A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP11305.

HExpPredict: *In Vivo* Exposure Prediction of Human Blood Exposome Using a Random Forest Model and Its Application in Chemical Risk Prioritization

Fanrong Zhao,[1,2,3] Li Li,[4] Penghui Lin,[3] Yue Chen,[5] Shipei Xing,[6] Huili Du,[3,7] Zheng Wang,[5] Junjie Yang,[3] Tao Huan,[6] Cheng Long,[5] Limao Zhang,[3] Bin Wang,[8,9] and Mingliang Fang[1,3,10]

[1]Department of Environmental Science and Engineering, Fudan University, Shanghai, P.R. China
[2]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[3]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore
[4]School of Community Health Sciences, University of Nevada, Reno, Reno, Nevada, USA
[5]School of Computer Science and Engineering, Nanyang Technological University, Singapore
[6]Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada
[7]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, P.R. China
[8]Institute of Reproductive and Child Health, Peking University/Key Laboratory of Reproductive Health, National Health Commission of the People's Republic of China, Beijing, P.R. China
[9]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, P.R. China
[10]Institute of Eco-Chongming, Shanghai, P.R. China

Background: Due to many substances in the human exposome, there is a dearth of exposure and toxicity information available to assess potential health risks. Quantification of all trace organics in the biological fluids seems impossible and costly, regardless of the high individual exposure variability. We hypothesized that the blood concentration (C_B) of organic pollutants could be predicted via their exposure and chemical properties. Developing a prediction model on the annotation of chemicals in human blood can provide new insight into the distribution and extent of exposures to a wide range of chemicals in humans.

Objectives: Our objective was to develop a machine learning (ML) model to predict blood concentrations (C_Bs) of chemicals and prioritize chemicals of health concern.

Methods: We curated the C_Bs of compounds mostly measured at population levels and developed an ML model for chemical C_B predictions by considering chemical daily exposure (DE) and exposure pathway indicators (δ_{ij}), half-lives ($t_{1/2}$), and volume of distribution (V_d). Three ML models, including random forest (RF), artificial neural network (ANN) and support vector regression (SVR) were compared. The toxicity potential or prioritization of each chemical was represented as a bioanalytical equivalency (BEQ) and its percentage (BEQ%) estimated based on the predicted C_B and ToxCast bioactivity data. We also retrieved the top 25 most active chemicals in each assay to further observe changes in the BEQ% after the exclusion of the drugs and endogenous substances.

Results: We curated the C_Bs of 216 compounds primarily measured at population levels. RF outperformed the ANN and SVF models with the root mean square error (RMSE) of 1.66 and 2.07 μM, the mean absolute error (MAE) values of 1.28 and 1.56 μM, the mean absolute percentage error (MAPE) of 0.29 and 0.23, and R^2 of 0.80 and 0.72 across test and testing sets. Subsequently, the human C_Bs of 7,858 ToxCast chemicals were successfully predicted, ranging from 1.29×10^{-6} to 1.79×10^{-2} μM. The predicted C_Bs were then combined with ToxCast *in vitro* bioassays to prioritize the ToxCast chemicals across 12 *in vitro* assays with important toxicological end points. It is interesting that we found the most active compounds to be food additives and pesticides rather than widely monitored environmental pollutants.

Discussion: We have shown that the accurate prediction of "internal exposure" from "external exposure" is possible, and this result can be quite useful in the risk prioritization. https://doi.org/10.1289/EHP11305

Introduction

Because many chemical substances have been developed and used in commerce over numerous recent decades, there is a dearth of exposure and toxicity information available to assess potential health risks of most of these chemicals to humans.[1,2] To address concerns over the potential health effects of untested chemicals, high-throughput screening (HTS) assessments that incorporate both exposure and toxicity data are needed for risk-based screening and prioritization.[1–4] The U.S. Environmental Protection Agency (U.S. EPA) has developed the ToxCast program to provide *in vitro* bioactivity data that may inform chemical toxicity.[5,6] However, to use the *in vitro* bioactivity data of ToxCast to evaluate the potential risk to human health, chemical blood concentration (C_B) is essential to link the internal exposure to external human exposure.[7]

One challenge to chemical exposure and risk assessments has been the demand for a large number of chemical C_B measurements.[8] Clearly, experimental quantification is cumbersome and time-consuming. The standards used for analysis are also costly or difficult to obtain. In addition, the concentrations of most compounds are too low to be detectable.[9,10] Moreover, there is high variability in chemical levels between biospecimens from different people, sometimes even for samples collected from the same donors on different days in cases of exposure to rapidly metabolized chemicals.[11,12] The National Health and Nutrition Examination Survey (NHANES) has spent years monitoring several hundred chemicals, which is still insufficient for the evaluation of chemical exposure risk in the era of the exposome. Therefore, without extensive direct measurements of chemicals at the population level, there is an urgent need to explore whether we can develop *in silico* methods to predict the C_Bs of chemicals. Although the U.S. EPA has also developed the ExpoCast program to predict human exposure to the large number of chemicals with the balanced

Address correspondence to Bin Wang, Institute of Reproductive and Child Health, Peking University/National Health Commission's Key Laboratory of Reproductive Health, Beijing 100191, China. Email: binwang@pku.edu.cn. And, Mingliang Fang, Department of Environmental Science and Engineering, Fudan University, Shanghai, China, 200433. Email: mlfang@ntu.edu.sg

Supplemental Material is available online (https://doi.org/10.1289/EHP11305).

The authors declare they have no actual or potential competing financial interests.

Received 25 March 2022; Revised 15 December 2022; Accepted 14 February 2023; Published 13 March 2023.

accuracies of the source-based exposure pathway models ranging from 73% to 81% and with a coefficient of determination (R^2) between predictions and biomonitoring-based inferences of 0.8,[3] the ExpoCast can only predict the intake rates, which is an indicator of external exposure. Because different chemicals have different bioavailability and clearance, to assess health risks using ToxCast activity test data, it is necessary to convert the external exposure data into internal concentration in bodily fluids.[7] Previous efforts built quantitative approaches to translate *in vitro* toxicity potencies to equivalent *in vivo* doses using *in vitro*–*in vivo* extrapolation (IVIVE) techniques.[13] These approaches used pharmacokinetic equations to estimate steady-state plasma concentrations (C_{SS}) using the High-Throughput Toxicokinetic (HTTK) the open-source R package (version 4.2.1; R Development Core Team).[13] However, the C_B values predicted by the HTTK model were derived by assuming steady-state and 100% oral bioavailability under a dose rate of 1 mg/kg/d, which did not consider the exposure and the corresponding uncertainty; and chemicals such as perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS), which were thought to be actively resorbed by the kidney, were not captured by the current HTTK model.[13] In addition, for the recent studies, the high-throughput PROduction-To-EXposure (PROTEX-HT) model developed by Li et al. could already predict the Css without assuming 100% oral absorption,[2] and the Physiologically based Toxicokinetic (PBTK) model developed by Armitage et al. could already capture the renal clearance and reabsorption of ions such as polyfluoroalkyl substances (PFAS).[14] However, most of those theoretical methods used to predict the chemical Css resulting from repeated daily exposure were limited to oral route of exposure.[4,15,16]

We hypothesized that the C_B of organic pollutants could be predicted via their exposure and chemical properties, especially for those with similar exposure routes and physicochemical parameters. We seek to increase the prediction accuracy of C_B using machine learning (ML) methods. In this study, we curated the C_Bs of pollutants in the general population from available databases and literature and applied ML algorithms for C_B predictions by optimizing the key parameters that mediate the C_B. We compared three ML algorithms, including random forest (RF), artificial neural network (ANN) and support vector regression (SVR), based on the publicly available experimental data. The best-performing RF model was then used to predict the C_Bs of >7,500 ToxCast chemicals. The predicted C_B values were further combined with ToxCast *in vitro* bioassays to prioritize those ToxCast chemicals in terms of C_B/AC_{50} ratios, using different assay end points. This advanced human internal exposure prediction (HExpPredict) approach provides the ability to evaluate and prioritize chemicals for potential risk to human health.

Methods

A detailed data processing workflow is depicted in Figure 1. Key parameters and models regarding the models developed for this study are described in the following sections. According to the pharmacokinetics and toxicokinetics models, factors that are known or expected to influence the relationship between external exposure and the chemical C_B are the elimination half-life, bioavailability, volume of distribution (V_d), dosage, and dosing interval.[17] When defining dosing interval equal to 1 day, the maintenance dose refers to the daily exposure (DE, milligrams per kilogram body weight per day). Because of the lack of data, the parameters such as renal clearance half-life and bioavailability were treated as an unknown parameter and trained by ML model. As one of the major pathways of elimination, the predictable biotransformation half-life ($t_{1/2}$) was included in our prediction model.

