
Machine Learning on Toxicogenomic Data Reveals a Strong Association Between the Induction of Drug-Metabolizing Enzymes and Centrilobular Hepatocyte Hypertrophy in Rats

Kazuki Ikoma , [Takuomi Hosaka](#) , [Akira Ooka](#) , [Ryota Shizu](#) , [Kouichi Yoshinari](#) *

Posted Date: 17 March 2025

doi: 10.20944/preprints202503.1106.v1

Keywords: enzyme induction; hepatocyte hypertrophy; machine learning; toxicity mechanism



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Machine Learning on Toxicogenomic Data Reveals a Strong Association Between the Induction of Drug-Metabolizing Enzymes and Centrilobular Hepatocyte Hypertrophy in Rats

Kazuki Ikoma, Takuomi Hosaka, Akira Ooka, Ryota Shizu and Kouichi Yoshinari *

Laboratory of Molecular Toxicology, School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526 Japan

* Correspondence: yoshinari@u-shizuoka-ken.ac.jp; +81-54-264-5685

Abstract: Centrilobular hepatocyte hypertrophy is frequently observed in animal studies for chemical safety assessment. Although its toxicological significance and precise mechanism remain unknown, it is considered an adaptive response resulting from the induction of drug-metabolizing enzymes (DMEs). This study aimed to elucidate the association between centrilobular hepatocyte hypertrophy and DME induction using machine learning on toxicogenomic data. Utilizing publicly available gene expression data and pathological findings from rat livers of 134 compounds, we developed six different types of machine learning models to predict the occurrence of centrilobular hepatocyte hypertrophy based on gene expression data as explanatory variables. Among these, a LightGBM-based model demonstrated the best performance with an accuracy of approximately 0.9. With this model, we assessed each gene's contribution to predicting centrilobular hepatocyte hypertrophy using mean absolute SHAP values. The results revealed that *Cyp2b1* had an extremely significant contribution, while other DME genes also displayed positive contributions. Additionally, enrichment analysis of the top 100 genes based on mean absolute SHAP values identified "Metabolism of xenobiotics by cytochrome P450" as the most significantly enriched term. In conclusion, the current results suggest that the induction of multiple DMEs, including CYP2B1, is crucial for the development of centrilobular hepatocyte hypertrophy. (197 words)

Keywords: enzyme induction; hepatocyte hypertrophy; machine learning; toxicity mechanism

1. Introduction

Hepatomegaly and hepatocyte hypertrophy are commonly observed effects of chemical substances in toxicological studies involving experimental animals [1–3]. Hepatocyte hypertrophy can occur in specific liver regions, particularly in the centrilobular region. For example, phenobarbital, an antiepileptic drug, induces centrilobular hepatocyte hypertrophy in mice and rats, accompanied by the induction of microsomal drug-metabolizing enzymes (DMEs), including cytochrome P450s (P450s), along with the proliferation of smooth endoplasmic reticulum [2–6]. The underlying mechanism of centrilobular hepatocyte hypertrophy remains poorly understood. However, since many P450 inducers, such as phenobarbital, cause hepatocyte hypertrophy specifically in the centrilobular region, where they primarily induce DMEs, it is hypothesized that centrilobular hepatocyte hypertrophy occurs as a secondary effect of DME induction [2–6]. Nevertheless, there is no direct evidence to suggest that DME induction directly causes centrilobular hepatocyte hypertrophy.

In chemical toxicity studies, hepatocyte hypertrophy associated with DME induction is regarded as an adaptive change rather than a toxic effect, unless hepatic necrosis or elevated liver injury markers, such as alanine aminotransferase, γ -glutamyl transpeptidase, and alkaline phosphatase, are

present [2,5]. However, a relationship between hepatocyte hypertrophy and hepatocarcinogenesis has been suggested [7–9]. For example, long-term administration of WY-14643, an agonist for peroxisome proliferator-activated receptor alpha (PPAR α), or phenobarbital, which indirectly activates the nuclear receptor constitutive androstane receptor (CAR) through phosphorylation/dephosphorylation signaling [10], leads to hepatocyte hypertrophy, hepatocyte proliferation, and subsequent liver cancer in rodents [11–15]. Furthermore, kava root extract, derived from a plant in the Piperaceae family, has been reported to induce both hepatocyte hypertrophy and elevated serum γ -glutamyl transpeptidase and alkaline phosphatase levels in rats [4]. These findings suggest that hepatocyte hypertrophy induced by certain chemical compounds is linked to hepatotoxicity, necessitating further exploration of its mechanisms and toxicological significance.

Open TG-GATEs (Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System) [16] is a comprehensive toxicogenomics database containing gene expression data from livers and kidneys, as well as toxicity evaluation data, including hepatic and renal pathological findings and blood biochemistry results, from rats treated with over 100 chemical substances across multiple doses and durations. This data is publicly available through the Toxygates web platform [17]. Nagata et al. [18] used Open TG-GATEs data to explore genes contributing to increased relative liver weight, while Liu et al. [9] identified characteristic genes associated with hepatocyte hypertrophy leading to hepatocarcinogenesis.

Machine learning methods such as LightGBM [19] and XGBoost [20] have been developed and applied across various fields for prediction and classification tasks, achieving higher accuracy than conventional approaches. However, their predictive rationale often remains a black box. To address this limitation, SHapley Additive exPlanations (SHAP) [21] has been proposed as an explainable artificial intelligence (XAI) method that calculates the contribution of each feature to predictions based on game theory. In chemical toxicity prediction, SHAP has been utilized to identify critical structural features associated with toxicity [22] and to explore toxicity biomarkers [23].

In the present study, we aimed to elucidate the mechanisms of hepatocyte hypertrophy using machine learning approaches. Specifically, we developed a prediction model for chemical-induced centrilobular hepatocyte hypertrophy based on gene expression data from rat studies in Open TG-GATEs. Furthermore, we analyzed the model's predictive rationale using SHAP to quantitatively evaluate the contribution of each gene to the prediction and to identify genes crucial for the development of centrilobular hepatocyte hypertrophy.

2. Results

2.1. Dataset

We constructed a dataset by merging data from two sources. First, we collected liver pathological findings from rat repeated-dose toxicity studies in Open TG-GATEs, which included 134 compounds administered at three doses (low, medium, and high) over four different durations (3, 7, 14, or 28 days). We then combined this with the corresponding gene expression data from individual rats, obtained from Toxygates by using the sample ID as a matching key. The resulting dataset comprised dosing results from a total of 5,055 rats.

In this dataset, the positive rate for centrilobular hepatocyte hypertrophy across all samples was approximately 10%, indicating an imbalanced dataset with relatively few positive cases (Table 1). However, this positive rate increased with higher doses and longer administration periods, showing considerable variation, with rates ranging from 0% to 28.0% among the 12 groups stratified by dose and administration period (Table 1).

