitro* data for each pair of compounds by using the PRank method (**Figures 2A, B**). 2) Then, we counted the number of Dice's coefficients of each compound-pair, revealing different tendencies in *in vivo* (original and "drug-responding component") and *in vivo* data among 7021 compound-pairs. A different tendency means that Dice's coefficients are both lower than the cut-off or both higher than the cut-off in the *in vivo* and *in vitro* data (**Figure 2C**). 3) Additionally, in order to investigate the differences in detail, for each compound-pair, we utilized the absolute difference values of Dice's coefficients to determine how close the gap between the *in vivo* and *in vitro* data was narrowed by deconvoluting the data. Based on 7021 compound-pairs for each system, absolute difference values were obtained by subtracting Dice's coefficients between "drug-responding component" data and *in vitro* data or between the original *in vivo* data and *in vitro* data. Then, the distribution of absolute differences for each system for single-dose and repeat-doses data could be observed and compared, and the mean absolute difference (MAD) of each system could be calculated as well (**Figure 2D**).

For single-dose *in vivo* data, the similarity of "drug-responding component" VS *in vitro* data was 9% higher than the single-dose original *in vivo* data VS *in vitro* data. The PRank score rose from 0.56 before deconvolution to 0.65 after
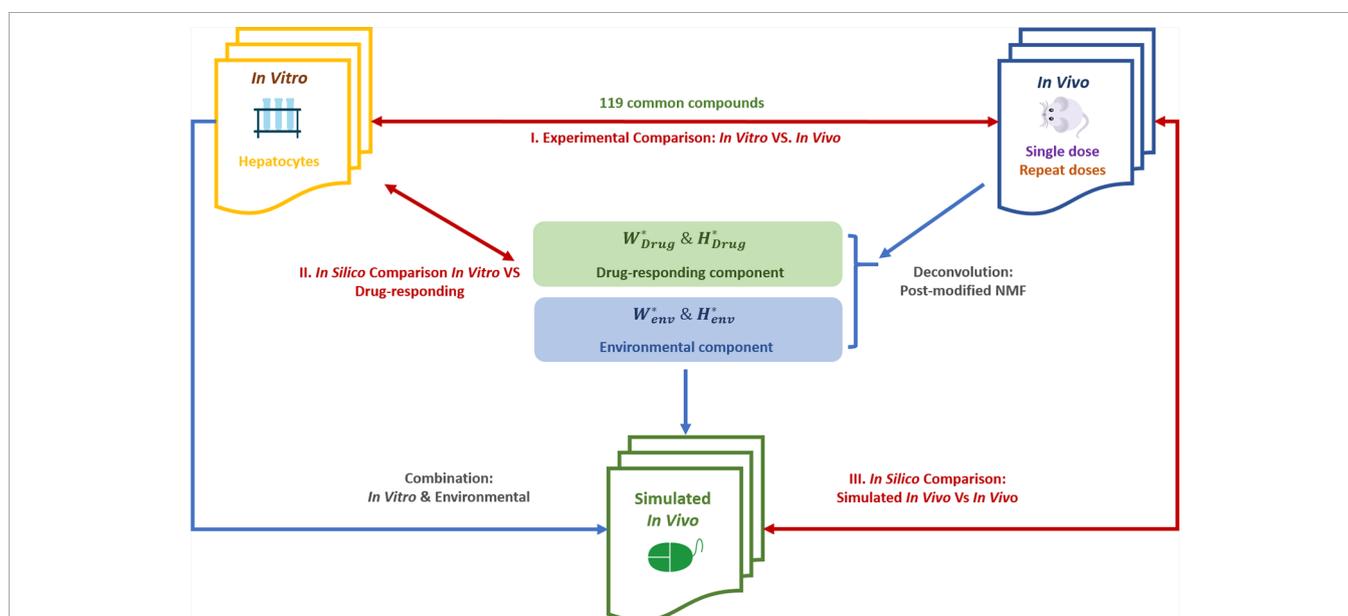
FIGURE 1 | The flowchart of our study design. Post-modified NMF, post-modified non-negative matrix factorization; comparisons accomplished by using the pair ranking method (PRank score) and other indicators. W^*_{drug}, H^*_{drug}, W^*_{env}, H^*_{env} were deconvoluted from experimental *in vivo* data.

FIGURE 2 | The comparisons among original *in vivo* data, *in vitro* data, and the drug-responding component from the *in vivo* data. For all four subplots, purple represents the single-dose data, and yellow color represents the repeat-doses data. In the comparison with original *in vivo* data, the drug-responding component achieved higher consistencies based on both **(A)** a single dose and **(B)** repeat doses. **(C)** The comparison of the number of inconsistent compound-pairs between the *in vivo* and *in vitro* data among each system. **(D)** The violin plots of the absolute difference values between *in vivo* and *in vitro* (|Dice's coefficient of *in vivo* data—Dice's coefficient of *in vitro* data|) systems.

generating the "drug-responding component" (**Figure 2A**). Additionally, an extra 73 compound-pairs were found when comparing the "drug-responding component" with the *in vitro* data. The number of inconsistent compound-pairs was reduced from 193 to 120 (**Figure 2C**, purple). Furthermore, the MAD values were calculated by subtracting Dice's coefficients. The "drug-responding component" VS *in vitro* comparison decreased by 0.02 ("drug-responding component" VS *in vitro* = 0.10, and original *in vivo* data VS *in vitro* = 0.08) (**Figure 2D**, purple). The *p-value* of the significance test (Cohen, 1992) (Student t-test) for the absolute difference values between the "drug-responding component" VS *in vitro* and the original *in vivo* VS *in vitro* data was less than 10×10^{-10}.

Similar results were achieved with repeat-doses *in vivo* data; the score reached 0.79 for the compression of repeat-doses *in vivo* data "drug-responding component" VS *in vitro* data, while the score obtained from repeat-doses *in vivo* data original VS *in vitro* was 0.70, with 9% higher similarity score (**Figure 2B**). Sixteen extra compound-pairs were found when comparing "drug-responding component" with *in vitro* data. The number of inconsistent compound-pairs was reduced from 126 to 110 among 7021 compound-pairs (**Figure 2C**, yellows). Moreover, a lower tendency was observed when comparing the violin plot of absolute differences obtained by subtracting Dice's coefficients between the *in vitro* and *in vivo* assays. The MAD value was reduced from 0.10 to 0.08, the gap between the *in vivo* and

in vitro data decreased by 0.02, and the *p-value* was less than 10×10^{-10} (**Figure 2D**, yellows) when comparing the absolute difference value obtained for the "drug-responding component" VS *in vitro* data and the absolute difference value obtained for the original *in vivo* data VS *in vitro* data, respectively.

To summarize the results of this stage, in both single-dose and repeat-doses data, the "drug-responding component" deconvoluted by post-modified NMF from *in vivo* data showed better results in the investigation among the three indicators than the original *in vivo* data when compared with *in vitro* data. The results indicated that post-modified NMF is able to efficiently factorize the inner-environmental factors and can be used to optimize the consistency between *in vivo* data and *in vitro* data at the gene expression level.

In Vivo "Drug-Responding Component" Can Be Replaced by *In Vitro* Data

In the simulation stage, two group comparisons based on the *in vivo* systems were carried out: *in vivo* single VS original *in vivo* and *in vivo* repeat VS original *in vivo*. Unlike the "drug-responding component," the "environmental component" tended to reflect attributions that are more related to the inner-environment or biological processing of a living body. Thus, the "environmental component" could be regarded as the main factor for examining the differentiation and inconsistency between *in vivo* data and *in vitro* data. When the "environmental component" was confirmed, simulated *in vivo* data were obtained by combining "environmental component" with *in vitro* data. In other words, *in vitro* data were first used to replace (swap) the "drug-responding component" line in W^*, and H^* was integrated with the new W^* to generate the simulated *in vivo* data.

The consistency of the "drug-responding component" and *in vitro* was already confirmed in the validation stage by using the PRank score and Dice's coefficient. However, when replacing (swapping) the "drug-responding component" with *in vitro* data, consistency at the genetic quantification level still needs to be verified to ensure this replacement was feasible. Thus, we investigated the consistency of the "drug-responding component" with *in vitro* data gene by gene for each compound (**Figures 3** and **4**). For each compound, out of 119 compounds, we used gene expression profiles from the "drug-responding component" and *in vitro* data to plot the intensity scatter, then calculated the R^2 and root-mean-square error (RMSE) for evaluation. Each scatter showed the corresponding tendencies of an expressed gene from the "drug-responding component" and *in vitro* profiles.

Both single-dose data and repeat-doses data showed good consistency at the genetic level; most of the dots were arranged around the diagonal line in the intensity scatter chart (**Figures 3A–D**). For single-dose *in vivo* data, the average of R^2 was 0.73, R^2 (**Figure 4A**) ranged from 0.63–0.81, and the average RMSE (**Figure 4B**) reached 3.95. For repeat-doses *in vivo* data, the mean R^2 reached 0.72, R^2 ranged from 0.62–0.80, and the average RMSE was 4.00. Furthermore, the linear regression process could be applied based on the "drug-responding component" and *in vitro* data to obtain a better correlation if necessary (discussed in *How to Improve the Consistency Between In Vitro and "Drug-Responding Component" Data In Vivo*).