Chemical Selection

The chemicals were selected based on a subset of the ToxCast Database (version 3.0, publicly released October 2018) in this study, for which the exposure data and *in vitro* bioactivity assay data were readily available.[18] The U.S. EPA's ToxCast chemical list includes more than 9,000 compounds, including industrial chemicals, pesticides, consumer product ingredients, and pharmaceuticals. The full list of chemicals considered is

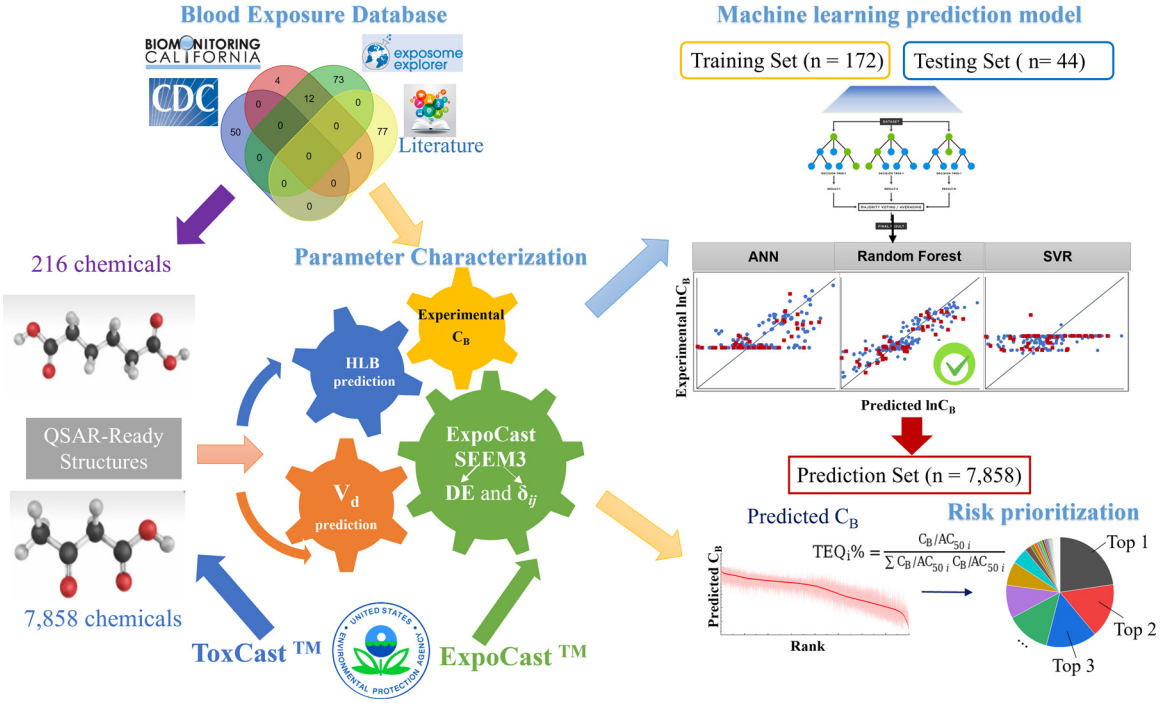

Figure 1. Overview of framework for human C_B prediction (HExpPredict) modeling and risk prioritization in this study. Note: C_B, blood concentration.

available in Excel Table S1. All chemical descriptors including CAS registry number, chemical name, Simplified Molecular Input Line Entry Specification (SMILES), molecular formula, average mass, and monoisotopic mass are available through the U.S. EPA's CompTox Chemicals Dashboard (version 2.1.1; https://comptox.epa.gov/dashboard/batch-search).[19]

Exposure Estimates

The median of estimated DE level (milligrams per kilogram body weight per day) with uncertainty [95% confidence interval (CI)] for the ToxCast chemical as shown in Excel Table S1 was acquired from the U.S. EPA's ExpoCast exposure estimates, which were developed using the General Population Consensus Model (SEEM3).[3,20] The exposure pathway indicators (δ_{ij}) for four source-based pathways (far-field pesticide use, nonpesticide dietary exposure, far-field industrial exposure, and consumer) in the SEEM3 model were also included in our prediction model.[3] The δ_{ij} is an estimated probability of whether a given pathway j is relevant to a given chemical i.

Chemical Biotransformation Half-Life Prediction

The predicted half-life values ($t_{1/2}$) for the ToxCast chemicals were taken from the Human Exposome and Metabolite Database (HExpMetDB).[21] The prediction was based on the quantitative structure−activity relationship (QSAR) approach called Iterative Fragment Selection (IFS).[22]

The Distribution Volume (V_d) Prediction

The V_d values were predicted by a comprehensive exposure model named Risk Assessment, IDentification And Ranking-Indoor and Consumer Exposure (RAIDAR-ICE) according to previous study.[23]

Molecular Descriptors and QSAR Parameter Calculation

The QSAR parameters such as Log K_{OW} and Log K_{OA} were calculated using solute descriptors provided by the online UFZ-LSER Database.[24] Water solubility (WS) and substructure molecular descriptors were calculated by the Toxicity Estimation Software Tool (TEST, version 5.1.1).[25]

Chemical C_B Search

To investigate the occurrence and levels of xenobiotics in human blood, we conducted a database and literature search on chemicals in human blood. The measured C_Bs of xenobiotics in this study were first retrieved from the NHANES 2003–2017,[26] the California Environmental Contaminant Biomonitoring Program (also known as Biomonitoring California),[27] or the Exposome-Explorer.[28] We excluded drugs and endogenous compounds by filtering the U.S. EPA's CompTox Chemicals Dashboard Drugbank list (https://comptox.epa.gov/dashboard/chemical-lists/DRUGBANK) and manually searching the chemical category through PubChem (https://pubchem.ncbi.nlm.nih.gov/). When a given chemical was present in both of these databases, we used the NHANES concentrations. To further obtain concentration data for more compounds, we performed a literature search on typical pollutants that were not in the databases, based on the chemicals of concerns previously summarized in our research.[29] The National Center for Biotechnology Information (NCBI) PubMed database (https://pubmed.ncbi.nlm.nih.gov/) was searched from the year 2005 to 2022. The keywords used to search the PubMed database included those describing sample types "blood," "plasma," or "serum" and terms for the typical pollutant classes summarized in our previous study,[29] including "perfluorinated compounds," "volatile organic compound," "pesticide," "organophosphorus flame retardant," or "polycyclic aromatic hydrocarbons," together with keywords including "exposome," "exposure," "detection," "level" or "concentration." We included only the studies from healthy human populations using a mass spectrometry–based analytical method during our manual screening of the possible literature hits. We also excluded the studies from polluted areas or special environment areas. The C_B of each compound was calculated based on the sample size weighted geometric mean (GM, if provided) or median concentrations measured in serum, plasma, or blood. To develop models for different age and sex groups, we also collected the GMs of C_Bs for different age and sex groups from the NHANES Database ($n = 48$).

ML Models

Methods of random search and 5-fold cross-validation were used for parameter optimization to train three ML models (i.e., RF, ANN, and SVR) with various prediction features of DE, δ_{ij}, V_d, $t_{1/2}$, and other chemical properties, of which the optimal parameters was evaluated by and root mean square error (RMSE). The publicly available data sets Exposome-Explorer database,[28] the Fourth National Report on Human Exposure to Environmental Chemicals,[26] and the California Environmental Contaminant Biomonitoring Program[27] were searched for experimentally measured human *in vivo* C_B values. Literature mining was performed by manually searching reviews or articles as mentioned above. The measured C_Bs were employed to train ML models for *in silico* C_B prediction. We excluded the drug and endogenous compounds by filtering the U.S. EPA's CompTox Chemicals Dashboard Drugbank list and manually searching the chemical category through PubChem (https://pubchem.ncbi.nlm.nih.gov/), because our model only considers the C_Bs produced by external exposures. For the collected experimental C_Bs and predicted $t_{1/2}$, V_d, and DE values, we unified their units into micromolar, day, L, and micromole per day, respectively, and normalized the right-skewed data by natural logarithmic transformation before feeding them to a ML model. The training and testing splits were 80:20 to train and test RF, ANN, and SVR models. Training and testing set chemicals were randomly selected. In this work, the RMSE, mean absolute error (MAE), mean absolute percentage error (MAPE), and fitness degree R^2 of the three models were compared. Finally, the trained model was used to predict C_B for the ToxCast compounds. All analyses were performed in R (version 4.2.1; R Development Core Team). All chemical predictors are provided in Excel Table S1. To improve the applicability of our model, the R script and tutorial for users are also available in the Supplemental File HExpPredict_scripts.rar and Supplemental Material, "Text S1," as well as at https://github.com/FangLabNTU/HExpPredict.