The distribution of samples across the five folds was well balanced (Table 2), with each fold containing between 991 and 1,039 administration results (samples) and between 28 and 30 unique compounds. The positive rates were also relatively consistent across the folds, ranging from 8.0% to 11.1%, which is close to the overall positive rate of 10.1% for the complete dataset. This balanced

distribution of both samples and positive cases across the folds helped ensure the reliability of our cross-validation results by minimizing potential bias from data splitting.

Supplementary Table S1 provides a comprehensive overview of the dataset, including compound names, administration durations, dose levels, labels (e.g., presence or absence of centrilobular hepatocyte hypertrophy), and assignments to the five folds for cross-validation.

Table 1. Summary of in vivo experimental results obtained from TG-GATEs. The numbers and proportions of positive samples for centrilobular hepatocyte hypertrophy across 134 compounds, three dosing levels, and four administration durations.

Dose level	Treatment duration (d)	Total samples	Positive samples	Positive rates (%)
High	28	393	110	28.0
	14	420	97	23.1
	7	429	67	15.6
	3	429	40	9.3
Middle	28	423	73	17.3
	14	423	44	10.4
	7	423	26	6.1
	3	423	15	3.5
Low	28	423	24	5.7
	14	423	9	2.1
	7	423	3	7.0
	3	423	0	0
All	–	5055	508	10.1

Table 2. Distribution of data across each fold for five-fold cross-validation. The total number of samples, the number of positive samples, positive rates, and the number of compounds included in each fold are presented.

Fold	Total samples	Positive samples	Positive rates (%)	Number of unique compounds
0	1004	104	10.4	28
1	1023	114	11.1	30
2	1039	108	10.4	29
3	991	79	8.0	28
4	998	103	10.3	28

2.2. Training and Validation of Prediction Models for Centrilobular Hepatocyte Hypertrophy

Using the constructed dataset, we trained and validated six different machine learning models: logistic regression, support vector machines (SVM), random forest (RF), XGBoost, LightGBM, and k-nearest neighbors (kNN). The presence or absence of centrilobular hepatocyte hypertrophy served as the objective variable, while gene expression ratios (log₂ fold changes) relative to the vehicle control group were used as explanatory variables. We performed stratified 5-fold cross-validation, dividing the 134 compounds into five groups to minimize the differences in positive rates among the groups (Supplementary Table S2).

Figure 1 presents the evaluation results for each prediction model. Both LightGBM and XGBoost, which are gradient boosting algorithms, demonstrated superior performance across various accuracy metrics compared to the other models. Specifically, LightGBM achieved the highest area under the receiver operating characteristic curve (ROC-AUC). After implementing early stopping, the mean ROC-AUC for LightGBM slightly improved from 0.885 to 0.887. The SVM exhibited the highest balanced accuracy (BA) of 0.78, while the RF showed a notably high recall of 0.82. However, it is important to note that BA and recall require the binarization of prediction results, and these evaluation metrics can vary significantly depending on the selected binarization threshold. Additionally, precision and F1-score exhibited greater variability across different splits in the 5-fold cross-validation than the other evaluation metrics for all models.

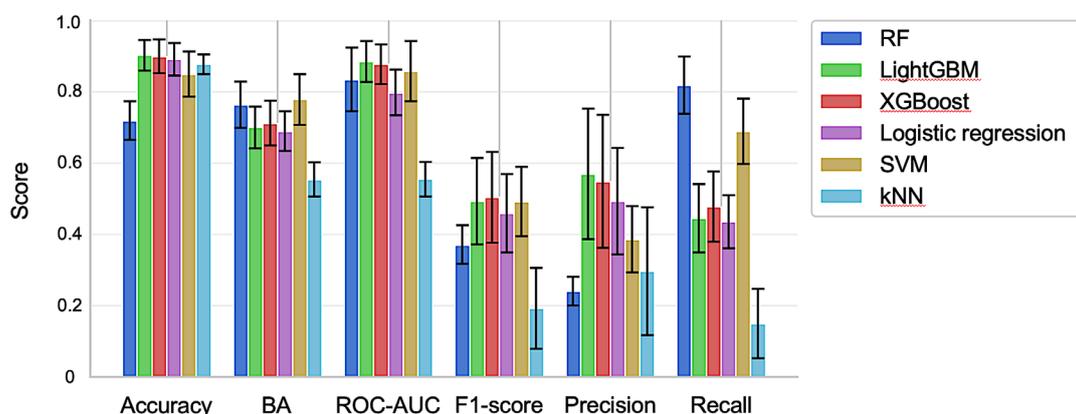
Figure 1

Figure 1. Performance of the machine learning models. The evaluation results for each prediction model are shown, with error bars representing the standard deviation of each metric.

2.3. Genes Contributing to the Prediction of Centrilobular Hepatocyte Hypertrophy

To identify the genes most influential in predicting centrilobular hepatocyte hypertrophy, we focused on the LightGBM model, which achieved the highest ROC-AUC score. We evaluated each gene's average contribution to the model's predictions using mean absolute SHAP values. Out of the 14,037 genes analyzed, 4,946 had mean absolute SHAP values greater than 0, indicating that approximately one-third of the genes were involved in the prediction process. The names and mean absolute SHAP values of these contributing genes are provided in Supplementary Table S3. Notably, the gene with the most significant contribution to the predictions was *Cyp2b1*, a CAR-target rat P450 gene, with a mean absolute SHAP value considerably higher than that of other genes.

Several of the top 10 genes, ranked by mean absolute SHAP values, were DME genes, including *Akr7a3* and *Aldh1a1* (Figure 2(a)). These top 10 genes showed a positive correlation between their fold-change values on a logarithmic scale with base 2 ($\text{Log}_2(\text{fold-change})$; Log_2FC) and SHAP values (Figure 2(b)). These findings suggest that the upregulation of DMEs, particularly CYP2B1, significantly contributes to the development of centrilobular hepatocyte hypertrophy.