In Vivo Data Can Be Simulated Better by Combining *In Vitro* Data With "Environmental Component"

After confirming that "drug-responding component" was replaceable by using *in vitro* data, we determined that combining the correct factor with upcoming *in vitro* data allowed the simulated *in vivo* data to be obtained based on *in vitro* assays without carrying out *in vivo* assays. Since the combined data (simulated *in vivo* data) took the inner-environment into consideration, the simulation demonstrated higher compatibility with the original *in vivo* data than using the *in vitro* data alone (**Figure 5**). The "environmental component" was combined with the corresponding *in vitro* data (details are shown in the *Methods* section) for each compound. Later, the PRank score was applied again to investigate the correlation between simulated *in vivo* data and original *in vivo* data.

For both single-dose and repeat-doses *in vivo* data, the PRank scores were generated based on 7,021 compound-pairs, and high consistencies were observed, with similarity PRank scores higher than 0.72 (**Figures 5A**, **B**). The PRank score between the simulated single-dose data *in vivo* and the original single-dose *in vivo* data was 0.72 *versus* only 0.56 for the *in vitro* data and the single-dose *in vivo* data. For repeat-doses *in vivo* data, the PRank score was 0.70 when directly comparing the *in vitro* data with the original *in vivo* data and rose to 0.75 when comparing the simulated *in vivo data* with the original *in vivo* data. These results indicated that the simulated *in vivo* data had higher consistency with real *in vivo* data than did the *in vitro* data with real *in vivo* data, especially for single-dose *in vivo* systems (discussed in *Extension of Strategy Application*). Furthermore, once the attribution of two components has been confirmed, we not only can use the "drug-responding component" to obtain a better understanding of the mechanism for the drug effect but also use the "environmental component" to make the *in vitro* data become more accessible for real situations in living objects. With the storage of the "environmental component," the upcoming *in vitro* data can always be adjusted and turned into simulated *in vivo* data.

Validation of the Deconvolution Strategy

To further investigate the utilization of the strategy in this study, we applied the strategy to another individual dataset, which including 15 common compounds. Factor analysis also applied to validation dataset, two factors observed from *in vivo* data set and only one main factor observed from the *in vitro* data (**Figure S2**). Thus, the number of factors we used at post-modified NMF of validation is 2.

Similar to the results that observed from using TG-GATEs, when comparing original *in vivo* data with *in vitro* directly, the PRank score is 0.44 (discussed in *Further Perspective and Challenges*). And the Prank score raised to 0.50 when comparing drug-responding component from *in vivo* data with *in vitro* data. Later, simulated *in vivo* data obtained from

FIGURE 3 | The intensity scatter plots reflecting the consistency between *in vivo* and *in vitro* data. **(A** and **C)** Intensity scatter plots for the compounds that have max R^2 (R^2 = 0.81, RMSE = 3.96) and min R2 (R^2 = 0.63, RMSE = 3.97) in single-dose *in vivo* data. **(B** and **D)** Plots for the compounds that have max R^2 (R^2 = 0.80, RMSE = 4.00) and min R2 (R^2 = 0.62, RMSE = 3.96) in repeat-doses *in vivo* data. 9[#], 20[#], 97[#], and 40[#] are acetaminophen, captopril, quinidine, and disopyramide, respectively. Purple color and yellow color in this figure indicate data from the single-dose and repeat-doses systems, respectively.

validated dataset showed higher similarity (Prank score = 0.68) than the original comparison. Notably, after applying the linear regression, PRank score between original *in vivo* data and simulated data reached to 0.77 (discussed in *How to Improve the Consistency Between In Vitro and "Drug-Responding Component" Data In Vivo*) (**Figure S3**). The results implied that, when dealing with the data generated from different platforms (GPL341 VS GPL1355) and at different times (2008 VS 2018), the linear regression is necessary.

All in all, the performance evaluation we have shown above is positive for single-dose *in vivo* data and repeat-doses *in vivo* data in the verification stage and simulation stage, especially for single-dose systems (discussed in *Extension of Strategy Application*). The "drug-responding component" deconvoluted from *in vivo* data achieved higher similarities than the original *in vivo* data when compared with *in vitro* data. The two factors, the "drug-responding component" and the "environmental component," generated from the original *in vivo* data were able to reveal different drug-interaction and inner-environment

functions inside living bodies at the gene expression level. Additionally, this result showed more potential for extending the usage of post-modified NMF to apply to heterogeneous fields, which include the cancer clinical data that we discussed in a previous study (Liu et al., 2018a) and the large-scale toxicogenomics field that we utilized in this study.

DISCUSSION

Confirm the *Number of Factors* (*k* Value) for Deconvoluting by Factor Analysis

By using factor analysis for *in vivo* data, both single-dose and repeat-doses data showed (**Figure S1**) that there were two main factors for a gene expression profile, whereas *in vitro* data only had one factor. Additionally, in terms of the factors of *in vivo* single-dose data, the first factor had a higher score than the second factor. On the other hand, the two main factors for *in vivo* repeat-doses data had almost the same score. This result suggests

FIGURE 4 | The attributions of simulated *in vivo* data. Purple color and yellow color in this figure always indicate data from the single-dose and repeat-doses systems, respectively. **(A** and **B)** The distribution of R^2 among 119 compounds between the original *in vivo* and the simulated *in vivo* data for single-dose and repeat-doses data, respectively. **(C** and **D)** The distribution of RMSE among 119 compounds between the original *in vivo* and the simulated *in vivo* data for single-dose and repeat-doses data, respectively.

the second factor might be "environmental component." Since a compound remains in the repeat-doses *in vivo* data situation for a longer time, the inner-environment would influence the data for repeat-doses more than it influences single-dose *in vivo* data.

Moreover, we confirmed that the number of main factors is 1 for *in vitro* data and 2 for *in vivo* data. As shown in **Figure S1-a**, although there are some minor factors behind the highest one or two columns, the cut-off already contributes more than 99% of factors **Figure S1-b**, specifically 99.61% for the first factor of the *in vitro* data and 99.56% and 99.65% for the first and second factors of the single-dose and repeat-doses *in vivo* data, respectively. Thus, we used k = 2 when we applied post-modified NMF for deconvolution.

How to Improve the Consistency Between *In Vitro* and "Drug-Responding Component" Data *In Vivo*

Technically, if more samples are involved, higher accuracy can be generated during formulization. However, within the Open TG-

GATEs database, for each compound at a different dose and different duration, there are three biological repetitions for *in vivo* data and two biological repetitions for *in vitro* data. The amount of sample that we can utilize in this study was large but still limited. Although this database is capable of helping us prove our verification and launch the simulation, the number of samples is far from sufficient. To pursue better performance and durability, the number of *in vivo* samples used to obtain the "drug-responding component" and the "environmental component" should be as high as possible. Additionally, once the two components have been obtained, infinite *in vitro* data can always be modified and further optimized based on those two components.

Notably, in the Open TG-GATEs database, all the data were generated by the same microarray chip (Affymetrix Support by Product for GeneChip® Rat Genome 230 2.0 Array) in the same relative location. Only minor improvements for single-dose data (0.72–0.72) and repeat-doses data (0.75–0.76) (**Figure S4-a** and **b**) were achieved by applying linear regression in this study, which
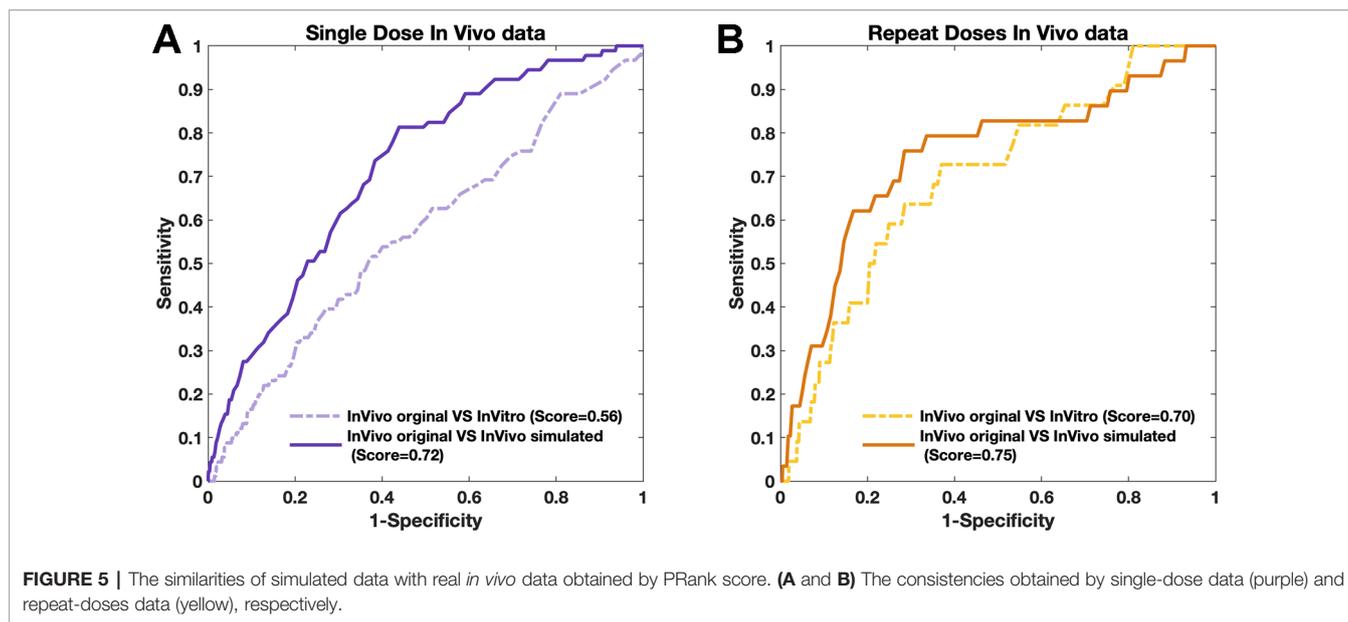
FIGURE 5 | The similarities of simulated data with real *in vivo* data obtained by PRank score. **(A** and **B)** The consistencies obtained by single-dose data (purple) and repeat-doses data (yellow), respectively.