Monte Carlo (MC) Simulation and Parameter Distributions

MC simulation was implemented to simulate the impact of DE and $t_{1/2}$ uncertainty on calculating the C_B 10,000 times, using a similar model as in our previous studies.[21,30,31] Three separate MC simulations were performed referring to previous studies: DE prediction uncertainty only, $t_{1/2}$ prediction uncertainty only, and both DE and $t_{1/2}$ prediction uncertainty.[20,21] For each chemical, the C_B was calculated 10,000 times for three separate MC simulations respectively, allowing estimation of the 5th, median, and 95th percentiles.

In Vitro *Bioactivity Data*

All ToxCast *in vitro* HTS data (version 3.0, publicly released October 2018)[18] were downloaded from the U.S. EPA's CompTox Chemicals Dashboard (version 2.1) Assay Endpoints

List (https://comptox.epa.gov/dashboard/assay-endpoints?filtered) to estimate the endocrine-related activity. The 12 targeted assays covering the estrogen receptor alpha (ERα) (TOX21_Era_BLA_Agonist_ratio and TOX21_Era_BLA_Antagonist_ratio), androgen receptor (AR) (TOX21_AR_BLA_Agonist_ratio, Tox21_AR_LUC_MDAKB2_Agonist, TOX21_AR_BLA_Antagonist_ratio and TOX21_AR_LUC_MDAKB2_Antagonist_0.5nM_R1881), peroxisome proliferator–activated receptor gamma (PPARγ) (Tox21_PPARg_BLA_Agonist_ratio, TOX21_PPARg_BLA_Agonist_ch2, TOX21_PPARg_BLA_Antagonist_ch1 and TOX21_PPARg_BLA_antagonist_viability), and thyroid hormone receptor (TR) (TOX21_TR_LUC_GH3_Agonist and TOX21_TR_LUC_GH3_Antagonist) were chosen for further study. The bioactivity potential or prioritization of each chemical was represented as C_B-to-AC_{50} ratio (C_B/AC_{50}). We used the concentration at 50% of maximum activity (AC_{50}) estimates from the U.S. EPA's CompTox Chemicals Dashboard (version 2.1) ToxCast Assay Endpoints List[32] provided by the ToxCast program[18] as well as the predicted C_B to calculate the C_B/AC_{50} ratios of ToxCast chemicals.

The relative ranking of C_B/AC_{50} can be used for priority setting; that is, higher C_B/AC_{50} can be considered to be a higher priority. The toxicity potential or prioritization of each chemical was represented as a bioanalytical equivalency (BEQ). The BEQ values of each chemical and its percentage in the total BEQ (BEQ%) were estimated based on the below equations[29]:

$$BEQ_i = C_{B\,i}/AC_{50\,i} \times AC_{50\,ref} \quad (1)$$

$$BEQ_i\% = BEQ_i \div \sum BEQ_i \times 100\%, \quad (2)$$

where $C_{B\,i}$ is the predicted blood concentration of compound i; $AC_{50\,i}$ is the concentration of compound i that causes 50% response; and $AC_{50\,ref}$ is the concentration of the reference compound (the compound with the minimum AC_{50} for each assay) that causes 50% response. We further retrieved the applications of the top 25 most active chemicals of each assay from the NCBI PubMed databases (https://pubmed.ncbi.nlm.nih.gov).

Results

A total of 7,858 chemicals were selected in this study from 9,403-chemical U.S. EPA's ToxCast Database.[18] The chemicals that were not selected (1,545) comprised those that did not have available DE data through ExpoCast SEEM3 and those that were categorized as ionogenic chemicals, organic mixtures, or chemicals with molecular weights over 1,000 Da and therefore unable to be used by the iterated function system (IFS) algorithm.

Chemical C_B Search

To investigate the occurrence and levels of the selected chemicals in human blood, we conducted a database[26–28] and literature search,[10,33–37] extracting C_B from identified data and studies. In total, the measured C_Bs of 216 chemicals were documented for the further ML modeling. In general, the C_Bs of the documented chemicals ranged from 1.65×10^{-8} to 1.59 μM staggering 8 orders of magnitude. The final list is presented in Excel Table S2, including CAS registry number, chemical name, formula, average mass, monoisotopic mass, weighted C_B, and data sources for our RF model. Overall, the NHANES, the Exposome-Explorer Database, and the literature search were the dominant contributors to the training set and contributed to 23%, 39%, and 36% of the data set, respectively. Due to limited measured C_B data of the population, we collected data from only 48 chemicals for which the age- and sex-specific geometric means of measured C_B was available from NHANES Database. The GM C_B ranges for individuals age 12–19 y and those older than 20 y were 4.65×10^{-6}–5.32×10^{-3} μM and 1.11×10^{-5}–9.00×10^{-3} μM, respectively. The GM C_B ranges were 8.31×10^{-6}–1.07×10^{-2} μM for males and 8.31×10^{-6} –6.84×10^{-3} μM for females (Excel Table S3).

Human Exposure Evaluation

The predicted exposure values of 7,858 chemicals were obtained from ExpoCast (Excel Table S1). The estimated human chemical DE ranged from 3.17×10^{-15} (95% CI: 3.82×10^{-17}, 4.19×10^{-13}) to 4.92 (95% CI: 1.65×10^{-7}, 2.21×10^{5}) mg/kg body weight/d, spanning across 15 orders of magnitude. The δ_{ij} values of 7,858 chemicals ranged from 0 to 1 for the four pathways (Excel Table S1; i.e., far-field pesticide use, nonpesticide dietary exposure, far-field industrial exposure, and consumer), with values near zero indicating low probability and values near one indicating high probability exposure to the chemical.

Chemical $t_{1/2}$ Evaluation

The $t_{1/2}$ of 7,858 chemicals listed in Excel Table S1 were successfully predicted using the IFS approach. Of these 7,858 chemicals, the median $t_{1/2}$ was predicted to be 4.64 h (h). Rolitetracycline was predicted to have the shortest $t_{1/2}$ of 0.05 h, and mirex was predicted to have the longest $t_{1/2}$ of 2,020,000 h with a wide range of 8 orders of magnitude.

Chemical V_d Prediction

We used the RAIDAR-ICE model to predict the V_d values of 7,858 chemicals (Excel Table S1). The median V_d was predicted to be 14.4 L/kg whole blood. The V_ds span over 3 orders of magnitude, from 7.36×10^{-1} to 20.3 L/kg whole blood.

C_B Prediction ML Modeling

We developed a workflow to use experimental C_B data to train and test ML models (Figure 1). Such models were then applied to the 7,858 chemicals from U.S. EPA ToxCast Program for which *in vitro* bioactivity data were available. We collected available experimentally measured human C_B values through publicly available databases and literature to train ML models for *in silico* C_B prediction. After excluding the drug and endogenous compounds, a total of 216 experimental C_B data points were included in the ML model (Figure 2A). We randomly divided the 216 data points into 172 compounds for training and 44 compounds for further testing (i.e., 80%:20%). We downloaded the chemical QSAR-ready SMILES from the U.S. EPA's CompTox Chemicals Dashboard Batch Search (version 2.1.1),[38] which we used to predict the V_d, and $t_{1/2}$. Chemical-specific inputs to ML models included DE, V_d, $t_{1/2}$, and δ_{ij} for parameter tuning.