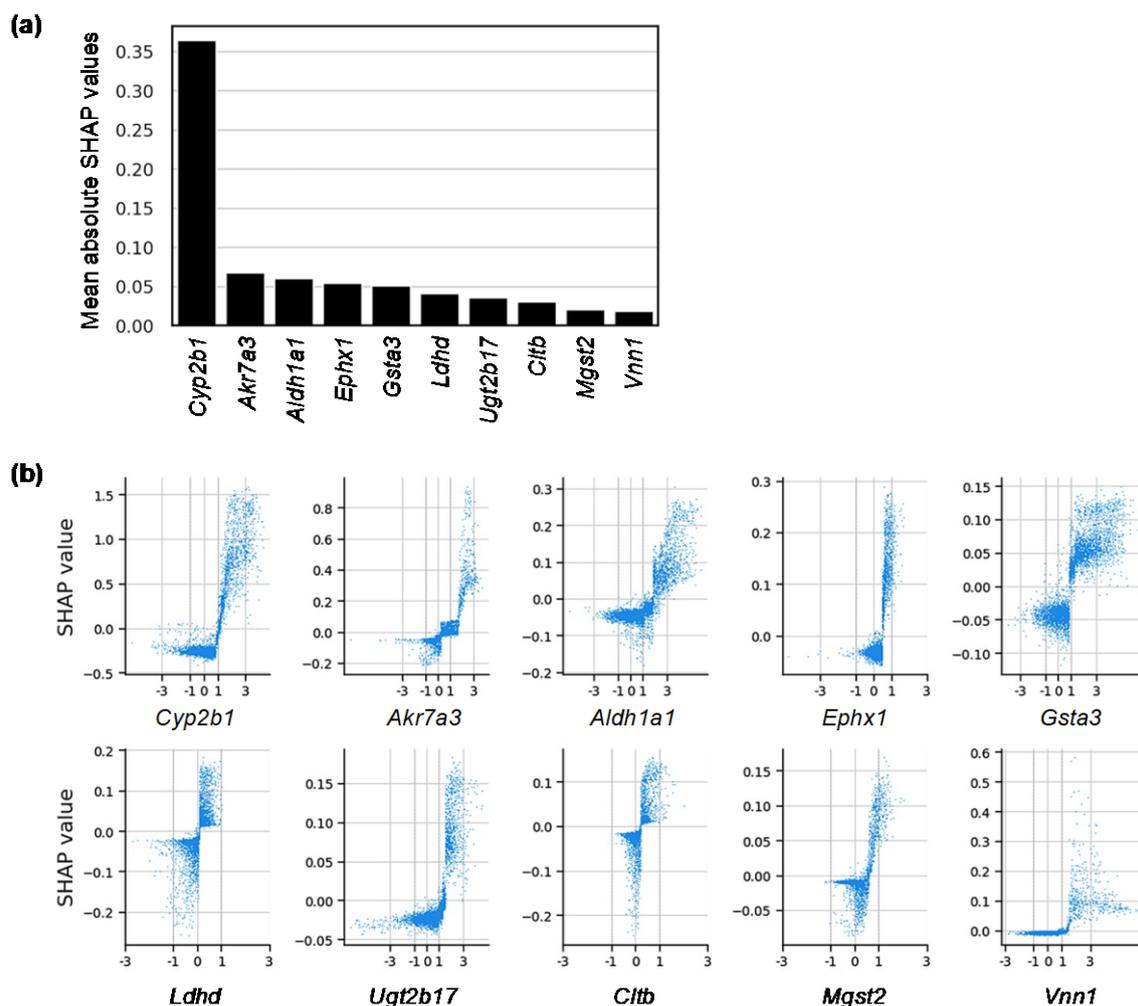
Figure 2

Figure 2. SHAP analysis of genes' contributions to model prediction. (a) A bar graph of the top 10 genes with mean absolute SHAP values is shown. (b) Relationships between SHAP values (*y*-axis) and Log₂FC (*x*-axis) for the top 10 genes for mean absolute SHAP values are shown as scatter plots.

2.4. Enrichment Analysis Using the Top 100 Genes that Contributed to the Prediction

To gain biological insights into the genes that significantly contribute to predicting centrilobular hepatocyte hypertrophy, we conducted an enrichment analysis using the top 100 genes ranked by mean absolute SHAP values from the LightGBM model (Figure 3). The most significantly enriched term was "Metabolism of xenobiotics by cytochrome P450," which included five of the top ten genes: *Cyp2b1*, *Akr7a3*, *Ephx1*, *Gsta3*, and *Ugt2b17*. This finding strongly suggests that the induction of DMEs, including P450 enzymes, plays a central role in the development of centrilobular hepatocyte hypertrophy.

In addition to the P450-related term, several other enriched terms were associated with drug exposure and metabolism, such as "Response to xenobiotic stimulus" and "Monocarboxylic acid metabolic process." These findings support the idea that the induction of DMEs, particularly P450s, is a crucial factor in the pathogenesis of centrilobular hepatocyte hypertrophy.

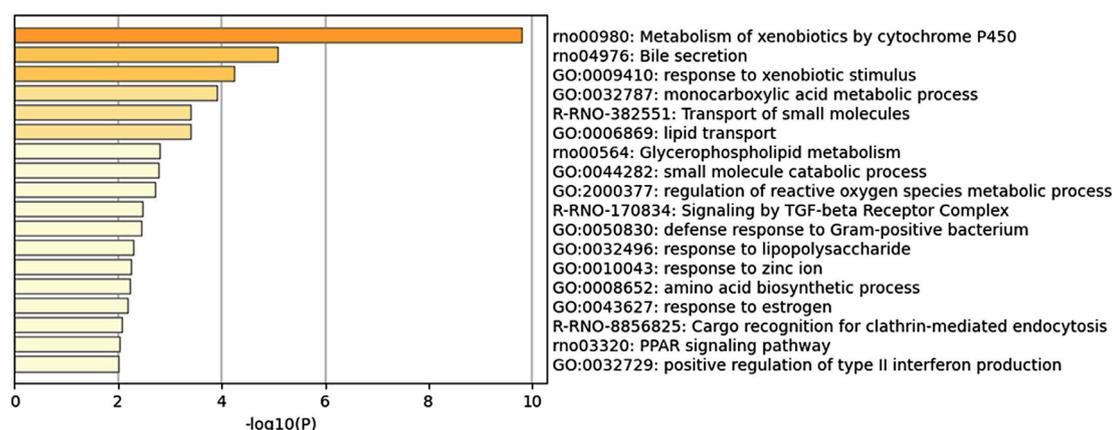
Figure 3

Figure 3. Enrichment analysis using the top 100 genes for mean absolute SHAP values. A bar graph of the enriched terms is shown.

2.5. Case Studies of Samples Showing CYP2B1 Induction with or Without Centrilobular Hepatocyte Hypertrophy

To further investigate the relationship between CYP2B1 induction and centrilobular hepatocyte hypertrophy, we analyzed individual samples, including those with CYP2B1 induction, with or without centrilobular hepatocyte hypertrophy. For example, samples treated with the typical enzyme inducer phenobarbital at high doses for 28 days (Phenobarbital/High/28-day) showed a *Cyp2b1* Log₂FC value of 3.828 and exhibited centrilobular hepatocyte hypertrophy. In contrast, samples treated with nifedipine at high doses for 28 days (Nifedipine/High/28-day) demonstrated a similar *Cyp2b1* Log₂FC value of 3.710 but did not exhibit centrilobular hepatocyte hypertrophy. Notably, nifedipine administration did not induce centrilobular hepatocyte hypertrophy at any dose level or duration (Supplementary Table S1).