indicated that the data we used in this study were similar in their order of magnitude. As for the validation stage, applying linear regression has improved 9% (from 0.68 to 0.77) on PRank score than that directly combine *in vivo* data with *in vitro* data. For further simulation processing, the "environmental component" might be combined with the upcoming *in vitro* data issued from other batches, other laboratories or even other chips. The expression level of the "environmental component" might be influenced by the experimental batch. Therefore, applying linear regression is suggested to reduce potential bias at the numerical and batch effect levels.

Linear regression ($y = Ax + B$, A is the slope and B is the intercept) was based on "drug-responding component" (y) and *in vitro* data (x) to correct the bias at the numerical and batch effect levels between *in vitro* and *in vivo* systems. This study also demonstrated the distributions of slope and intercept for each compound to investigate the correlation between *in vivo* "drug-responding component" and *in vitro* data. Most of the slopes were in the range of 0.8–0.9 in both the single-dose and repeat-doses data (**Figure S4 c**), and the average slopes for single-dose and repeat-doses *in vivo* data were both 0.86. Additionally, the average intercepts for single-dose and repeat-doses data were 0.31 and 0.35, respectively (**Figure S4 d**). Most of the intercepts had positive values, showing that the data obtained *in vitro* had slightly higher measurements overall than corresponding the "drug-responding component" *in vivo* data. This result indicated that, compared with the "drug-responding component," the concentration of the corresponding compound was attenuated by biological processes *in vivo*, and this bias was observed at the gene expression level by using our strategy.

Repeat-Doses Data Achieved Better Similarity Than Single-Dose Data

According to previous reports, when comparing single-dose data (24 h) and repeat-doses data (28 days) from an *in vivo* assay,

there was a high concordance between the two *in vivo* assay systems, indicating the potential to use a short-term *in vivo* assay for some endpoints, saving time and money (Liu et al., 2018b). However, for comparisons between *in vitro* and *in vivo* data, the *in vitro* TGx data set had a relatively higher similarity to the repeat-doses *in vivo* data (0.70) than did the single-doses *in vivo* data (0.56), suggesting better correlation of the *in vitro* assays with the longer-term *in vivo* assays. Specifically, gene activities associated with the survival of hepatocytes reflect a level of adaptation that resembles those under 28-day repeated dosing conditions. One explanation could be that 24 h (single dose) *in vivo* is simply the time frame in which the liver responds to a new chemical stressor and the inner-environment sways the expression level more than drug effect. Conversely, after 28 days (repeat doses) of *in vivo* treatment, there is a greater drug effect on the expression level with multiple treatments, making repeat doses more equivalent to the response of hepatocytes in cell culture (*in vitro* assay). These differences can also be observed by the factor analysis procedure (**Figure S1**); the two main factors generated by a repeat-doses data set were more even than the factors generated by a single-dose data set. Notably, after applying the simulation strategy, the similarities of the simulated data generated from the *in vitro* data set with single-dose and repeat-doses *in vivo* data were improved 16% and 5%, respectively. This result indicated that the deconvolution strategy can be an efficient way to improve IVIVE, especially in terms of assisting with the utility of short-term *in vivo* assays.

Extension of Strategy Application

Technically speaking, all kinds of data (kidney, liver *etc.*) that meet follow criteria can be processed by the deconvoluting strategy in this study: a. There are normal samples as blank control (to identify drug effect and inner-environmental components respectively). b. The number of samples is bigger than the number of components (in this study is 2). c. The *in*

vitro data which is going to participate the simulation, is generated by applying same chemical compound. Additionally, same sequencer (platform), and post-processing of *in vivo* and *in vitro* data are highly recommended.

In this study, the data involved a large number of different compounds and treatment durations, also applied to relatively small dataset (validation), but we did not find any available sequencing data that matched our study strategy. The lack of large-scale sequencing data in this context makes this struggling for has not demonstrated an approach for sequencing data. However, we performed RNA sequencing to prove the stability and reliability of post-modified NMF in our previous study, which contained sequencing data obtained from multiple human tissues and human cancers (Liu et al., 2018a). This strategy has promising potential to be utilized with toxicogenomic sequencing data in the future. We will continue collecting related data and try to apply the strategy to sequencing data once we find an appropriate and available dataset. Additionally, it is reasonable to believe that with the development of emerging genetic technologies, the strategy that we proposed might be able to utilize miRNA, ncRNA, and other kinds of genetic data.

Further Perspective and Challenges

In this study, a computational strategy is presented to improve IVIVE problem *in silico*. The strategy is able to extract drug effect and inner-environmental components from the original *in vivo* data, further by combining *in vitro* data with inner-environmental component to narrow the gap between two experiential systems. Moreover, by keeping inner-environmental components that been obtained, more simulations can be perform with upcoming *in vitro* data.

Nevertheless, several directions still remain to investigate. As for NMF algorithm which is the focus of this study, reference-free is an advantage for NMF but also can be regarded as a delicate factor during iterations. Iterations need to avoid local optimize and also to guarantee the stability and robustness (*Post-Modified NMF Deconvolution Method*). For biological part, as we mentioned, due to conventions at processing, the number of replications of samples is lacking for mathematical calculation. And the quality control of batch affect, environmental difference, multi-platform operations are also need to count into consideration.

For this study, the different inter-lab and different inter-platform problem still needs to be further discussed. Thus, for the 119 compounds and 15 compounds that we performed in application and validation parts, respectively, we shared tables for the "environmental component" (W_{env}^{*}), and the average of the corresponding weight matrix ($H_{in\ vivo}^{*}$) obtained in this study for further research in the supplementary materials (**Tables S3–S5**). Therefore, if any available *in vitro* data generated with the corresponding compounds is included in this study, the list can be utilized to transfer the *in vitro* data into simulated *in vivo* data (details shown in *Generation of Simulated In Vivo Data Based on In Vitro Data*). Additionally, we believe that with the utilization of our strategy, the variation caused by inter-lab and inter-platform situations can be further investigated.

MATERIALS AND METHODS

Materials

Toxicogenomics Database

The rat data were downloaded from a large-scale toxicogenomics database named Open TG-GATEs (Open Toxicogenomics Project-Genomics Assisted Toxicity Evaluation Systems, http://toxico.nibiohn.go.jp/english/) (Uehara et al., 2010; Igarashi et al., 2014). This is a database that stores gene expression profiles and traditional toxicological information obtained *in vivo* (rat) and *in vitro* (primary rat hepatocytes). In total, 170 compounds with multiple doses, measure-times, and treatment durations were involved. In more detail, there are two kinds of treatment durations for *in vivo* data: one treatment duration is a single-dose trial, and the other treatment durations is a repeat-doses trail. Three systems in this study specifically included gene expression data obtained *in vitro* and *in vivo* (single dose/repeat doses). In other words, each compound involved *in vitro* data, single-dose *in vivo* data, and repeat-doses *in vivo* data.

For each compound, the *in vitro* data consisted of three doses (low, medium, and high) and three treatment time points (2, 4, and 24 h). The *in vivo* data were obtained from adult rats that were 6 weeks old. For single-dose data, the rats were treated with three doses (low, medium, and high), and liver tissue was collected at four time points (3, 6, 9, and 24 h) after treatment. For *in vivo* repeat-doses data, the rats were treated with three doses (low, medium, and high) and with different treatment durations (3, 7, 14, and 28 days). After the last exposure and dose the animals were killed 24 h later and that liver tissue was collected and isolated. Every time point (or duration) had corresponding control samples.