Model Validation

To optimize the C_B prediction model performance by training set, tuning parameters including maximum depth (5–100), mtry ratio (0.2–0.8), number of trees (10–500), maximum tuning times (20), and tuning method ("random_search") were executed using the learner "ranger" of "mlr3" learning platform (https://github.com/mlr-org/mlr3). The RMSE, MAE, MAPE, and R^2 were calculated to compare the predicted and experimental C_B in the test data set. We investigated three widely used ML models (RF, ANN, and SVR) for C_B predictions with seven basic variables, including DE, V_d, $t_{1/2}$, and four δ_{ij}s. RF outperformed the other two models with RMSE values of 1.66 and 2.07 μM, MAE of 1.28 and 1.56 μM, MAPE of 0.29 and 0.23, and R^2 of 0.80 and 0.72 across training and testing predictions of C_B, respectively

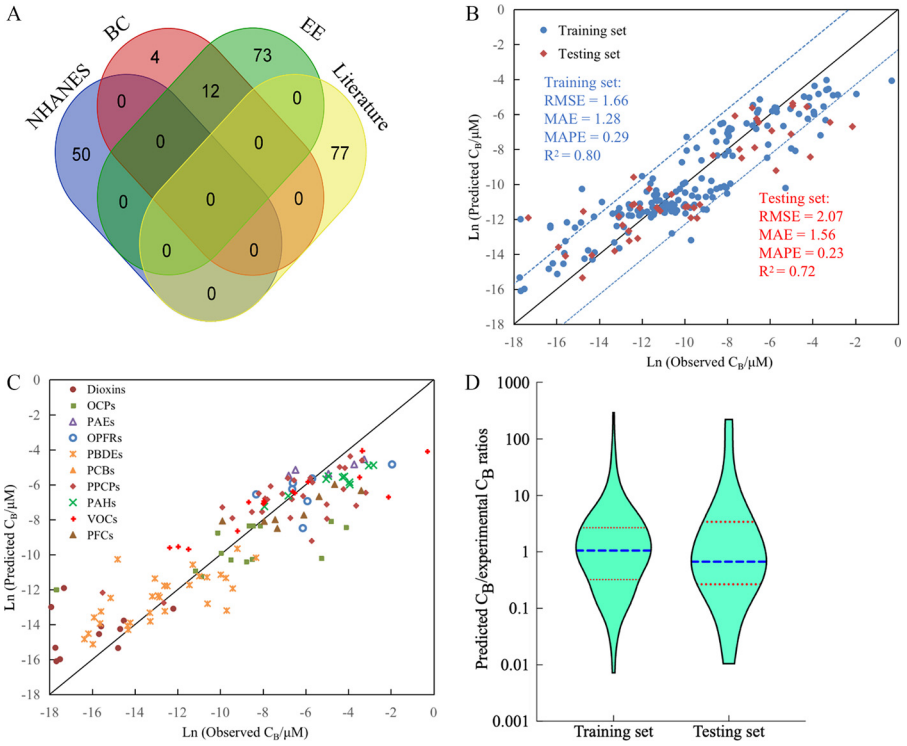

Figure 2. (A) Overlapping analysis of major sources for measured C_B used in machine learning training. (B) Prediction performance of RF ML model for training ($n = 172$) and testing ($n = 44$) sets (referring to the data in Excel Table S4); Black line is the $y = x$ line, and blue dotted lines are 10-fold boundaries; (C) Prediction performance of RF ML model for different groups of chemicals (referring to the data in Excel Table S2); Black line is the $y = x$ line. (D) Violin plots for training and testing set prediction errors by calculating the ratio between measured and predicted concentration from RF ML model (referring to the data in Excel Table S4). Blue dashed lines are the median line, and red dotted lines are quartiles. Note: BC, Biomonitoring California; C_B, blood concentration; EE, Exposome-Explorer; ML, machine learning; NHANES, National Health and Nutrition Examination Survey; OPFRs, organophosphorus flame retardants; OP, organochlorine pesticide; PAE, phthalate ester; PBDE; polybrominated diphenyl ether; PCB, polychlorinated biphenyl; PFC, perfluorinated compounds; PPCP, personal care and consumer product; RF, random forest; VOC, volatile organic compound.

(Table 1). In comparison, ANN and SVR showed less robustness, with RMSE values of 2.83 and 3.07 μM, MAE of 2.13 and 2.56 μM, MAPE of 0.39 and 0.35, and R^2 of 0.41 and 0.39 for ANN, and with RMSE values of 3.52 and 3.79 μM, MAE of 2.81 and 3.22 μM, MAPE of 0.69 and 0.47, and R^2 of 0.08 and 0.06 for SVR across training and testing sets (Table 1), respectively. Approximately 90% (165 of 174) and 84% (37 of 44) predicted C_B values of training and testing sets showed to be within the 10-fold boundary when compared with measured C_B values (Figure 2B), showing much better regression than ANN and SVR models (Figure S1, referring to the data in Excel Table S4). To further optimize the RF model, we considered adding sex, age, and variables of varying complexity including Log K_{OW}, Log K_{OA}, WS, and additional molecular descriptors to our RF model. However, the prediction performance was not dramatically improved when more parameters were included into the RF model. Detailed results were provided in the Supplemental Material, "Text S2."

Good prediction performance of the RF model were observed for some typical environmental pollutants, such as polychlorinated biphenyls (PCBs), dioxins, phthalate esters (PAEs), dioxins, polycyclic aromatic hydrocarbons (PAHs), perfluorinated compounds (PFCs), organophosphorus flame retardants (OPFRs), and volatile organic compounds (VOCs) (Figure 2C), with the RMSE of 0.64, 0.70, 0.71, 0.73, 0.83, 0.85, and 0.86, respectively (Table S1). In contrast, some substances, like personal care and consumer products (PPCPs) and organochlorine pesticides (OPs), showed relatively poor prediction performance, with the RMSE of 1.18 and 1.68, respectively. The RF model covered 50% compounds within 0.32 to 2.6 and 0.24 to 3.4 times of predicted C_B/experimental C_B ratios for training set and testing set, respectively (Figure 2D).

Using the final RF model, C_Bs were determined for each of the 7,858 ToxCast chemicals. In general, the predicted human blood C_B of 7,858 ToxCast chemicals ranged from 1.02×10^{-6} to 3.25×10^{-2} μM (Excel Table S1), ranging four orders of magnitude (Figure 3).

Table 1. The prediction performance of three C_B prediction machine learning models.

	Training set ($n = 172$)				Testing set ($n = 44$)			
Model	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
Random forest	1.66	1.28	0.29	0.80	2.07	1.56	0.23	0.72
Artificial neural network	2.83	2.13	0.39	0.41	3.07	2.56	0.35	0.39
Support vector regression	3.52	2.81	0.69	0.08	3.79	3.22	0.47	0.06

Note: C_B, blood concentration; MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean square error.

Uncertainty Analysis

Three MC simulations (DE prediction uncertainty alone, $t_{1/2}$ prediction uncertainty alone, and both) were performed to determine the predicted C_B upper 95th percentile. The ratio of the C_B for the 95th percentile to the median indicates the relative contribution uncertainty, with larger ratios indicating greater uncertainty. We observed that the ratio value of median $t_{1/2}$ prediction uncertainty (1.17) was close to DE prediction uncertainty (1.28). The ratio value of both uncertainty (2.17) was close to the sum of $t_{1/2}$ and DE, which indicated that the prediction of $t_{1/2}$ and DE contributed approximately the same degree of uncertainty to the prediction model.

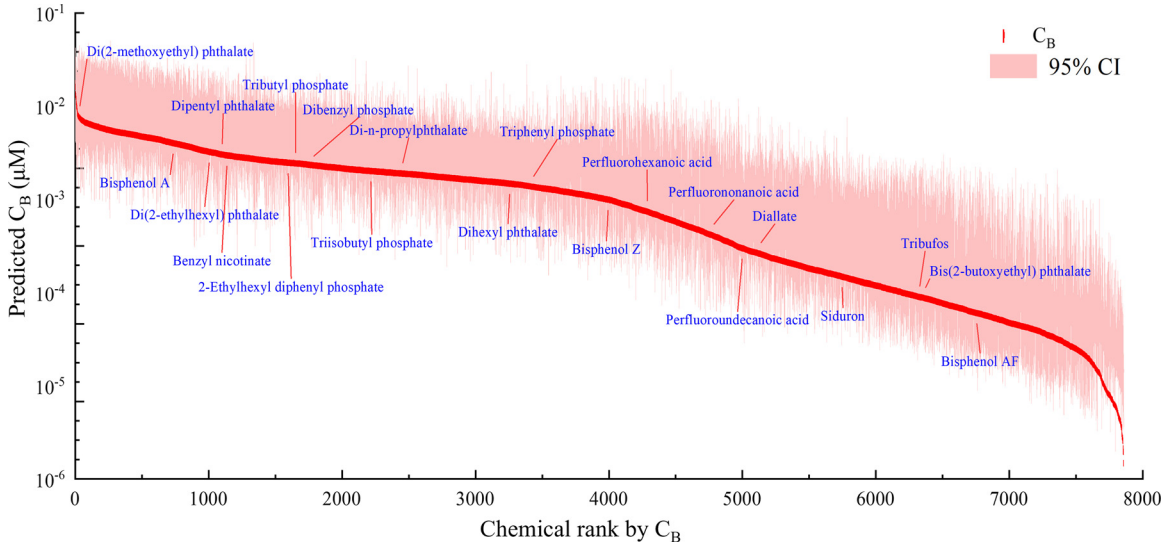

Figure 3. The cumulative distribution of chemical predicted C_B using RF model ($n = 7,858$). The bar indicates the median predicted C_B for each chemical; the pink area represents the predicted C_B range (5%–95%) derived from the Monte Carlo simulations. Some typical environmental pollutants are labeled. Note: C_B, blood concentration; RF, random forest.