To investigate why nifedipine did not induce centrilobular hepatocyte hypertrophy despite inducing *Cyp2b1* to levels comparable to phenobarbital, we quantified each gene's contribution to the prediction using SHAP value analysis for these two specific treatment conditions. The SHAP value analysis for Phenobarbital/High/28-day revealed positive values not only for *Cyp2b1* but also for other DME genes, including *Akr7a3*, *Ephx1*, and *Mgst2* (Figure 4(a), (c)), indicating that these DMEs positively contributed to the prediction of centrilobular hepatocyte hypertrophy in the LightGBM model. In contrast, the SHAP value analysis for Nifedipine/High/28-day indicated that although *Cyp2b1* and *Ephx1* had positive values signifying positive contributions to the prediction, these values were smaller than those in Phenobarbital/High/28-day, suggesting a lesser degree of contribution (Figure 4B,D). Additionally, *Akr7a3*, which showed significant positive contributions in Phenobarbital/High/28-day, demonstrated minimal contribution in Nifedipine/High/28-day, while *Mgst2*, which contributed positively in Phenobarbital/High/28-day, exhibited a negative contribution in Nifedipine/High/28-day (Figure 4(b), (d)).

This pattern of significant *Cyp2b1* contribution, along with limited involvement from other DME genes, was also observed with chemicals such as disopyramide and clomipramine, which induce CYP2B1 without leading to centrilobular hepatocyte hypertrophy. The Log₂FC values for *Cyp2b1* in high-dose, 28-day samples were 3.29 for disopyramide and 3.63 for clomipramine (Supplementary Figure S1).

These results suggest that although the induction of CYP2B1 significantly contributes to the development of centrilobular hepatocyte hypertrophy, it is insufficient on its own; multiple DMEs must also be induced for this to occur.

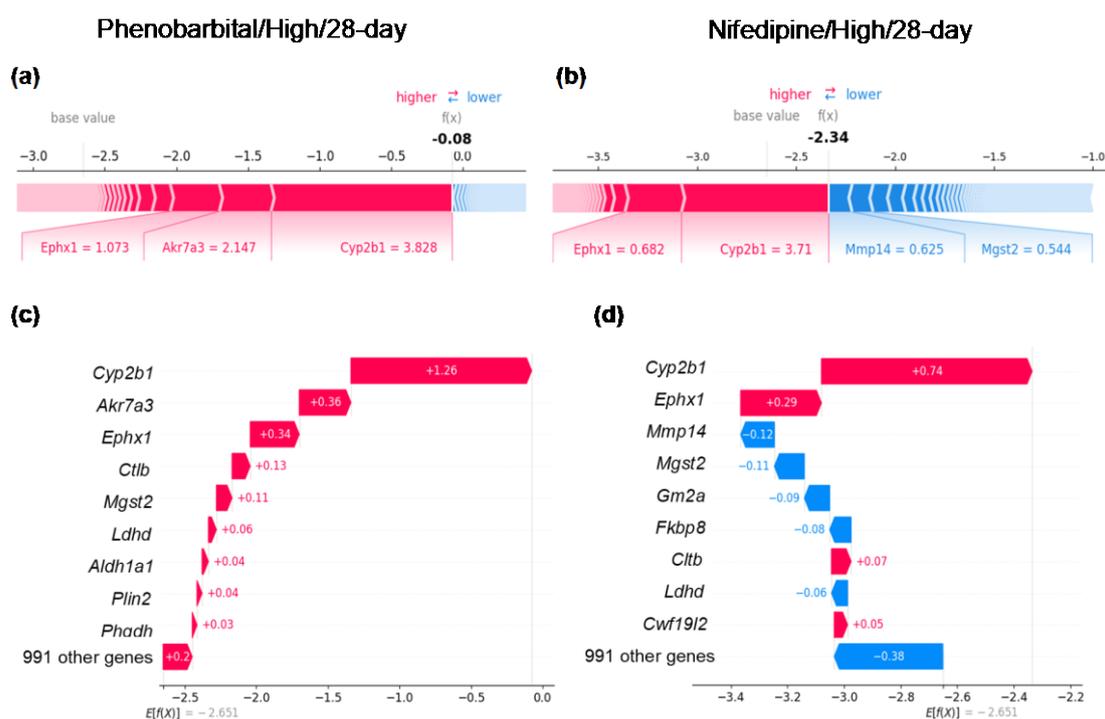
Figure 4

Figure 4. Case studies of SHAP analyses on gene contributions to model predictions. (a) and (b): SHAP value plots illustrating the impact of individual genes on model predictions for Phenobarbital/high/28-day (high-dose phenobarbital administration for 28 days) and Nifedipine/high/28-day (high-dose nifedipine administration for 28 days). Red bars indicate genes that positively contribute to the predictions, while blue bars represent genes that show negative contributions. The values shown are the log2FC of gene expression. (c) and (d): Waterfall plots visualizing the cumulative impact of genes on model predictions. The x-axis represents SHAP values, indicating the magnitude of each gene's contribution. The value presented is the SHAP value of the gene.

3. Discussion

In our LightGBM model for predicting centrilobular hepatocyte hypertrophy, *Cyp2b1* exhibited the most significant contribution to the prediction (Figure 2A). Other DME genes, including *Aldh1a1* and *Akr7a3*, also demonstrated substantial contributions, as samples with higher induction ratios of these DMEs tended to contribute more strongly to the prediction (Figure 2B). Furthermore, enrichment analysis of the top 100 genes with the highest contributions revealed that several DME-related terms were enriched (Figure 3). These findings suggest a close involvement of DME induction in the development of centrilobular hepatocyte hypertrophy. Chemical-induced CYP2B/3A induction is primarily observed in the centrilobular region, and many of these enzyme-inducing chemicals cause centrilobular hepatocyte hypertrophy [6]. Given that enzyme induction and hepatocyte hypertrophy are observed in the same region of the liver lobule, centrilobular hepatocyte hypertrophy is regarded as a secondary effect resulting from DME induction [2–6]. Our machine learning results using gene expression data support this hypothesis.

CYP2B induction by phenobarbital, a typical enzyme-inducing drug, is mediated through the activation of the nuclear receptor CAR [24,25]. Therefore, centrilobular hepatocyte hypertrophy can occur via different signals associated with CAR activation, beyond those involved in enzyme induction. While CAR activation promotes hepatocyte proliferation in rodents [11–14,26,27], the enrichment analysis did not reveal terms related to cell proliferation (Figure 3). These results suggest that the development of centrilobular hepatocyte hypertrophy is strongly linked to DME induction but has a limited association with other CAR functions, supporting the idea that, at least for CAR activators, centrilobular hepatocyte hypertrophy is a secondary phenomenon associated with DME

induction resulting from CAR activation, rather than the diverse signals resulting from CAR activation.