To reduce the influence of irrelevant variables on the measurements to guarantee external validity, the data with the highest doses, longest timepoints, and longest durations, as well as their corresponding control data, were used for our investigation. Specifically, the "*in vitro*" data are *in vitro* data that were obtained with the high dose at 24 h, "*in vivo* single" data are single-dose *in vivo* data that were obtained with the high dose at 24 h, and "*in vivo* repeat" are the repeat-doses *in vivo* data generated under the high dose at 28 days. A total of 24,023 biological samples can be utilized in Open TG-GATEs. Based on 119 common compounds (**Table S1**), 7,021 pairwise combinations for compounds were generated.

The Generation of Gene Expression Profiles

For application data, the microarray data downloaded from Open TG-GATEs were processed by Factor analysis for Robust Microarray Summarization (FARMS) (Hochreiter et al., 2006). For each compound, there were three and two replicate samples for *in vivo* and *in vitro* data for every time point, respectively. Every dosing sample had its matched control samples as well. Specifically, there were three control samples *in vivo* and two control samples *in vitro* for each time point for each compound. After quantile normalization of the probe-level data, we calculated the probe intensity ratios by referencing the corresponding control measurement for the blank cell culture

(without compound) to correct the compound batch. Next, the probe-level names were transformed into their corresponding gene-level names by using a CDF (Version 15.1.0, ENTREZG) file (Dai et al., 2005). Later, in order to obtain expression values per gene, intensity ratios at the probe set level were summarized.

The original *in vivo* single-dose data were stored as a matrix 13,934 rows and 715 columns, and each column represented the group of 13,934 gene expression profiles for one biological sample. Similarly, the *in vivo* repeat-doses data were a matrix with 13,934 rows and 706 columns (control samples in repeat doses of some compounds had no biological repetition), and *in vitro* data were 13,934 rows and 476 columns. We marked the gene expression profiles of the *in vitro* data, single-dose *in vivo* data, and repeat-doses *in vivo* data as "*in vitro*," "*in vivo* single," and "*in vivo* repeat," respectively.

Validation Datasets

The validation rat data contains two microarray datasets from different years and platforms, and can be both downloaded from NCBI-GEO database (Barrett et al., 2012). The *in vivo* data is obtained from data set (GSE68110) at 2008 (Ellinger-Ziegelbauer et al., 2008) on Affymetrix Rat Expression 230A Array platform. 30 compounds included with different dosages (low/middle/high). And the *in vitro* data set (GSE119933) was generated coordinate with *in vivo* at 2018 (Grinberg et al., 2018) on Rat Genome 230 2.0 array, 29 compounds with different treatment durations (1d/3d/7d/14d) are involved. Similar to the usage of GT_GATE, only the compounds that have control samples and at the highest dosage or longest treatment would be utilized for this study. That is, samples with high dosage group from *in vivo* data, and 14d treatment from *in vitro* data are involved in this study (**Table S2**).

Data preprocessing and all subsequent analyses were performed using MATLAB as well. After filtering, 15 common compounds (**Table S2**) are selected. Each compound has three experimental samples with three controls for *in vivo* and *in vitro* data, respectively. And 10324 common genes left after overlapping two platforms, 105 pairwise combinations for compounds were generated.

Methods

Decision of the Number of Main Factors

Factor analysis was first released by J. Pearson (Holgado–Tello et al., 2010) and C. Spearman (Woolley et al., 2010) to determine the number of species; then, R. M. Wallace further added the matrix rank method to address multicomponent systems (Wallace, 1960; Wallace and Katz, 1964). We used factor analysis to decide upon the number of main factors in our study (discussed in *Confirm the Number of Factors (k Value) for Deconvoluting by Factor Analysis*). This statistical method (Bartholomew et al., 2008) is able to describe a group of main components generated among observed, correlated variables. Components can be regarded as several factors that potentially can lower the number of unobserved variables and maintain the main characters of the variables in the meanwhile. For example, in our case, it is possible that variations in multiple observed variables mainly reflect the variations in two underlying variables. Factor analysis searches for such joint factors in

response to unobserved latent variables. Hence, factor analysis helps to deal with data sets in which there are large numbers of observed variables that can be reflected by a smaller number of latent variables.

Factor analysis was first used in psychometrics and then commonly used in chemistry (Subbarao et al., 1996), biology (Meng et al., 2011; Love et al., 2004), personality theories (Ford et al., 1986; Cattell, 1987; Kahn, 2006), and marketing (Stewart, 1981; Churchill and Iacobucci, 2006; Polit and Beck, 2008). Factor analysis can compress large data to achieve higher data quality, investigate a significant explanation, and simplify completed problems. The original data matrix is marked as *V*. By recombining observed variables linearly, the original matrix *V* can be represented with a group of new underlying variables. To generate the factors of *V*, the covariance matrices *Z* is calculated by *V*.

$$Z = V'V \tag{1}$$

The size of *V* is $r \times c$; then, the size of *Z* is $c \times c$. Then, we diagonalize the covariance matrix *Z*, and the diagonalized matrix λ is generated.

$$Q^{-1}ZQ = \begin{bmatrix} \lambda_1, 0, \dots, 0 \\ 0, \lambda_2, \dots, 0 \\ \dots \\ 0, 0, \dots, \lambda_c \end{bmatrix} = \lambda \tag{2}$$

$Q=[q_1,q_2,\dots,q_c]$ is the matrix consisting of eigenvectors and meeting the orthogonality ($Q^{-1} = Q^\tau$, the symbol "τ" in formulas represents the transposition of the corresponding matrix), Q^{-1} is the inverse matrix of Q, and λ_i is the eigenvalue that meets the criteria:

$$Zq_j = \lambda_i q_j, \; where \; j = 1, \dots, c. \tag{3}$$

Then, we can utilize λ to calculate reduced eigen value (REV) based on *V* by the following equation:

$$REV_j = \lambda_j / (r - j + 1)(c - j + 1) \tag{4}$$

Furthermore, a list of ratios could be derived by the REVs:

$$ratio_t = \frac{REV_t}{REV_{t+1}}, \; where \; t = 1 \dots j - 1. \tag{5}$$

Eventually, the maximum ratio can be regarded as the most significant factor, and the minimum ratio can be regarded as the least significant factor. In this study, the number of factors that added more than 99% cumulative contribution were considered main factors (details shown in *Confirm the Number of Factors (k Value) for Deconvoluting by Factor Analysis*). Those main factors were able to reflect *V* and could be visualized to obtain better observations.

Post-Modified NMF Deconvolution Method

Non-negative matrix factorization (NMF) is a series of unsupervised learning methods that is able to factorize a nonnegative matrix *V* into 2 nonnegative matrices (*W* and *H*) (Lee and Seung, 1999). The post-modified NMF that we used in

this study is an unsupervised learning algorithm that is capable of estimating the gene expression profiles and contents of the major components in samples without any prior reference knowledge. NMF is frequently used in blind source separation because of its nonnegative conception. The ultimate objective of NMF can be described as: given a matrix V, NMF finds nonnegative matrices W and H. $W \cdot H$ is a lower-rank approximation of V. NMF estimates the basis matrix W ($m \times k$ nonnegative matrices) and the coefficient matrix H ($k \times n$ nonnegative matrices) from the original matrix V ($m \times n$ nonnegative matrix). In practical application, the factorization rank k should meet the conditions that $k \ll \text{Min}(m, n)$. After deconvoluting and iterating from the original matrix V to the sum of the loss function F (V, WH) (Gaujoux and Seoighe, 2010), a regularization function F (W, H) is reached at the minimum, and the optimized W and H are generated (Hoyer, 2004; Kim and Park, 2007). The original NMF algorithm can be written as:

$$V \approx W \cdot H \qquad (6)$$

$$V_j = \sum_{i=1}^{k} W_j H_{ij} \, (j = 1, 2, …, the\ number\ of\ columns\ in\ V) \qquad (7)$$

The matrices W and H were estimated using an alternating least squares (ALS) algorithm, which was first proposed by Paatero and Tapper (Paatero and Tapper, 1994) and improved by Paatero and Albright et al. (Paatero, 1999). In the ALS algorithm, W is initialized as a random dense matrix and used to solve H using a least squares calculation step. The negative elements in H are set to 0. The loss function F (V, WH) is applied to measure the result of each iteration. In this function, the factors W and H are chosen to minimize the root-mean-squared residual D (the cost function) between V and $W \cdot H$. That is, when loss function F (W, H) is reached at the minimum, the optimized W and H are generated. The iteration procedure is iterated until W and H can minimize the cost function D (Berry et al., 2007). The loss function F algorithm can be written as:

$$F(W, H) = \| V - WH \|_F^2 \qquad (8)$$

Moreover, for better biological utilization of a practical situation, a normalized H^* can be generated by normalizing the H after deconvoluting. Specifically, by restricting the sum of the values in each column of H to 1, a normalized H^* is generated (k components × n samples).

$$H = \left[h_1, h_2, …, h_j \, (j\ is\ the\ number\ of\ columns\ in\ V) \right] \qquad (9)$$

$$H_{ij}^* = \frac{H_{ij}}{\sum_{l=1}^{k} H_{lj}} \qquad (10)$$

After H being normalized as H^*, with known H^* and original matrix V, W^* can be generated accordingly. Notably, as the W^* might contain a tiny amount of negative values (less than 0.01% in all compounds), we set the negative values to zero to keep the matrix nonnegative. Similar to NMF, the essence of post-modified NMF (Liu et al., 2018a) is a multivariate linear

model. Each column of V is approximately expressed as a linear combination of the column vectors in W^* and the coefficient matrix H^* of corresponding columns. The procedure results in a new W^* as follows:

$$W^* = V H^{*\tau} \left(H^* H^{*\tau} \right)^{-1} \qquad (11)$$

The symbol "τ" in formulas represents the transposition of the corresponding matrix.