Chemical Prioritization Using the U.S. EPA's ToxCast Database

We evaluated bioactivity potential for each chemical across 12 *in vitro* assays from ToxCast using AC_{50}/C_B ratios calculated as ToxCast AC_{50}/C_B ratios. The 12 ToxCast *in vitro* HT screening assays,[18] including the targets of ERα, AR, PPAR-γ and TR, were chosen as case studies. The total 12 assays covered two AR agonists, two AR antagonists, one ERα agonist, one ERα antagonist, two PPARγ agonists, two PPARγ antagonists, one TR agonist, and two TR antagonist assays (Excel Table S5). The C_B/AC_{50} ratios across all the 12 assays are listed in Excel Table S6, and the distribution (BEQ%) of each target assay result is shown in Excel Table S7. We found that each end point had obviously different chemical toxicity prioritization and had its own dominant contributor(s). For different assays of the same receptor toxicity end point, the results varied widely due to the distinct compounds tested by the different assays. It was interesting to find that drugs or endogenous chemicals were dominant contributors with the top C_B/AC_{50} ratios for most assays. For example, salidroside and *N*-vinyl-2-pyrrolidone were the most dominant contributors for Tox21_AR_LUC_MDAKB2_Agonist and TOX21_AR_LUC_MDAKB2_Antagonist_0.5 nM_R1881 assays, respectively, with the high C_B/AC_{50} ratios of 2,288, and BEQ% of 24.0% and 37.0%, respectively (Excel Table S7; Figure S2). Salidroside is a major component of *Rhodiola rosea*, which has been used in traditional Chinese medicine[39] and *N*-vinyl-2-pyrrolidone is used for treatment of infectious conjunctivitis.[40] For the TOX21_PPARg_BLA_antagonist_viability assay, the top contributor was ribavirin (C_B/AC_{50}: 327, BEQ: 79.1%), followed by ramipril (36.6, 8.84%) and diphenoxylate hydrochloride (31.0, 7.49%). Drugs like acipimox, 5-methyl-1-phenyl-2(1H)-pyridone, piconol, triacetin, and hexylcaine hydrochloride were the most abundant chemicals, accounting for 9.17%, 19.9%, 11.5%, 10.7%, and 4.05% in TOX21_AR_BLA_Agonist_ratio, TOX21_ERa_BLA_Agonist_ratio, TOX21_ERa_BLA_Antagonist_ratio, TOX21_PPARg_BLA_Agonist_ch2, and TOX21_TR_LUC_GH3_Agonist assay, respectively.

Because the predicted C_B in this study was based only on the internal C_B generated by the external exposure, we further excluded endogenous chemicals and drugs, and performed the analysis on the remaining 4,893 chemicals. After excluding endogenous chemicals and drugs, methyl formate, di(2-methoxyethyl) phthalate, propylammonium nitrate, 2,3-butanedione, and (3,5-dimethyl-1H-pyrazol-1-yl)methanol became the most dominant chemicals in TOX21_AR_BLA_Antagonist_ratio, TOX21_AR_LUC_MDAKB2_Antagonist_0.5nM_R1881, TOX21_PPARg_BLA_Antagonist_ch1, TOX21_PPARg_BLA_antagonist_viability, and TOX21_TR_LUC_GH3_Antagonist assay, with the BEQ% of 22.1%, 23.8%, 51.7%, 61.4%, and 46.3%, respectively (Excel Table S8). 2-Acetylpyrrole, thiamine thiozole, and aminopyridine a showed the highest C_B/AC_{50} ratios of 549, 494, and 401, respectively (Excel Table S6), which was the dominant contributor with the BEQ% of 10.3%, 9.31%, and 7.56%, for the TOX21_AR_BLA_Agonist_ratio assay (Excel Table S8), suggesting that they had a relatively high potential risk of androgen disruption. In the Tox21_AR_LUC_MDAKB2_Agonist, the largest contributions were 3,3′-(ethylenedioxy)dipropiononitrile (C_B/AC_{50} ratio: 275, BEQ%: 10.3%), 1,3-dichloropropanone (226, 8.42%), and MCPB (214, 7.97%). The dominant contributor of the TOX21_AR_BLA_Antagonist_ratio assay was methyl formate (1218, 22.1%), followed by 1-bromoheptadecane (644, 11.7%), and 1,2-dimethylhydrazine dihydrochloride (488, 8.83%). In the TOX21_AR_LUC_MDAKB2_Antagonist_0.5nM_R1881 assay, di(2-methoxyethyl) phthalate (7.71, 23.8%), FD&C yellow 5 (7.50, 23.1%), and acetone (7.43, 22.9%) contributed the most.

In the TOX21_ERa_BLA_Agonist_ratio assay, the major contributions came from 1,1':4',1"-terphenyl (538, 22.8%), sodium nicotinate (379, 16.0%), and sodium 2,5-dimethylbenzenesulfonate (361, 15.3%). In the TOX21_ERa_BLA_Antagonist_ratio assay, 2-bromo-1-ethanol (550, 13.6%), benzyl nicotinate (379, 9.32%), and ethyl bromoacetate (224, 5.51%) were the dominant contributors. The dominant contributors were 1-bromopentadecane (328, 24.1%), beta-nitrostyrene (221, 16.2%), and (6Z)-non6-en-1-ol (216, 15.9%) for the Tox21_PPARg_BLA_Agonist_ratio assay; triacetin (1132, 18.2%), succinic anhydride (841, 17.2%), 2-(2-aminoethoxy)ethanol (680, 13.9%), and 2-pyrrolidinone (510, 10.5%) for the TOX21_PPARg_BLA_Agonist_ch2 assay; propylammonium nitrate (599, 51.7%), citronellol (246, 21.2%), geranyl formate (188, 16.3%), and isopentyl benzoate (39.5, 3.41%) for the TOX21_PPARg_BLA_Antagonist_ch1assay; and 2,3-butanedione (7.47, 61.4%), 3-acetyldihydro-2(3H)-furanone (2.61, 21.4%) and

3-mercaptopropyltrimethoxysilane (0.75, 6.15%) for the TOX21_PPARg_BLA_antagonist_viability assay.

The top contributions of the TOX21_TR_LUC_GH3_Agonist and assay were (4-methoxyphenyl)methanol (549, 6.20%), 2-butene-1,4-diol (528, 5.96%), 2,3-butanedione (523, 5.90%), and ethyl phthalyl ethyl glycolate (487, 5.50%). In the TOX21_TR_LUC_GH3_Antagonist assay, the dominant contributors were (3,5-dimethyl-1H-pyrazol-1-yl)methanol (638, 46.5%), phenethyl anthranilate (183, 13.4%), dimethyl isophthalate (89.5, 10.3%), and sodium 2-mercaptobenzothiolate (73.2, 6.91%).

We further retrieved the applications of the top 25 chemicals of each assay from the NCBI PubMed databases (https://pubmed.ncbi.nlm.nih.gov) (Excel Table S8), and we recalculated their BEQ% values after excluding drugs and endogenous substances: Food additives such as 2,3-butanedione, methyl formate, and FD&C Yellow 5 are used as flavoring agents or colorants, with the BEQ% values of 61.4%, 22.1%, and 23.1% in TOX21_PPARg_BLA_antagonist_viability, TOX21_AR_BLA_Antagonist_ratio, and TOX21_AR_LUC_MDAKB2_Antagonist_0.5nM_R1881 assay, respectively. Plasticizers such as dimethyl isophthalate (6.50%), diisobutyl phthalate (4.51%), and diethyl phthalate (4.16%), which are defined as U.S. Food and Drug Administration indirect additives used in food-contact substances, also showed significant activity after excluding drugs and endogenous substances in TOX21_TR_LUC_GH3_Antagonist, TOX21_PPARg_BLA_Agonist_ch2, and TOX21_AR_BLA_Antagonist_ratio assays, respectively. Chemicals such as propylammonium nitrate (51.7%) and (3,5-dimethyl-1H-pyrazol-1-yl)methanol (46.3%), used for solvents and cosmetic products, were the top contributors in TOX21_PPARg_BLA_Antagonist_ch1 and TOX21_TR_LUC_GH3_Antagonist assay, respectively.