While our dataset included 44 P450 genes from the drug-metabolizing CYP1-CYP4 subfamilies, only *Cyp2b1* significantly contributed to the prediction of centrilobular hepatocyte hypertrophy. The induction of P450 involves xenobiotic-responsive transcription factors such as CAR, PXR, and AHR, with PXR primarily involved in CYP3A induction, CAR in CYP2B induction, and AHR in CYP1A induction [6,28,29]. Furthermore, the crosstalk between PXR and CAR in cooperatively inducing target gene expression is well documented [30]. Previous studies have shown that many activators of these transcription factors induce centrilobular hepatocyte hypertrophy, suggesting that multiple P450 forms could contribute to predicting this condition. However, the contribution of P450 to the prediction was almost exclusively limited to CYP2B1.

When examining the median Log2FC values for each P450 gene across all the samples, *Cyp2b1* showed the highest value, significantly exceeding the median values of other P450 forms (Supplementary Table S4). These results suggest that among P450s, CYP2B1 induction may be particularly important for the development of centrilobular hepatocyte hypertrophy. However, it is also possible that many of the enzyme-inducing compounds in our dataset were CAR activators. This may explain why the LightGBM model indicated significant contribution only from *Cyp2b1*, a representative CAR target gene.

We conducted a comparative analysis of SHAP values for two samples, Phenobarbital/High/28-day and Nifedipine/High/28-day, which exhibited similar levels of *Cyp2b1* induction but differed in the occurrence of centrilobular hepatocyte hypertrophy (Figure 4). Since phenobarbital and nifedipine are thought to induce *Cyp2b1* through the activation of CAR and PXR, respectively [31,32], this difference in induction mechanisms may influence the development of centrilobular hepatocyte hypertrophy. However, since other compounds that activate PXR also cause centrilobular hepatocyte hypertrophy [6], the difference in *Cyp2b1* induction mechanisms was not regarded as the primary factor in the development of hepatocyte hypertrophy.

The quantification of each gene's contribution to prediction through SHAP value analysis revealed that Nifedipine/High/28-day exhibited smaller SHAP values for genes with positive contributions, including *Cyp2b1*, compared to Phenobarbital/High/28-day. Furthermore, some DME genes, such as *Mgst2*, which showed positive contributions in Phenobarbital/High/28-day, demonstrated negative contributions in Nifedipine/High/28-day. These results suggest that while the Phenobarbital/High/28-day and Nifedipine/High/28-day samples demonstrated similar levels of CYP2B1 induction, the former showed the induction of additional DMEs, whereas the latter did not.

Given that cell size correlates with intracellular protein content [33–35], it is suggested that the induction of CYP2B1 alone is insufficient to cause centrilobular hepatocyte hypertrophy and that increasing intracellular protein levels through the induction of multiple DMEs is essential.

In this study, after evaluating multiple machine learning models based on various principles for predictive model construction, LightGBM demonstrated higher accuracy (including ROC-AUC) than other prediction models (Figure 1). Previous reports have also shown that LightGBM exhibits greater predictive accuracy than other machine learning models [36,37]. In particular, models based on gradient boosting algorithms, such as LightGBM, tend to show high predictive performance with large-scale datasets [38]. The dataset used in this study was medium-sized, comprising approximately 10,000 explanatory variables and about 5,000 data points, which likely explains why complex algorithms like LightGBM achieved high accuracy. Our current results are consistent with previous studies on toxicity prediction using machine learning.

We aimed to identify the genes involved in centrilobular hepatocyte hypertrophy by developing a toxicity prediction model and evaluating the toxicity mechanism based on the contribution of each feature to the predictions. This method is likely valuable for other analyses that utilize large-scale toxicogenomics databases. For example, Open TG-GATEs contains measurement values for liver injury markers, such as alanine aminotransferase, as well as information on various hepatic

pathological findings, including hepatocellular necrosis. Therefore, the approach employed in this study can estimate the mechanisms underlying different types of adverse effects.

In conclusion, by developing machine learning-based prediction models for centrilobular hepatocyte hypertrophy and analyzing the rationale behind these predictions, our current results suggest that the induction of CYP2B1 and other DMEs is essential for the development of centrilobular hepatocyte hypertrophy. These findings reinforce the hypothesis that centrilobular hepatocyte hypertrophy arises from DME induction. While the induction of these DMEs is considered a key factor in hepatocyte hypertrophy due to their role in increasing intracellular protein levels, further research is necessary to clarify the detailed mechanisms involved in this process.

4. Materials and Methods

4.1. Dataset

Pathological findings of the liver in rat repeated-dose toxicity studies were obtained from Open TG-GATEs [16], while gene expression data were obtained from Toxygates [17]. Open TG-GATEs provides gene expression data and toxicity evaluation results (including liver and kidney pathological findings and blood biochemistry results) from in vitro tests using rat and human primary hepatocytes, as well as in vivo tests using rats, where 170 compounds were administered at three dose levels over multiple durations. For in vivo studies, liver and kidney gene expression data were collected using the Affymetrix GeneChip system (Rat Genome 230 2.0 Array). These data were obtained in triplicate under identical dosing conditions (compound, dose, administration period). In Toxygates, the gene expression data from Open TG-GATEs was preprocessed and presented as logarithmic values with base 2 ($\text{Log}_2(\text{fold-change})$) relative to the vehicle control group.

Since these databases manage data using identical IDs to identify samples, representing data such as gene expression, pathological findings, and blood biochemistry results obtained from an individual rat under specific dosing conditions, we obtained liver pathological finding data from rat repeated-dose toxicity studies via Open TG-GATEs and the corresponding gene expression data from Toxygates. Then, using the IDs as keys, we combined the two datasets to construct the final dataset containing both pathological findings and gene expression data.

We obtained gene information for the 31,014 probes included in the Rat Genome 230 2.0 A and limited our analysis to those where one probe corresponded to a single gene. In cases where multiple probes were assigned to the same gene, we randomly selected one probe to represent the gene expression level for our analysis. The final number of probes (genes) included in the study was 14,037.

The presence or absence of centrilobular hepatocyte hypertrophy was determined by whether Liver/Centrilobular/Hypertrophy appeared in the pathological findings of Open TG-GATEs. For analytical convenience, we assigned a label of 1 for positive cases and 0 for negative cases. The labeling of centrilobular hepatocyte hypertrophy for each sample is presented in Supplementary Table S1.