In our study, for each compound in every system, the matrix V was the original gene expression data, including the expression of m genes in n mixed samples ($m \times n$ matrix), where m represents the number of genes, and n represents the number of samples. W^* was a deconvoluted signal matrix, including the expression profiles of m genes in k individual component-types ($m \times k$ matrix), and H^* was a weight matrix that included the relative weight of k component types in n mixtures ($k \times n$ matrix). Matrices V, W^*, and H^* were all nonnegative.

Since the algorithm for generating W^* and H^* needs iterations for parameter optimization, which might lead to local minima, repeated factorizations with random initial W^* and H^* may yields different W and H pairs, which might include less optimal results. To ensure reliable and conforming results, the deconvolution procedure was repeated 100,000 times in total for every compound. After 1,000 inner-iterations with a different initial W^* and H^*, to select the results with the minimum root-mean-squared error, yielded from (Boess et al., 2003), and then repeat the whole processing 100 times to get the average result. The averaged results were used for further analysis to ensure reliability and optimization.

After factor analysis in *Decision of the Number of Main Factors*, 2 factors were yielded from the *in vivo* data. We sought to determine which factor represents the inner-environment of the inner body and which factor represents the response of the drug effect. After deconvoluting by post-modified NMF, the matrix of the gene expression profiles of each compound for *in vivo* data was factorized into two matrices: W^* represented the gene expression profiles, and H^* represented the weights of two factors. We compared each column of W^* with the corresponding control sample for each compound by using pairwise Pearson correlation coefficients. The column with the higher coefficient was considered the "environmental component," while the other column with the lower coefficient from the control sample was considered the "drug-responding component." Thus, each column of W^* in formula (Cattell, 1987) (*Post-Modified NMF Deconvolution Method*) represented the estimated expression levels of m genes in a component, which could be compared with the corresponding expression levels of the matched control sample.

Identification of Biologically Significant Genes
The biologically significant genes identified between the dosing samples and control samples may be potentially correlated with phenotypic differences. For each compound in each assay system, fold change (FC) values were generated by comparing the dosing group VS the matched control group. Then, we ranked the genes from the highest FC value to the lowest, and significant genes

were selected from the top and bottom of the ranked list. To find the best cut-off number of significant genes for our study, we investigated the stability of ranked FC lists of each compound by cutting at different amounts (from 50 to 550 by step 50). A stable trend appeared after 200 genes from the ranked FC list. Hence, we used 200 as the cut-off for selecting significant genes that were both up- and downregulated. Furthermore, based on the significant genes for each compound (**Figure S5**), the pairwise similarity could be calculated for PRank processing.

Pair Ranking (Prank) Method

The Pair Ranking (PRank) method was used to investigate the consistency among the three rat TGx assay systems (Liu et al., 2017; Liu et al., 2018b). The PRank score was utilized to measure the similar extent between every system in the validation part as well as in the application. Similar to the definition of the ROC curve, the similarity score ranged from 0 to 1, and the higher score was better.

Biologically significant genes were ranked according to their fold change value. The procedure utilized to find the threshold was the same as that in Liu's work, and the threshold of 200 was kept afterwards. That is, biologically significant genes were obtained for each compound by finding the top and bottom 200 ranked genes by their fold change values (**Figure S3**).

Dice's coefficient was employed to calculate the similarity between the gene expression profiles of compounds, as suggested

by the SEQC I study (Wang et al., 2014; Xu et al., 2016). The pairwise compound similarity of any two compounds within a system was calculated by using the total number of 400 genes obtained for each compound. Notably, the regulated direction of overlapping genes was taken into consideration in this situation.

$$\text{Dice's coefficient} = \frac{N_{i,j,up} + N_{i,j,down}}{400} \quad (12)$$

$N_{i,j,up}$ and $N_{i,j,down}$ indicate the number of overlapping up- and downregulated genes, respectively, between compound i and j.

The pairwise similarities are ranked from highest to lowest in each system separately. Eventually, by using a receiver operating (ROC) curve and the area under the curve (AUC), the PRank score between any two systems can be calculated. To obtain the PRank score, the ranked Dice's coefficient is transferred into a binary model (0/1) with 0.4 as the cut-off. This cut-off is close to the 95% quantile. The built-in function *perfcurve* in MATLAB R2018b was applied for ROC-AUC calculation.

Generation of Simulated *In Vivo* Data Based on *In Vitro* Data

After deconvolution and component confirmation of *in vivo* data, we determined that the *in vitro* gene expression profile had high similarity with the "drug-responding component" deconvoluted from the *in vivo* data. Additionally, the "environmental component" tends to reflect more attribution

FIGURE 6 | Step by step illustration of generating the simulated *in vivo* data by combining *in vitro* data with environmental component from *in vivo* data.
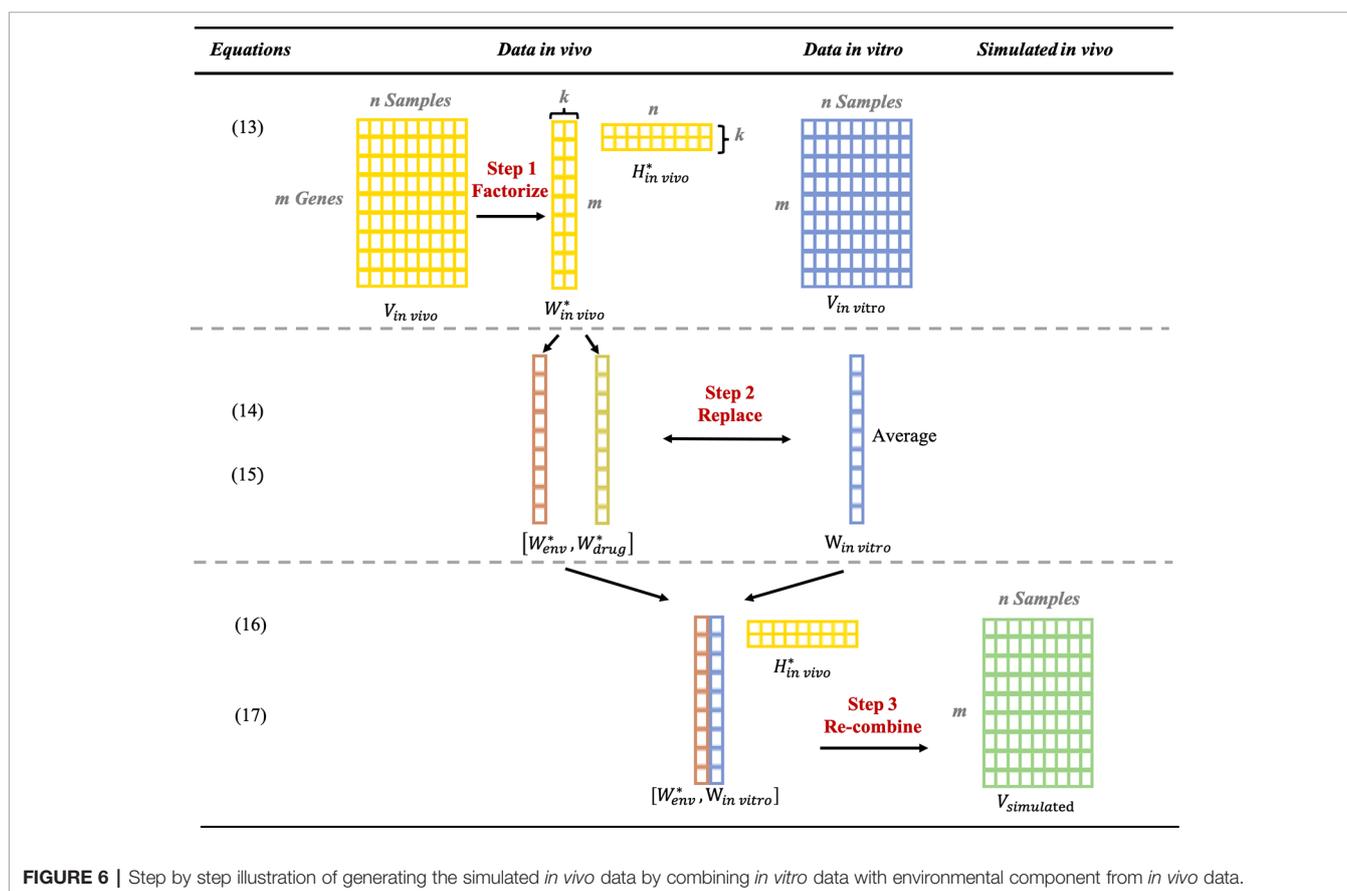
of the inner body. Thus, once the "environmental component" was obtained, we obtained the simulated *in vivo* data based on the *in vitro* expression profile by replacing the "drug-responding component" with the *in vitro* data. The replacement procedure can be formulized as follows:

$$V_{in\ vivo} = W^{*}_{in\ vivo} H^{*}_{in\ vivo} \tag{13}$$

$V_{in\ vivo}$ is the original gene expressed profile for *in vivo* data which contains m genes (rows) × n sample (columns). The gene expression profile $V_{in\ vivo}$ is deconvoluted into $W^{*}_{in\ vivo}$ ($m \times k$ factors) and $H^{*}_{in\ vivo}$ (k factors × n) by using post-modified NMF (**Figure 6**). $W^{*}_{in\ vivo}$ is composed of two columns of gene expression profiles, W^{*}_{env} and W^{*}_{drug}, corresponding to the "environmental component" and the "drug-responding component," respectively; and $H^{*}_{in\ vivo}$ is composed of two rows of gene expression profiles, H^{*}_{env} and H^{*}_{drug}, corresponding to the weight of "environmental component" and the "drug-responding component," respectively; the order of these two columns depends on their similarity with matched control samples.

$$W^{*}_{in\ vivo} = \left[W^{*}_{env}, W^{*}_{drug} \right] \tag{14}$$

To prepare for replacement, the *in vitro* data V_{invivo} is formed into the same size (1 column × the number of genes) as W^{*}_{drug}. The average of the *in vitro* data for each compound is utilized, which we named $w_{in\ vivo}$.

$$W_{in\ vitro} = avg\left(V_{in\ vitro} \right) \tag{15}$$

Later, by replacing W^{*}_{drug} with $W_{in\ vitro}$, $W_{modified}$ is generated.

$$W_{modified} = \left[W^{*}_{env}, W_{in\ vitro} \right] \tag{16}$$

Eventually, with the recombination of $w_{modified}$ and $H^{*}_{in\ vivo}$, the simulated data $V_{simulated}$ are obtained.

$$V_{simulated} = W_{modified} H^{*}_{in\ vivo} \tag{17}$$

According to (Cohen, 1992) and (Dai et al., 2006), by keeping $H^{*}_{in\ vivo}$ and w_{env}, the simulated data can always be obtained based on *in vitro* data that are currently available or *in vitro* data that are upcoming.

Note that if linear regression is applied to correct $w_{in\ vitro}$ in equation (Cohen, 1992), then $W_{in\ vitro}$ would be modified by W^{*}_{drug}. Thus, in this situation, $H^{*}_{in\ vivo}$ and $W^{*}_{in\ vivo}$ and (W^{*}_{env} and W^{*}_{drug}) are needed.

All calculation procedures and the identification of significant genes were conducted in MATLAB (MathWorks®, R2018b). In order to make Post-modified NMF and corresponding strategy easier to use, we have released a functional MATLAB package with detailed tutorial at github. (https://github.com/annlyuan/Post_modified_NMF).

CONCLUSIONS

In this study, an *in silico* strategy based on post-modified NMF was proposed to factorize the inner-environmental factor from *in*

vivo assays at the gene expression level. Drug effect and inner-environmental components were obtained from the *in vivo* data. This strategy first verified its applicability to TGx data and then simulated the *in vivo* data by correcting the *in vitro* data. Similarities between real *in vivo* data and simulated data were higher than those obtained by directly comparing real *in vivo* data with *in vitro* data. The results indicated that this strategy could promptly generate substitutions for *in vivo* TGx. Additionally, a simulation can be generated by using *in vitro* data to reduce the launch of live-animal tests. Eventually, the gap between *in vivo* and *in vitro* data at the gene expression level is effectively narrowed.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://toxico.nibiohn.go.jp/open-tggates/english/search.html.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2019.01489/full#supplementary-material

FIGURE S1 | The bar graphs for the factors obtained from each system after factor analysis. **(A)** Represents the reduced eigen values (REVs) calculated by factor analysis of each factor in *in vitro* (left), single-dose *in vivo* (middle) and repeat-doses *in vivo* (right) data. **(B)** Represents the ratios of contribution calculated based on REVs in *in vitro* (left), single-dose *in vivo* (middle) and repeat-doses *in vivo* (right) data.

FIGURE S2 | The bar graphs for the factors obtained from each system of validation data set after factor analysis. **(A)** Represents the reduced eigen values (REVs) calculated by factor analysis of each factor in *in vivo* (upper), and the ratios of contribution calculated based on REVs of *in vivo* (down) data. **(B)** Represents the reduced eigen values (REVs) calculated by factor analysis of each factor in *in vitro* (upper), and the ratios of contribution calculated based on REVs of *in vitro* (down) data.

FIGURE S3 | Results of validation dataset. **(A)** In the comparison with original *in vivo* data (light blue), the drug-responding component achieved higher

consistencies. **(B)** The comparison of applying linear regression or not. **(C)** The violin plots of the absolute difference values between *in vivo* and *in vitro* (\Dice's coefficient of *in vivo* data - Dice's coefficient of *in vitro* data\) of validation datasets.

FIGURE S4 | The attributions of applying linear regression to *in vitro* and "drug-responding compound" *in vivo* data. Purple color and yellow color in this figure always indicate the data from single-dose and repeat-doses systems, respectively. **(A)** and **(B)** The consistencies between original *in vivo* data and simulated *in vivo* data obtained by PRank score. The gray lines indicate the scores obtained based on the simulated data (which are derived from applied linear regression between "drug-responding compound" and *in vitro* data) compared with original *in vivo* data. **(C)** and **(D)** The slope and intercept distributions, respectively, after launching linear regression between original *in vivo* and simulated *in vivo* data.

FIGURE S5 | The stability of the ranked similarity list for each assay and system. The data points were calculated by comparing the ranked similarity list with different numbers of differentially expressed genes by using Spearman's correlation coefficient.

TABLE S1 | The list of 119 common compounds.

TABLE S2 | The sample information of validation dataset (NA: no information available, h: hour, d: day).

TABLE S3 | The W_{env}^* and the average of H_{env}^* obtained by single-dose in vivo data (format as .xlsx).

TABLE S4 | The W_{env}^* and the average of H_{env}^* obtained by repeat-doses in vivo data (format as .xlsx).

TABLE S5 | The W_{env}^* and the average of H_{env}^* obtained by validation in vivo data (format as .xlsx).

REFERENCES

Aardema, M. J., and MacGregor, J. T. (2003). "Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies," in *Toxicogenomics* (Tokyo: Springer), 171–193. doi: 10.1007/978-4-431-66999-9_22

Abbott, A. (2005). Animal testing: more than a cosmetic change. *Nature* (London: Nature Publishing Group). 438, 144. doi: 10.1038/438144a

Aigner, A., Buesen, R., Gant, T., Gooderham, N., Greim, H., Hackermüller, J., et al. (2016). Advancing the use of noncoding RNA in regulatory toxicology: report of an ECETOC workshop. *Regul. Toxicol. Pharmacol.* 82, 127–139. doi: 10.1016/j.yrtph.2016.09.018

Alam, M. N., Bristi, N. J., and Rafiquzzaman, M. (2013). Review on *in vivo* and *in vitro* methods evaluation of antioxidant activity. *Saudi Pharm. J.* 21 (2), 143–152. doi: 10.1016/j.jsps.2012.05.002

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2012). NCBI GEO: archive for functional genomics data sets—update *Nucleic Acids Res.* 41, D1, D991–D995. doi: 10.1093/nar/gks1193

Bartholomew, D. J., Steele, F., Galbraith, J., and Moustaki, I. (2008). *Analysis of multivariate social science data* (Tokyo and Francis: Chapman and Hall/CRC).

Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., and Plemmons, R. J. (2007). Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data Anal.* 52 (1), 155–173. doi: 10.1016/j.csda.2006.11.006

Boess, F., Kamber, M., Romer, S., Gasser, R., Muller, D., Albertini, S., et al. (2003). Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the *in vivo* liver gene expression in rats: possible implications for toxicogenomics use of *in vitro* systems. *Toxicol. Sci.* 73 (2), 386–402. doi: 10.1093/toxsci/kfg064

Boverhof, D. R., and Zacharewski, T. R. (2005). Toxicogenomics in risk assessment: applications and needs. *Toxicol. Sci.* 89 (2), 352–360. doi: 10.1093/toxsci/kfj018

Bumgardner, G. L., Matas, A. J., Chen, S., Cahill, D., Cunningham, T. R., Payne, W. D., et al. (1990). Comparison of *in vivo* and *in vitro* immune response to purified hepatocytes. *Transplantation* 49 (2), 429–436. doi: 10.1097/00007890-199002000-00039

Cattell, R. B. (1987). *Intelligence: its structure, growth and action* (New York: Elsevier).