Discussion

The framework described in this study provides several implications for HT chemical screening and prioritization. We used an HT machine learning algorithm for C_B predictions with key parameters, including DE, δ_{ij}, V_d, and $t_{1/2}$. This HT HExpPredict approach can rapidly relate environmental chemical exposures to *in vitro* bioactivity, helping drive priorities based on risk potential.

Based on direct comparison of RMSE, MAE, MAPE, and R^2 between models, we concluded that the RF model showed better performance than the other models. The ML model developed in this study was based on the physical and chemical properties and exposure of the chemicals. The input data of the ML models only included the key parameters DE, δ_{ij}, V_d, and $t_{1/2}$, and we used the ML models to combine these variables to make predictions without the other parameters, such as bioavailability and plasma protein binding data. We noted that only 10.3% and 15.9% of our evaluation chemicals were predicted to be over the 10-fold boundary for the RF training and testing sets, respectively, showing good predictive ability. To build this ML model, some well-performed predictive models including the IFS approach and SEEM3 were applied. Although these models were evaluated and tested, it is important to note that these prediction models can continue to be improved with the generation of more data, which could also improve our present ML model in the future.

Uncertainty in predicting C_B can be accounted for in risk prioritization if the degree of uncertainty can be predicted for each chemical. According to the results of the three MC simulations, the prediction uncertainties of $t_{1/2}$ and DE contributed approximate uncertainty to the ML prediction model. However, the uncertainty of the ML model was underestimated because of the lack of the V_d uncertainty. Although $t_{1/2}$ and DE contributed approximately the same degree of uncertainty, some chemicals out of the model's applicability domain, such as chemicals that contain silicon, were observed to have large standard errors in the prediction, which leads to high uncertainties for the $t_{1/2}$.

The C_B of phthalates such as di(2-methoxyethyl) phthalate, dipentyl phthalate, dipropyl phthalate, dihexyl phthalate, and bis(2-butoxyethyl) phthalate were predicted to be 7.86×10^{-3} (2.75×10^{-3}–1.99×10^{-2}), 3.08×10^{-3} (7.31×10^{-4}–5.55×10^{-3}), 1.93×10^{-3} (4.71×10^{-4}–3.01×10^{-3}), 1.50×10^{-3} (2.95×10^{-4}–3.01×10^{-3}) and 9.07×10^{-5} (2.97×10^{-5}–3.18×10^{-4}) μM, respectively. A phthalate metabolite such as monobutyl phthalate was predicted to be with the C_B of 2.12×10^{-3} (6.64×10^{-4}–3.82×10^{-3}) μM [i.e., 0.47 (0.15–0.85) ng/mL], which was consistent with the concentration of 0.5 ng/mL observed in the previous study.[41] Because the exposure of phthalates is usually characterized by monitoring the concentrations of their metabolites in the urine, our model can HT predict the C_B of these easily metabolized substances, which is convenient for subsequent HT prioritization of their toxicity and risk. The C_Bs of bisphenol A (BPA) alternatives, such as bisphenol AF (BPAF), were predicted to be 0.020 (0.019–0.021) ng/mL, which was similar to the GM concentration of 0.01 ng/mL determined in the previous study.[42] Perfluorinated compounds such as perfluorononanoic acid (PFNA) and perfluoroundecanoic acid (PFUnA) were predicted to have the C_Bs of 0.21 (0.19–0.52) and 0.17 (0.13–0.27) ng/mL, respectively, which were within the GM concentration ranges of 0.11–1.88 and 0.07–1.38 ng/mL, respectively, as observed in the general populations in 13 Chinese cities.[43] However, for perfluorohexanoic acid (PFHxA), the predicted C_B value (0.24; 95% CI: 0.21–0.26 ng/mL) was a little bit higher than the GM concentration range of 0.02–0.21 ng/mL of the 13 Chinese cities' general populations.[43] The estimated concentration can be very useful in the exposure or toxicity prioritization or even the mixture effect of blood exposome.[31,44,45] In this study, the potential health effects and the causal compounds of ToxCast were summarized, revealing several key biomarker assays. A total of 12 ToxCast assays were used to assess the health effects of 4,893 chemicals, which showed different risk-based prioritization patterns. In addition to the top risk substances listed in the "Results" section, we found it interesting that some typical AR agonists, such as 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin with the C_B/AC_{50} ratio of 0.12 for Tox21_AR_LUC_MDAKB2_Agonist assay, also showed relative higher (97th of 4,893 chemicals) AR agonist activity owing to its extremely low AC_{50} (6.45×10^{-5} μM). In contrast, due to its low C_B (7.91×10^{-6} μM), the BEQ% was only 0.0045%. Nonylparaben showed relatively strong AR antagonist activity, with the C_B/AC_{50} ratio of 2.49 and BEQ% of 0.05% in the TOX21_AR_BLA_Agonist_ratio assay, and diethyl phthalate showed very strong AR antagonist activity, with the C_B/AC_{50} ratio of 230 and BEQ% of 3.59% in the TOX21_AR_BLA_Antagonist_ratio assay.

Due to the very low AC_{50} values, some pesticides such as siduron and tribufos were observed to have relatively strong ER agonist activity in the TOX21_ERa_BLA_Agonist_ratio assay, with the C_B/AC_{50} ratios of 15.1 and 5.92, and BEQ% of 30.49% and 0.19%, respectively, and benzyl nicotinate (379, 9.26%) and diallate (19.5, 0.48%) were found to have strong antagonist activity in the TOX21_ERa_BLA_Antagonist_ratio assay, with the C_B/AC_{50} ratios of 379 and 19.5 and BEQ% of 9.26% and 0.48%, respectively. In the ER agonist and antagonist assays, phthalates, BPA, and BPA alternatives were negligible due to their relatively higher AC_{50}. For example, the C_B/AC_{50} ratios of BPA in the TOX21_ERa_BLA_Agonist_ratio assay and di(2-ethylhexyl) phthalate (DEHP) in the TOX21_ERa_BLA_Antagonist_ratio assay were only 4.22×10^{-3} and 5.07×10^{-4}, respectively, due to their higher AC_{50} of 0.96 and 6.46 μM, respectively, although they had relative high C_B values of 4.06×10^{-3} and 3.27×10^{-3} μM,

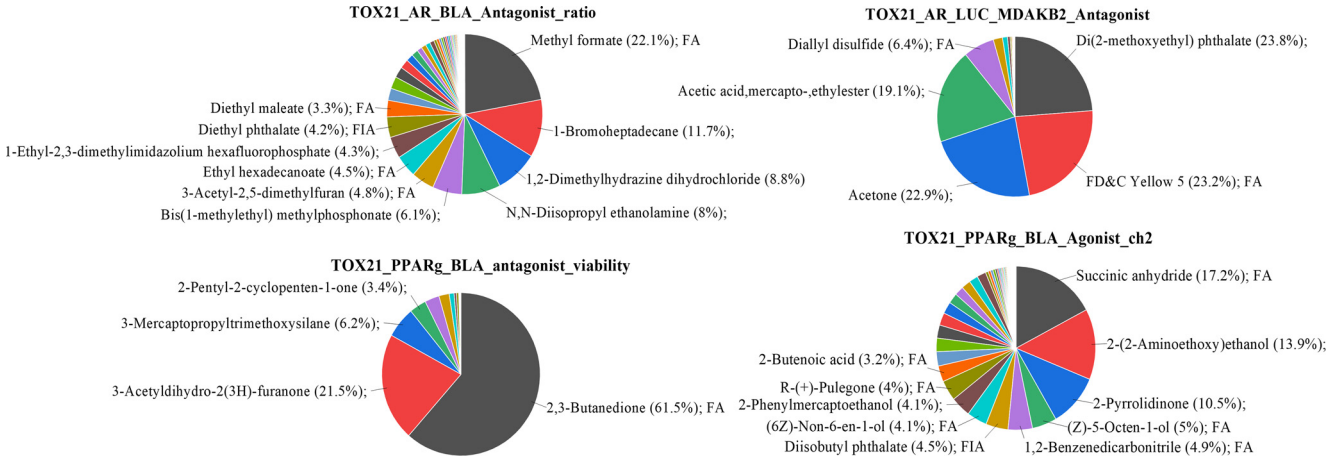

Figure 4. Toxicity contributions (percentage) of ToxCast chemicals (excluding the drugs and endogenous compounds) in assays of AR and PPARγ as examples (referring to the data in Excel Table S8). Note: AR, androgen receptor; FA, food additive; FIA, food indirect additive; PPARγ, peroxisome proliferator–activated receptor.

respectively. For the organophosphate compounds, the C_B/AC_{50} ratios of dibenzyl phosphate and triisobutyl phosphate were 202 and 13.6, respectively, in the TOX21_TR_LUC_GH3_Agonist assay, showing strong TR agonist activity. Triisobutyl phosphate also showed strong PPAR agonist activity, with the C_B/AC_{50} ratio of 41.2 in the Tox21_PPARg_BLA_Agonist_ratio assay.