4.2. Development of Prediction Models for Centrilobular Hepatocyte Hypertrophy

We trained six models: logistic regression [39], SVM [40], RF [41], LightGBM [19], XGBoost [20], and kNN [42], using gene expression Log_2FC values as explanatory variables and the presence or absence of centrilobular hepatocyte hypertrophy as the objective variable. These models were selected based on their established track record in various toxicity predictions [36,43–45].

All model training was conducted in a Python 3.10 environment using Scikit-learn (1.2.2) for logistic regression, SVM, RF, and kNN, along with LightGBM (4.1.0) and XGBoost (2.0.3). For initial model comparison, we used the default hyperparameters set in the libraries. For example, logistic regression employed L2 regularization (strength: 1.0) as the regularization parameter, and SVM utilized the RBF kernel.

After comparing the performance of all six models, LightGBM achieved the highest ROC-AUC. For this best-performing model, we implemented early stopping to further improve generalization

performance by preventing overfitting and reducing computation time. Specifically, we monitored the ROC-AUC values for validation data at each learning step. If the ROC-AUC value did not improve for 100 consecutive learning steps, we saved the model from the learning step that recorded the best ROC-AUC and terminated the training.

To evaluate the prediction accuracy of the established machine learning models, we performed 5-fold cross-validation, dividing the compounds into five groups using Scikit-learn's StratifiedGroupKFold. This method ensured minimal differences in positive rates among the folds and that results from the same compound were not included in different splits. The assignment of splits for the 5-fold cross-validation is documented in Supplementary Table S1.

4.3. Evaluation of Predictive Models

We performed validation by defining four of the five folds as training data and designating the remaining fold as validation data. This process was repeated five times to ensure that each fold served as validation data once. The prediction accuracy of each model was assessed by calculating the mean of the prediction accuracy metrics across the five validation folds. The evaluation metrics used were: 1) accuracy (the proportion of correctly predicted samples among all samples, assessing the model's overall performance), 2) BA (the average proportion of actual positive samples correctly predicted as positive and the proportion of actual negative samples accurately predicted as negative, a robust metric even for imbalanced data), 3) precision (the ratio of actual positive samples among those predicted as positive, crucial for minimizing false positives), 4) recall (the ratio of samples predicted as positive among actual positive samples, essential for minimizing false negatives), 5) F1-score (the harmonic mean of precision and recall, which considers the balance between these two metrics), and 6) ROC-AUC (a metric unaffected by the threshold for positives in predictions). These evaluation metrics were calculated using the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad \text{BA} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right), \quad \text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. A sample was considered positive if the predicted probability exceeded the overall positive rate of 0.1005.

We conducted a SHAP value analysis using the model that has the highest ROC-AUC value, as positive determination criteria do not impact this metric.

4.4. Calculation of SHAP Values

To evaluate each gene's contribution to the prediction of centrilobular hepatocyte hypertrophy, we employed SHAP [21]. SHAP utilizes the Shapley value concept from cooperative game theory in machine learning, allowing us to quantify and interpret how each feature (gene expression levels in this study) influences the model's predictions. A positive SHAP value for a given gene indicates that its expression level contributes to an increase in the predicted probability of centrilobular hepatocyte hypertrophy, while a negative value contributes to a decrease. We used TreeExplainer from the Python SHAP library (0.44.0) for SHAP value estimation, which is optimized for rapid computation with tree-structured models such as LightGBM.

In this study, we evaluated the average magnitude of each gene's contribution to prediction by using the mean absolute SHAP value. To calculate this value, we first computed SHAP values for each gene across all samples using the five models trained in the 5-fold cross-validation. We then calculated their absolute values (absolute SHAP values). Finally, we determined the mean absolute SHAP value from these absolute SHAP values across all samples and the five LightGBM models for each gene.

4.5. Enrichment Analysis

To perform a biological interpretation of genes with high mean absolute SHAP values, we conducted an enrichment analysis using Metascape [46]. *R. norvegicus* (rat) was selected as both the input and analysis species, and the top 100 genes ranked by mean absolute SHAP values from the LightGBM models were used as target genes for the analysis. For background genes, we employed all 14,037 genes as explanatory variables. The analysis incorporated three databases: Gene Ontology (GO) Biological Processes, Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway, and Reactome Pathway.

4.6. Computational Environment and Program Codes

The analysis was conducted using Google Colaboratory, where LightGBM and XGBoost were trained on a Tesla T4 GPU. The program codes and dataset used in this study are available upon reasonable request.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: SHAP analysis of genes' contributions to model predictions for clomipramine disopyramide.; Table S1: Administration conditions and labeling for centrilobular hepatocyte hypertrophy in the TG-GATEs dataset.; Table S2: Detailed performance metrics for each machine learning model across five-fold cross-validation.; Table S3: SHAP importance values for each gene across five-fold cross-validation in the LightGBM model: individual fold values, mean importance values, and gene names.; Table S4: SHAP importance values for each gene across five-fold cross-validation in the LightGBM model: individual fold values, mean importance values, and gene names.

Author Contributions: Conceptualization, Kazuki Ikoma and Kouichi Yoshinari; Data curation, Kazuki Ikoma; Formal analysis, Kazuki Ikoma; Investigation, Kazuki Ikoma, Takuomi Hosaka, Akira Ooka, Ryota Shizu and Kouichi Yoshinari; Project administration, Kouichi Yoshinari; Supervision, Takuomi Hosaka and Kouichi Yoshinari; Validation, Kazuki Ikoma; Visualization, Kazuki Ikoma; Writing – original draft, Kazuki Ikoma and Kouichi Yoshinari; Writing – review & editing, Kazuki Ikoma, Takuomi Hosaka, Akira Ooka, Ryota Shizu and Kouichi Yoshinari.. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The program codes are available at <https://github.com/Univ-Shizuoka-Eisei/Hypertrophy-Prediction> and the dataset used in this study is available upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Takeshita, J.; Goto, Y.; Yamamoto, S.; Sasaki, T.; Yoshinari, K., Comprehensive analysis of the toxicity-related findings from repeated-dose subacute toxicity studies of industrial chemicals in male rats. *Crit Rev Toxicol* **2024**, *54*, (10), 996-1010, 10.1080/10408444.2024.2427221.
2. Yoshida, M.; Umemura, T.; Kojima, H.; Inoue, K.; Takahashi, M.; Uramaru, N.; Kitamura, S.; Abe, K.; Tohkin, M.; Ozawa, S.; Yoshinari, K., [Basic principles of interpretation of hepatocellular hypertrophy in risk assessment in Japan]. *Shokuhin Eiseigaku Zasshi* **2015**, *56*, (2), 42-8, 10.3358/shokueishi.56.42.
3. Thoolen, B.; Maronpot, R. R.; Harada, T.; Nyska, A.; Rousseaux, C.; Nolte, T.; Malarkey, D. E.; Kaufmann, W.; Kuttler, K.; Deschl, U.; Nakae, D.; Gregson, R.; Vinlove, M. P.; Brix, A. E.; Singh, B.; Belpoggi, F.; Ward, J. M., Proliferative and nonproliferative lesions of the rat and mouse hepatobiliary system. *Toxicol Pathol* **2010**, *38*, (7 Suppl), 5S-81S, 10.1177/0192623310386499.
4. Clayton, N. P.; Yoshizawa, K.; Kissling, G. E.; Burka, L. T.; Chan, P. C.; Nyska, A., Immunohistochemical analysis of expressions of hepatic cytochrome P450 in F344 rats following oral treatment with kava extract. *Exp Toxicol Pathol* **2007**, *58*, (4), 223-36, 10.1016/j.etp.2006.08.002.