Chen, M., Zhang, M., Borlak, J., and Tong, W. (2012a). A decade of toxicogenomic research and its contribution to toxicological science. *Toxicol. Sci.* 130 (2), 217–228. doi: 10.1093/toxsci/kfs223

Chen, Y., Jin, J. Y., Mukadam, S., Malhi, V., and Kenny, J. R. (2012b). Application of IVIVE and PBPK modeling in prospective prediction of clinical pharmacokinetics: strategy and approach during the drug discovery phase with four case studies. *Biopharm. Drug Dispos.* 33 (2), 85–98. doi: 10.1002/bdd1769

Chen, M., Bisgin, H., Tong, L., Hong, H., Fang, H., Borlak, J., et al. (2014). Toward predictive models for drug-induced liver injury in humans: are we there yet? *Biomarkers Med.* 8 (2), 201–213. doi: 10.2217/bmm.13.146

Churchill, G. A., and Iacobucci, D. (2006). *Marketing research: methodological foundations* (New York: Dryden Press).

Cohen, J. (1992). Statistical power analysis. *Curr. Dir. Psychol. Sci.* 1 (3), 98–101. doi: 10.1111/1467-8721.ep10768783

Dai, M., Wang, P., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 33 (20), e175–e175. doi: 10.1093/nar/gni179

Dai, J. J., Lieu, L., and Rocke, D. (2006). Dimension reduction for classification with gene expression microarray data *Stat. Appl. Genet. Mol. Biol.* 5 (1). doi: 10.2202/1544-61151147

Dempsey, J. L., and Cui, J. Y. (2017). Long non-coding RNAs: a novel paradigm for toxicology. *Toxicol. Sci.* 155 (1), 3–21. doi: 10.1093/toxsci/kfw203

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., and Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* 95, 5–12. doi: 10.1093/toxsci/kfl103

Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A., and Ahr, H. J. (2008). Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term *in vivo* studies. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* 637 (1–2), 23–39. doi: 10.1016/j.mrfmmm.2007.06.010

Eskildsen, T., Taipaleenmäki, H., Stenvang, J., Abdallah, B. M., Ditzel, N., Nossent, A. Y., et al. (2011). MicroRNA-138 regulates osteogenic differentiation of human stromal (mesenchymal) stem cells *in vivo*. *Proc. Nat. Acad. Sci.* 108 (15), 6139–6144. doi: 10.1073/pnas.1016758108

Fagan, A. M., Mintun, M. A., Mach, R. H., Lee, S. Y., Dence, C. S., Shah, A. R., et al. (2006). Inverse relation between *in vivo* amyloid imaging load and cerebrospinal fluid Aβ42 in humans. *Ann. Neurol.* 59 (3), 512–519. doi: 10.1002/ana.20730

Ford, J. K., MacCallum, R. C., and Tait, M. (1986). The application of exploratory factor analysis in applied psychology: a critical review and analysis. *Personnel Psychol.* 39 (2), 291–314. doi: 10.1111/j.1744-6570.1986.tb00583.x

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinf.* 11 (1), 367. doi: 10.1186/1471-2105-11-367

Glasser, M. F., and Van Essen, D. C. (2011). Mapping human cortical areas *in vivo* based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* 31 (32), 11597–11616. doi: 10.1523/JNEUROSCI.2180-112011

Grinberg, M., Stöber, R. M., Albrecht, W., Edlund, K., Schug, M., Godoy, P., et al. (2018). Toxicogenomics directory of rat hepatotoxicants *in vivo* and in cultivated hepatocytes. *Arch. Toxicol.* 92, 12, 3517–3533. doi: 10.1007/s00204-018-2352-3

Gronthos, S., Mankani, M., Brahim, J., Robey, P. G., and Shi, S. (2000). Postnatal human dental pulp stem cells (DPSCs) *in vitro* and *in vivo*. *Proc. Nat. Acad. Sci.* 97 (25), 13625–13630. doi: 10.1073/pnas.240309797

Groothuis, F. A., Heringa, M. B., Nicol, B., Hermens, J. L. M., Blaauboer, B. J., and Kramer, N. I. (2015). Dose metric considerations in *in vitro* assays to improve quantitative *in vitro–in vivo* dose extrapolations. *Toxicology* 332, 30–40. doi: 10.1016/j.tox.2013.08.012

Hamburg, M. A. (2011). "Advancing regulatory science," in *American Association for the Advancement of Science*. doi: 10.1126/science.1204432

Harada-Shiba, M., Yamauchi, K., Harada, A., Takamisawa, I., Shimokado, K., and Kataoka, K. (2002). Polyion complex micelles as vectors in gene therapy–pharmacokinetics and *in vivo* gene transfer. *Gene Ther.* 9 (6), 407. doi: 10.1038/sj.gt.3301665

Hochreiter, S., Clevert, D.-A., and Obermayer, K. (2006). A new summarization method for Affymetrix probe level data. *Bioinformatics* 22 (8), 943–949. doi: 10.1093/bioinformatics/btl033

Holgado–Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., and Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual. Quant.* 44 (1), 153. doi: 10.1007/s11135-008-9190-y

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.* 5 (Nov), 1457–1469.

Hussain, T., Shukla, G. S., and Chandra, S. V. (1987). Effects of cadmium on superoxide dismutase and lipid peroxidation in liver and kidney of growing rats: *in vivo* and *in vitro* studies. *Pharmacol. Toxicol.* 60 (5), 355–358. doi: 10.1111/j.1600-0773.1987.tb01526.x

Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., et al. (2014). Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43 (D1), D921–D927. doi: 10.1093/nar/gku955

Iwatsubo, T., Hirota, N., Ooie, T., Suzuki, H., Shimada, N., Chiba, K., et al. (1997). Prediction of *in vivo* drug metabolism in the human liver from *in vitro* metabolism data. *Pharmacol. Ther.* 73 (2), 147–171. doi: 10.1016/S0163-7258(96)00184-2

Johnson, G. E., Slob, W., Doak, S. H., Fellows, M. D., Gollapudi, B. B., Heflich, R. H., et al. (2015). New approaches to advance the use of genetic toxicology analyses for human health risk assessment. *Toxicol. Res.* 4 (3), 667–676. doi: 10.1039/C4TX00118D

Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: principles, advances, and applications. *Couns. Psychol.* 34 (5), 684–718. doi: 10.1177/0011000006286347

Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations *via* alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23 (12), 1495–1502. doi: 10.1093/bioinformatics/btm134

Klein, L., Khazaie, K., and Von Boehmer, H. (2003). In vivo dynamics of antigen-specific regulatory T cells not predicted from behavior *in vitro*. *Proc. Nat. Acad. Sci.* 100 (15), 8886–8891. doi: 10.1073/pnas.1533365100

Knight, A., Bailey, J., and Balcombe, J. (2006). Animal carcinogenicity studies: 3, Alternatives to the bioassay. *Altern. Lab. Anim.* 34 (1), 39–48. doi: 10.1177/026119290603400119

Kramer, N. I., Di Consiglio, E., Blaauboer, B. J., and Testai, E. (2015). Biokinetics in repeated-dosing *in vitro* drug toxicity studies. *Toxicol. In Vitro* 30 (1), 217–224. doi: 10.1016/j.tiv.2015.09.005

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401 (6755), 788. doi: 10.1038/44565

Lee, D. D., and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems.* 556–562.

Liu, Z., Fang, H., Borlak, J., Roberts, R., and Tong, W. (2017). In vitro to *in vivo* extrapolation for drug-induced liver injury using a pair ranking method. *ALTEX-Altern. Anim. Exp.* 34 (3), 399–408. doi: 10.14573/altex.1610201

Liu, Y., Liang, Y., Kuang, Q., Xie, F., Hao, Y., Wen, Z., et al. (2018a). Post-modified non-negative matrix factorization for deconvoluting the gene expression profiles of specific cell types from heterogeneous clinical samples based on RNA-sequencing data. *J. Chemom.* 32 (11), e2929. doi: 10.1002/cem2929

Liu, Z., Delavan, B., Roberts, R., and Tong, W. (2018b). Transcriptional responses reveal similarities between preclinical rat liver testing systems. *Front. Genet.* 9, 74. doi: 10.3389/fgene.2018.00074

Liu, Z., Huang, R., Roberts, R., and Tong, W. (2018c). Toxicogenomics: a 2020 vision. *Trends Pharmacol. Sci.* 40 (2). doi: 10.1016/j.tips.2018.12.001

Love, D., Hallbauer, D., Amos, A., and Hranova, R. (2004). Factor analysis as a tool in groundwater quality management: two southern African case studies. *Phys. Chem. Earth Parts A/B/C* 29 (15-18), 1135–1143. doi: 10.1016/j.pce.2004.09.027

Martin, S. A., McLanahan, E. D., Bushnell, P. J., Hunter Iii, E. S., and El-Masri, H. (2014). Species extrapolation of life-stage physiologically-based pharmacokinetic (PBPK) models to investigate the developmental toxicology of ethanol using *in vitro* to *in vivo* (IVIVE) methods. *Toxicol. Sci.* 143 (2), 512–535. doi: 10.1093/toxsci/kfu246

Maruyama, R., Maeda, K., Moroda, H., Kato, I., Inoue, M., Miyakawa, H., et al. (2014). Detecting cells using non-negative matrix factorization on calcium imaging data. *Neural Networks* 55, 11–19. doi: 10.1016/j.neunet.2014.03.007

Meng, J., Chen, H.-I., Zhang, J., Chen, Y., and Huang, Y. (2011). Uncover cooperative gene regulations by microRNAs and transcription factors in glioblastoma using a nonnegative hybrid factor model". *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* 6012–6015.