An interesting finding was that when drugs and endogenous substances were excluded, food additives were the major contributors of BEQ% to the majority of assays (Figure 4; Figure S3) due to high predicted exposure by SEEM3. Food additives such as 2,3-butanedione, methyl formate, FD&C Yellow 5, and succinic anhydride showed a high potential receptor activity in AR or PPARγ (Figure 4). However, these substances are not typical pollutants, and there are little data on their biomonitoring in humans, raising concerns about their potential health risks. U.S. FDA indirect additives used in food-packing materials, such as dimethyl isophthalate, diisobutyl phthalate, and diethyl phthalate, also showed modest potential receptor activity in TR, PPARγ, and AR. The health risk of food additives and indirect food additives should be studied further. It should be noted that, besides nuclear receptors, adverse outcome pathways (AOP) (https://aopkb.oecd.org) with more toxicological end points could be further considered in future risk prioritization.

This study has several limitations. First, we could not predict the $t_{1/2}$ of chemicals with a metal atom or molecular weight over 1,000 using the IFS approach. In addition, some chemicals had extreme properties that were out of the model's applicability domain, such as silicon-containing chemicals. These chemicals were observed with large standard errors higher than the predicted mean. As far as we are concerned, no computational model can handle silicon-containing molecules at this point. Second, the ExpoCast database was unable to cover all the ToxCast compounds, and ExpoCast merely represents the exposure of typical Americans for their historical exposure. Because the amount of chemicals used varies with the year, the variation of chemical exposure and the year of blood collection has a certain impact on the predicted results. Our prediction should be periodically updated to incorporate new estimated exposure and measured C_Bs of chemicals in the future. Third, models based on subsets of measured data for chemical groups were not considered in the prediction model due to limited measured data. In the future, more accurate prediction models based on different chemical subsets could be built if we can collect sufficient data as a training set. In addition, we regarded the concentrations of blood, plasma, and serum as C_Bs and did not consider parameters such as plasma protein binding. Nonetheless, the predicted C_Bs in this study can still contribute to the concentrations' ranking of substances in human blood and the prioritization of potential biological effects. An accurate PBPK model could be combined with the C_B prediction model of this study in the future to predict concentrations in other organs, and animal experiments for validation of the model should be made in the future as well. Fourth, although the AC_{50} value has become a standard way to compare potencies of chemicals in *in vitro* pharmacology and toxicology studies, it may not be the best metric for prioritization or estimating toxicological risk based on well-designed *in vitro* tests. Fifth, the mode of toxic action (MOA), which was not considered in our prediction model, is related to the C_B and metabolism of the chemical. The MOA could be considered in future work to refine the model. Last, the present prioritization results based on ToxCast data have limitations in predicting the toxicities of the chemicals due to the limited assays adopted by the ToxCast exercise, and different chemicals were tested in different assays. Therefore, it is still impossible to systematically evaluate the contribution of one chemical in different toxicological end points.

In conclusion, we curated the C_Bs of 216 compounds and developed ML algorithms for C_B prediction, and our work improved HT risk prioritization for large numbers of environmental chemicals. Many of the high-risk chemicals in some assays were also unexpected. This study has implications for current efforts to overhaul existing chemical testing methods to address the disparity in the number of tested and untested chemicals. By using the HT method, chemicals could be screened in a cost-effective and efficient manner, which provides a better basis for informed decisions on chemical testing priorities and regulatory attention.

Acknowledgments

This work was funded by the National Key R&D Program (No. 2022YFC3702600 and 2022YFC3702601), the Singapore Ministry of Education Academic Research Fund Tier 1 (04MNP000567C120), and the Startup Grant of Fudan University (No. JIH 1829010Y).

In addition, to improve the applicability of our model, the R scripts are also provided at https://github.com/FangLabNTU/HExpPredict.