5. Hall, A. P.; Elcombe, C. R.; Foster, J. R.; Harada, T.; Kaufmann, W.; Knippel, A.; Kuttler, K.; Malarkey, D. E.; Maronpot, R. R.; Nishikawa, A.; Nolte, T.; Schulte, A.; Strauss, V.; York, M. J., Liver hypertrophy: a review of adaptive (adverse and non-adverse) changes--conclusions from the 3rd International ESTP Expert Workshop. *Toxicol Pathol* **2012**, *40*, (7), 971-94, 10.1177/0192623312448935.
6. Maronpot, R. R.; Yoshizawa, K.; Nyska, A.; Harada, T.; Flake, G.; Mueller, G.; Singh, B.; Ward, J. M., Hepatic enzyme induction: Histopathology. *Toxicol Pathol* **2010**, *38*, (5), 776-95, 10.1177/0192623310373778.
7. Allen, D. G.; Pearse, G.; Haseman, J. K.; Maronpot, R. R., Prediction of rodent carcinogenesis: an evaluation of prechronic liver lesions as forecasters of liver tumors in NTP carcinogenicity studies. *Toxicol Pathol* **2004**, *32*, (4), 393-401, 10.1080/01926230490440934.
8. Carmichael, N. G.; Enzmann, H.; Pate, I.; Waechter, F., The significance of mouse liver tumor formation for carcinogenic risk assessment: results and conclusions from a survey of ten years of testing by the agrochemical industry. *Environ Health Perspect* **1997**, *105*, (11), 1196-203, 10.1289/ehp.971051196.
9. Liu, S.; Kawamoto, T.; Morita, O.; Yoshinari, K.; Honda, H., Discriminating between adaptive and carcinogenic liver hypertrophy in rat studies using logistic ridge regression analysis of toxicogenomic data: The mode of action and predictive models. *Toxicol Appl Pharmacol* **2017**, *318*, 79-87, 10.1016/j.taap.2017.01.006.
10. Mutoh, S.; Sobhany, M.; Moore, R.; Perera, L.; Pedersen, L.; Sueyoshi, T.; Negishi, M., Phenobarbital indirectly activates the constitutive active androstane receptor (CAR) by inhibition of epidermal growth factor receptor signaling. *Sci Signal* **2013**, *6*, (274), ra31, 10.1126/scisignal.2003705.
11. Shizu, R.; Makida, N.; Sobe, K.; Ishimura, M.; Takeshita, A.; Hosaka, T.; Kanno, Y.; Sasaki, T.; Yoshinari, K., Interaction with YAP underlies the species differences between humans and rodents in CAR-dependent hepatocyte proliferation. *Toxicol Sci* **2024**, *198*, (1), 101-112, 10.1093/toxsci/kfad129.
12. Yoshinari, K., Role of nuclear receptors PXR and CAR in xenobiotic-induced hepatocyte proliferation and chemical carcinogenesis. *Biol Pharm Bull* **2019**, *42*, (8), 1243-1252, 10.1248/bpb.b19-00267.
13. Shizu, R.; Yoshinari, K., Nuclear receptor CAR-mediated liver cancer and its species differences. *Expert Opin Drug Metab Toxicol* **2020**, *16*, (4), 343-351, 10.1080/17425255.2020.1746268.
14. Yoshinari, K.; Shizu, R., Distinct Roles of the Sister Nuclear Receptors PXR and CAR in Liver Cancer Development. *Drug Metab Dispos* **2022**, *50*, (7), 1019-1026, 10.1124/dmd.121.000481.
15. Peters, J. M.; Cattley, R. C.; Gonzalez, F. J., Role of PPAR alpha in the mechanism of action of the nongenotoxic carcinogen and peroxisome proliferator Wy-14,643. *Carcinogenesis* **1997**, *18*, (11), 2029-33, 10.1093/carcin/18.11.2029.
16. Igarashi, Y.; Nakatsu, N.; Yamashita, T.; Ono, A.; Ohno, Y.; Urushidani, T.; Yamada, H., Open TG-GATES: a large-scale toxicogenomics database. *Nucleic Acids Res* **2015**, *43*, (Database issue), D921-7, 10.1093/nar/gku955.
17. Nyström-Persson, J.; Igarashi, Y.; Ito, M.; Morita, M.; Nakatsu, N.; Yamada, H.; Mizuguchi, K., Toxygates: interactive toxicity analysis on a hybrid microarray and linked data platform. *Bioinformatics* **2013**, *29*, (23), 3080-6, 10.1093/bioinformatics/btt531.
18. Nagata, K.; Washio, T.; Kawahara, Y.; Unami, A., Toxicity prediction from toxicogenomic data based on class association rule mining. *Toxicol Rep* **2014**, *1*, 1133-1142, 10.1016/j.toxrep.2014.10.014.
19. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. In *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017; Long Beach, CA, USA, 2017.
20. Chen, T.; Guestrin, C., XGBoost: A Scalable Tree Boosting System. In *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM: San Francisco, CA, USA, 2016; Vol. 11, pp 785 - 794.
21. Lundberg, S. M.; Lee, S.-I. In *A unified approach to interpreting model predictions*, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017; ACM: Long Beach, CA, USA, 2017; pp 4768 - 4777.