Niki, E. (2010). Assessment of antioxidant capacity *in vitro* and *in vivo*. *Free Rad. Biol. Med.* 49 (4), 503–515. doi: 10.1016/j.freeradbiomed.2010.04.016

Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C., and Afshari, C. A. (1999). Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinogenesis: Published In cooperation Univ. Texas MD Anderson Cancer Center* 24 (3), 153–159. doi: 10.1002/(SICI)1098-2744(199903)24:3<153::AID-MC1>3.0.CO;2-P

Paatero, P., and Tapper, U. (1994). Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5 (2), 111–126. doi: 10.1002/env.3170050203

Paatero, P. (1999). The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Graphical Stat.* 8 (4), 854–888. doi: 10.1080/10618600.1999.10474853

Pennie, W., Pettit, S. D., and Lord, P. G. (2004). Toxicogenomics in risk assessment: an overview of an HESI collaborative research program. *Environ. Health Perspect.* 112 (4), 417–419. doi: 10.1289/ehp6674

Pillay, J., den Braber, I., Vrisekoop, N., Kwast, L. M., de Boer, R. J., Borghans, J. A. M., et al. (2010). *In vivo* labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. *Blood* 116 (4), 625–627. doi: 10.1182/blood-2010-01-259028

Plate, K. H., Breier, G., Weich, H. A., and Risau, W. (1992). Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas *in vivo*. *Nature* 359 (6398), 845. doi: 10.1038/359845a0

Polit, D. F., and Beck, C. T. (2008). *Nursing research: generating and assessing evidence for nursing practice* (Philadelphia: Lippincott Williams & Wilkins).

Poma, A., and Di Giorgio, M. L. (2008). Toxicogenomics to improve comprehension of the mechanisms underlying responses of *in vitro* and *in vivo* systems to nanomaterials: a review. *Curr. Genomics* 9 (8), 571–585. doi: 10.2174/138920208786847962

Rollin, B. E. (2003). Toxicology and new social ethics for animals. *Toxicol. Pathol.* 31 (1_suppl), 128–131. doi: 10.1080/01926230390175011

Rostami-Hodjegan, A., and Tucker, G. T. (2007). Simulation and prediction of *in vivo* drug metabolism in human populations from *in vitro* data. *Nat. Rev. Drug Discovery* 6 (2), 140. doi: 10.1038/nrd2173

Russell, W. M. S., Burch, R. L., and Hume, C. W. (1959). *The principles of humane experimental technique* (London: Methuen).

Sager, J. E., Yu, J., Ragueneau-Majlessi, I., and Isoherranen, N. (2015). Physiologically based pharmacokinetic (PBPK) modeling and simulation approaches: a systematic review of published models, applications, and model verification. *Drug Metab. Dispos.* 43 (11), 1823–1837. doi: 10.1124/dmd.115.065920

Sayes, C. M., Reed, K. L., and Warheit, D. B. (2007). Assessing toxicity of fine and nanoparticles: comparing *in vitro* measurements to *in vivo* pulmonary toxicity profiles. *Toxicol. Sci.* 97 (1), 163–180. doi: 10.1093/toxsci/kfm018

Shourijeh, M. S., Flaxman, T. E., Benoit, D. L.Kinesiology. (2016). An approach for improving repeatability and reliability of non-negative matrix factorization for muscle synergy analysis. *J. Electromyogr. Kinesiol.* 26 36–43. doi: 10.1016/j.jelekin.2015.12.001

Stewart, D. W. (1981). The application and misapplication of factor analysis in marketing research. *J. Market. Res.* 18 (1), 51–62. doi: 10.1177/002224378101800105

Subbarao, C., Subbarao, N. V., and Chandu, S. N. (1996). Characterization of groundwater contamination using factor analysis. *Environ. Geol.* 28 (4), 175–180. doi: 10.1007/s002540050091

Suter, L., Babiss, L. E., and Wheeldon, E. B. (2004). Toxicogenomics in predictive toxicology in drug development. *Chem. Biol.* 11 (2), 161–171. doi: 10.1016/j.chembiol.2004.02.003

Suter, L., Schroeder, S., Meyer, K., Gautier, J.-C., Amberg, A., Wendt, M., et al. (2011). EU framework 6 project: predictive toxicology (PredTox)—overview and outcome. *Toxicol. Appl. Pharmacol.* 252 (2), 73–84. doi: 10.1016/j.taap.2010.10.008

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 (3), e1004847. doi: 10.1371/journal.pcbi.1004847

Swain, J. E., Lorberbaum, J. P., Kose, S., and Strathearn, L. (2007). Brain basis of early parent–infant interactions: psychology, physiology, and *in vivo* functional neuroimaging studies. *J. Child Psychol. Psychiatry* 48 (3–4), 262–287. doi: 10.1111/j.1469-7610.2007.01731.x

Tice, R. R., Agurell, E., Anderson, D., Burlinson, B., Hartmann, A., Kobayashi, H., et al.(2000). Single cell gel/comet assay: guidelines for *in vitro* and *in vivo* genetic toxicology testing. *Environ. Mol. Mutagen.* 35 (3), 206–221. doi: 10.1002/(SICI)1098-2280(2000)35:3<206::AID-EM8>3.0.CO;2-J

Tice, R. R., Austin, C. P., Kavlock, R. J., and Bucher, J. R. (2013). Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect.* 121 (7), 756–765. doi: 10.1289/ehp.1205784

Uehara, T., Ono, A., Maruyama, T., Kato, I., Yamada, H., Ohno, Y., et al. (2010). The Japanese toxicogenomics project: application of toxicogenomics. *Mol. Nutr. Food Res.* 54 (2), 218–227. doi: 10.1002/mnfr.200900169

Wallace, R. M., and Katz, S. M. (1964). A method for the determination of rank in the analysis of absorption spectra of multicomponent systems 1. *J. Phys. Chem.* 68 (12), 3890–3892. doi: 10.1021/j100794a511

Wallace, R. M. (1960). Analysis of absorption spectra of multicomponent systems 1. *J. Phys. Chem.* 64 (7), 899–901. doi: 10.1021/j100836a019

Wang, K., Zhang, S., Marzolf, B., Troisch, P., Brightman, A., Hu, Z., et al. (2009). Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Nat. Acad. Sci.* 106 (11), 4402–4407. doi: 10.1073/pnas.0813371106

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* 32 (9), 926. doi: 10.1038/nbt3001

Watanabe, T., Tobe, K., Nakachi, Y., Kondoh, Y., Nakajima, M., Hamada, S., et al. (2007). Differential gene expression induced by two genotoxic N-nitroso carcinogens, phenobarbital and ethanol in mouse liver examined with oligonucleotide microarray and quantitative real-time PCR. *Genes Environ.* 29 (3), 115–127. doi: 10.3123/jemsge.29.115

Waters, M. D., and Fostel, J. M. (2004). Toxicogenomics and systems toxicology: aims and prospects. *Nat. Rev. Genet.* 5 (12), 936. doi: 10.1038/nrg1493

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science* 330 (6004), 686–688. doi: 10.1126/science.1193147

Xu, J., Gong, B., Wu, L., Thakkar, S., Hong, H., and Tong, W. (2016). Comprehensive assessments of RNA-seq by the SEQC consortium: FDA-led efforts advance precision medicine. *Pharmaceutics* 8 (1), 8. doi: 10.3390/pharmaceutics8010008

Yang, X., Greenhaw, J., Shi, Q., Su, Z., Qian, F., Davis, K., et al. (2011). Identification of urinary microRNA profiles in rats that may diagnose hepatotoxicity. *Toxicol. Sci.* 125 (2), 335–344. doi: 10.1093/toxsci/kfr321

Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., and Alexander, D. C. (2012). NODDI: practical *in vivo* neurite orientation dispersion and density imaging of the human brain. *Neuroimage* 61 (4), 1000–1016. doi: 10.1016/j.neuroimage.2012.03.072