References

1. Shin H-M, Ernstoff A, Arnot JA, Wetmore BA, Csiszar SA, Fantke P, et al. 2015. Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays. Environ Sci Technol 49(11):6760–6771, PMID: 25932772, https://doi.org/10.1021/acs.est.5b00498.
2. Li L, Sangion A, Wania F, Armitage JM, Toose L, Hughes L, et al. 2021. Development and evaluation of a holistic and mechanistic modeling framework for chemical emissions, fate, exposure, and risk. Environ Health Perspect 129(12):127006, PMID: 34882502, https://doi.org/10.1289/EHP9372.
3. Ring CL, Arnot JA, Bennett DH, Egeghy PP, Fantke P, Huang L, et al. 2019. Consensus modeling of median chemical intake for the U.S. population based on predictions of exposure pathways. Environ Sci Technol 53(2):719–732, PMID: 30516957, https://doi.org/10.1021/acs.est.8b04056.
4. Wambaugh JF, Bare JC, Carignan CC, Dionisio KL, Dodson RE, Jolliet O, et al. 2019. New approach methodologies for exposure science. Curr Opin Toxicol 15:76–92, https://doi.org/10.1016/j.cotox.2019.07.001.
5. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5–12, PMID: 16963515, https://doi.org/10.1093/toxsci/kfl103.
6. Honda GS, Pearce RG, Pham LL, Setzer RW, Wetmore BA, Sipes NS, et al. 2019. Using the concordance of in vitro and in vivo data to evaluate extrapolation assumptions. PLoS One 14(5):e0217564, PMID: 31136631, https://doi.org/10.1371/journal.pone.0217564.
7. Blaauboer BJ. 2010. Biokinetic modeling and in vitro-in vivo extrapolations. J Toxicol Environ Health B Crit Rev 13(2–4):242–252, PMID: 20574900, https://doi.org/10.1080/10937404.2010.483940.
8. Wambaugh JF, Wetmore BA, Pearce R, Strope C, Goldsmith R, Sluka JP, et al. 2015. Toxicokinetic triage for environmental chemicals. Toxicol Sci 147(1):55–67, PMID: 26085347, https://doi.org/10.1093/toxsci/kfv118.
9. David A, Chaker J, Price EJ, Bessonneau V, Chetwynd AJ, Vitale CM, et al. 2021. Towards a comprehensive characterisation of the human internal chemical exposome: challenges and perspectives. Environ Int 156:106630, PMID: 34004450, https://doi.org/10.1016/j.envint.2021.106630.
10. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. 2014. The blood exposome and its role in discovering causes of disease. Environ Health Perspect 122(8):769–774, PMID: 24659601, https://doi.org/10.1289/ehp.1308015.
11. Jia S, Xu T, Huan T, Chong M, Liu M, Fang W, et al. 2019. Chemical isotope labeling exposome (CIL-EXPOSOME): one high-throughput platform for human urinary global exposome characterization. Environ Sci Technol 53(9):5445–5453, PMID: 30943026, https://doi.org/10.1021/acs.est.9b00285.
12. Zhao F, Kang Q, Zhang X, Liu J, Hu J. 2019. Urinary biomarkers for assessment of human exposure to monomeric aryl phosphate flame retardants. Environ Int 124:259–264, PMID: 30660026, https://doi.org/10.1016/j.envint.2019.01.022.
13. Sipes NS, Wambaugh JF, Pearce R, Auerbach SS, Wetmore BA, Hsieh J-H, et al. 2017. An intuitive approach for predicting potential human health risk with the Tox21 10k library. Environ Sci Technol 51(18):10786–10796, PMID: 28809115, https://doi.org/10.1021/acs.est.7b00650.
14. Armitage JM, Hughes L, Sangion A, Arnot JA. 2021. Development and intercomparison of single and multicompartment physiologically-based toxicokinetic models: implications for model selection and tiered modeling frameworks. Environ Int 154:106557, PMID: 33892222, https://doi.org/10.1016/j.envint.2021.106557.
15. Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, et al. 2015. Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. Toxicol Sci 148(1):121–136, PMID: 26251325, https://doi.org/10.1093/toxsci/kfv171.
16. Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, et al. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 125(1):157–174, PMID: 21948869, https://doi.org/10.1093/toxsci/kfr254.
17. Wadhwa R, Cascella M. Steady State Concentration. Treasure Island, FL: StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK553132/ [accessed 12 December 2022].
18. U.S. EPA (U.S. Environmental Protection Agency). Previously Published ToxCast Data. Updated data released October 2018. https://epa.figshare.com/articles/dataset/Previously_Published_ToxCast_Data/6062551/3 [accessed 4 March 2023].
19. Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, et al. 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform 9(1):61, PMID: 29185060, https://doi.org/10.1186/s13321-017-0247-6.
20. Wambaugh JF, Wetmore BA, Ring CL, Nicolas CI, Pearce RG, Honda GS, et al. 2019. Assessing toxicokinetic uncertainty and variability in risk prioritization. Toxicol Sci 172(2):235–251, PMID: 31532498, https://doi.org/10.1093/toxsci/kfz205.
21. Zhao F, Li L, Chen Y, Huang Y, Keerthisinghe TP, Chow A, et al. 2021. Risk-Based chemical ranking and generating a prioritized human exposome database. Environ Health Perspect 129(4):47014, PMID: 33929905, https://doi.org/10.1289/EHP7722.
22. Arnot JA, Brown TN, Wania F. 2014. Estimating screening-level organic chemical half-lives in humans. Environ Sci Technol 48(1):723–730, PMID: 24298879, https://doi.org/10.1021/es4029414.
23. Li L, Westgate JN, Hughes L, Zhang X, Givehchi B, Toose L, et al. 2018. A model for risk-based screening and prioritization of human exposure to chemicals from near-field sources. Environ Sci Technol 52(24):14235–14244, PMID: 30407800, https://doi.org/10.1021/acs.est.8b04059.
24. Ulrich N, Endo S, Brown TN, Watanabe N, Bronner G, Abraham MH, et al. 2017. *UFZ-LSER database v 3.2.1*. Leipzig, Germany: Helmholtz Centre for Environmental Research-UFZ.
25. U.S. EPA. 2022. Toxicity Estimation Software Tool (TEST). https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test [accessed 4 March 2023].
26. U.S. CDC (U.S. Centers for Disease Control and Prevention). 2022. National Report on Human Exposure to Environmental Chemicals. https://www.cdc.gov/exposurereport/ [accessed 8 December 2022].
27. Office of Environmental Health Hazard Assessment, California Department of Public Health. 2020. Explore Results: Biomonitoring California's Results Database. https://biomonitoring.ca.gov/results/explore [accessed 5 December 2021].
28. Neveu V, Moussy A, Rouaix H, Wedekind R, Pon A, Knox C, et al. 2017. Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. Nucleic Acids Res 45(D1):D979–D984, PMID: 27924041, https://doi.org/10.1093/nar/gkw980.
29. Dong T, Zhang Y, Jia S, Shang H, Fang W, Chen D, et al. 2019. Human indoor exposome of chemicals in dust and risk prioritization using EPA's ToxCast database. Environ Sci Technol 53(12):7045–7054, PMID: 31081622, https://doi.org/10.1021/acs.est.9b00280.
30. Jia S, Sankaran G, Wang B, Shang H, Tan ST, Yap HM, et al. 2019. Exposure and risk assessment of volatile organic compounds and airborne phthalates in Singapore's child care centers. Chemosphere 224:85–92, PMID: 30818198, https://doi.org/10.1016/j.chemosphere.2019.02.120.
31. Zhang Y, Liu M, Peng B, Jia S, Koh D, Wang Y, et al. 2020. Impact of mixture effects between emerging organic contaminants on cytotoxicity: a systems biological understanding of synergism between tris(1,3-dichloro-2-propyl)phosphate and triphenyl phosphate. Environ Sci Technol 54(17):10722–10734, PMID: 32786581, https://doi.org/10.1021/acs.est.0c02188.
32. U.S. EPA. 2022. CompTox Chemicals Dashboard (Version 2.1) Assay Endpoints List. https://comptox.epa.gov/dashboard/assay-endpoints?filtered= [accessed 8 December 2022].
33. Glynn A, Aune M, Darnerud PO, Cnattingius S, Bjerselius R, Becker W, et al. 2007. Determinants of serum concentrations of organochlorine compounds in Swedish pregnant women: a cross-sectional study. Environ Health 6:2, PMID: 17266775, https://doi.org/10.1186/1476-069X-6-2.
34. Koukoulakis KG, Kanellopoulos PG, Chrysochou E, Koukoulas V, Minaidis M, Maropoulos G, et al. 2020. Leukemia and PAHs levels in human blood serum: preliminary results from an adult cohort in Greece. Atmos Pollut Res 11(9):1552–1565, https://doi.org/10.1016/j.apr.2020.06.018.
35. Mathur H, Agarwal H, Johnson S, Saikia N. 2005. Analysis of Pesticide Residues in Blood Samples from Villages of Punjab. CSE/PML/PR-21, India, 1–15.
36. Mochalski P, King J, Klieber M, Unterkofler K, Hinterhuber H, Baumann M, et al. 2013. Blood and breath levels of selected volatile organic compounds in healthy volunteers. Analyst 138(7):2134–2145, PMID: 23435188, https://doi.org/10.1039/c3an36756h.
37. Zhao F, Wan Y, Zhao H, Hu W, Mu D, Webster TF, et al. 2016. Levels of blood organophosphorus flame retardants and association with changes in human sphingolipid homeostasis. Environ Sci Technol 50(16):8896–8903, PMID: 27434659, https://doi.org/10.1021/acs.est.6b02474.
38. U.S. EPA. 2022. CompTox Chemicals Dashboard (Version 2.1.1) Batch Search. https://comptox.epa.gov/dashboard/batch-search [accessed 4 March 2023].
39. Zhang X, Xie L, Long J, Xie Q, Zheng Y, Liu K, et al. 2021. Salidroside: a review of its recent advances in synthetic pathways and pharmacological properties. Chem Biol Interact 339:109268, PMID: 33617801, https://doi.org/10.1016/j.cbi.2020.109268.
40. National Library of Medicine, PubChem. PubChem Compound Summary for CID 6917, N-Vinyl-2-pyrrolidone. https://pubchem.ncbi.nlm.nih.gov/compound/N-Vinyl-2-pyrrolidone [accessed 22 January 2022].
41. Wang Y, Zhu H, Kannan K. 2019. A review of biomonitoring of phthalate exposures. Toxics 7(2):21, PMID: 30959800, https://doi.org/10.3390/toxics7020021.
42. Zhang B, He Y, Zhu H, Huang X, Bai X, Kannan K, et al. 2020. Concentrations of bisphenol A and its alternatives in paired maternal-fetal urine, serum and amniotic fluid from an e-waste dismantling area in China. Environ Int 136:105407, PMID: 31955035, https://doi.org/10.1016/j.envint.2019.105407.

43. Zhang S, Kang Q, Peng H, Ding M, Zhao F, Zhou Y, et al. 2019. Relationship between perfluorooctanoate and perfluorooctane sulfonate blood concentrations in the general population and routine drinking water exposure. Environ Int 126:54–60, PMID: 30776750, https://doi.org/10.1016/j.envint.2019.02.009.
44. Liu M, Jia S, Dong T, Zhao F, Xu T, Yang Q, et al. 2020. Metabolomic and transcriptomic analysis of MCF-7 cells exposed to 23 chemicals at human-relevant levels: estimation of individual chemical contribution to effects. Environ Health Perspect 128(12):127008, PMID: 33325755, https://doi.org/10.1289/EHP6641.
45. Fang M, Hu L, Chen D, Guo Y, Liu J, Lan C, et al. 2021. Exposome in human health: utopia or wonderland? Innovation (Camb) 2(4):100172, PMID: 34746906, https://doi.org/10.1016/j.xinn.2021.100172.