22. Mastrolorito, F.; Togo, M. V.; Gambacorta, N.; Trisciuzzi, D.; Giannuzzi, V.; Bonifazi, F.; Liantonio, A.; Imbrici, P.; De Luca, A.; Altomare, C. D.; Ciriaco, F.; Amoroso, N.; Nicolotti, O., TISBE: A Public Web Platform for the Consensus-Based Explainable Prediction of Developmental Toxicity. *Chem Res Toxicol* **2024**, *37*, (2), 323-339, 10.1021/acs.chemrestox.3c00310.
23. Pramudito, M. A.; Fuadah, Y. N.; Qauli, A. I.; Marcellinus, A.; Lim, K. M., Explainable artificial intelligence (XAI) to find optimal in-silico biomarkers for cardiac drug toxicity evaluation. *Sci Rep* **2024**, *14*, (1), 24045, 10.1038/s41598-024-71169-w.
24. Timsit, Y. E.; Negishi, M., CAR and PXR: the xenobiotic-sensing receptors. *Steroids* **2007**, *72*, (3), 231-246, 10.1016/j.steroids.2006.12.006.
25. Kodama, S.; Negishi, M., Phenobarbital confers its diverse effects by activating the orphan nuclear receptor car. *Drug Metab Rev* **2006**, *38*, (1-2), 75-87, 10.1080/03602530600569851.
26. Yamamoto, Y.; Moore, R.; Goldsworthy, T. L.; Negishi, M.; Maronpot, R. R., The orphan nuclear receptor constitutive active/androstane receptor is essential for liver tumor promotion by phenobarbital in mice. *Cancer Res* **2004**, *64*, (20), 7197-200, 10.1158/0008-5472.CAN-04-1459.
27. Kowalik, M. A.; Saliba, C.; Pibiri, M.; Perra, A.; Ledda-Columbano, G. M.; Sarotto, I.; Ghiso, E.; Giordano, S.; Columbano, A., Yes-associated protein regulation of adaptive liver enlargement and hepatocellular carcinoma development in mice. *Hepatology* **2011**, *53*, (6), 2086-96, 10.1002/hep.24289.
28. Hakkola, J.; Bernasconi, C.; Coecke, S.; Richert, L.; Andersson, T. B.; Pelkonen, O., Cytochrome P450 Induction and Xeno-Sensing Receptors Pregnane X Receptor, Constitutive Androstane Receptor, Aryl Hydrocarbon Receptor and Peroxisome Proliferator-Activated Receptor alpha at the Crossroads of Toxicokinetics and Toxicodynamics. *Basic Clin Pharmacol Toxicol* **2018**, *123* Suppl 5, 42-50, 10.1111/bcpt.13004.
29. Botts, S.; Ennulat, D.; Francke-Carroll, S.; Graham, M.; Maronpot, R. R.; Mohutsky, M., Introduction to hepatic drug metabolizing enzyme induction in drug safety evaluation studies. *Toxicol Pathol* **2010**, *38*, (5), 796-8, 10.1177/0192623310374330.
30. Oladimeji, P.; Cui, H.; Zhang, C.; Chen, T., Regulation of PXR and CAR by protein-protein interaction and signaling crosstalk. *Expert Opin Drug Metab Toxicol* **2016**, *12*, (9), 997-1010, 10.1080/17425255.2016.1201069.
31. Xiao, L.; Nickbarg, E.; Wang, W.; Thomas, A.; Ziebell, M.; Prosise, W. W.; Lesburg, C. A.; Taremi, S. S.; Gerlach, V. L.; Le, H. V.; Cheng, K. C., Evaluation of in vitro PXR-based assays and in silico modeling approaches for understanding the binding of a structurally diverse set of drugs to PXR. *Biochem Pharmacol* **2011**, *81*, (5), 669-79, 10.1016/j.bcp.2010.12.003.
32. Konno, Y.; Degawa, M., Gene activations of CYP2B1 and CYP3A1 by dihydropyridine calcium channel antagonists in the rat liver: the structure-activity relationship. *Biol Pharm Bull* **2004**, *27*, (6), 903-5, 10.1248/bpb.27.903.
33. Lanz, M. C.; Zatulovskiy, E.; Swaffer, M. P.; Zhang, L.; Ilerten, I.; Zhang, S.; You, D. S.; Marinov, G.; McAlpine, P.; Elias, J. E.; Skotheim, J. M., Increasing cell size remodels the proteome and promotes senescence. *Mol Cell* **2022**, *82*, (17), 3255-3269 e8, 10.1016/j.molcel.2022.07.017.
34. Rhind, N., Cell-size control. *Curr Biol* **2021**, *31*, (21), R1414-R1420, 10.1016/j.cub.2021.09.017.
35. Kempe, H.; Schwabe, A.; Cremazy, F.; Verschure, P. J.; Bruggeman, F. J., The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Mol Biol Cell* **2015**, *26*, (4), 797-804, 10.1091/mbc.E14-08-1296.
36. Zhang, J.; Mucs, D.; Norinder, U.; Svensson, F., LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity-Application to the Tox21 and Mutagenicity Data Sets. *J Chem Inf Model* **2019**, *59*, (10), 4150-4158, 10.1021/acs.jcim.9b00633.
37. Jaganathan, K.; Geethika, P. R.; Ramakrishnan, S.; Sundaram, D., Explainable AI and tree-based ensemble models: a comparative study in predicting chemical pulmonary toxicity. *The European Physical Journal Special Topics* **2024**, 10.1140/epjs/s11734-024-01291-3.
38. Seto, H.; Oyama, A.; Kitora, S.; Toki, H.; Yamamoto, R.; Kotoku, J.; Haga, A.; Shinzawa, M.; Yamakawa, M.; Fukui, S.; Moriyama, T., Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data. *Sci Rep* **2022**, *12*, (1), 15889, 10.1038/s41598-022-20149-z.

39. David W. Hosmer, J.; Lemeshow, S.; Sturdivant, R. X., *Applied Logistic Regression, Third Edition*. John Wiley & Sons, Inc.: Hoboken, New Jersey, USA, 2013.
40. Van Messem, A., Chapter 10 - Support vector machines: A robust prediction method with applications in bioinformatics. In *Handbook of Statistics*, Srinivasa Rao, A. S. R.; Rao, C. R., Eds. Elsevier: 2020; Vol. 43, pp 391-466.
41. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C.; Villa-Vialaneix, N., Random Forests for Big Data. *Big Data Research* **2017**, *9*, 28-46, <https://doi.org/10.1016/j.bdr.2017.07.003>.
42. Cover, T.; Hart, P., Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* **1967**, *13*, (1), 21-27, 10.1109/TIT.1967.1053964.
43. Cavasotto, C. N.; Scardino, V., Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point. *Acs Omega* **2022**, *7*, (51), 47536-47546, 10.1021/acsomega.2c05693.
44. Zhang, D.; Gong, Y., The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure. *IEEE Access* **2020**, *8*, 220990-221003, 10.1109/ACCESS.2020.3042848.
45. Wang, D.; Zhang, Y.; Zhao, Y., LightGBM: An Effective miRNA Classification Method in Breast Cancer Patients. In *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics*, Association for Computing Machinery: Newark, NJ, USA, 2017; pp 7-11.
46. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **2019**, *10*, (1), 1523, 10.1038/s41467-019-09234-6.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.