

## ARTICLE

# Non-targeted metabolomics for the identification of plasma metabolites associated with organic anion transporting polypeptide 1B1 function

Kreetta Hämäläinen<sup>1,2</sup>  | Päivi Hirvensalo<sup>1,2,3</sup>  | Mikko Neuvonen<sup>1,2</sup>  |  
Aleksi Tornio<sup>1,2,3,4</sup>  | Janne T. Backman<sup>1,2,5</sup>  | Marko Lehtonen<sup>6,7</sup>  | Mikko Niemi<sup>1,2,5</sup> 

<sup>1</sup>Department of Clinical Pharmacology, University of Helsinki, Helsinki, Finland

<sup>2</sup>Individualized Drug Therapy Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland

<sup>3</sup>Integrative Physiology and Pharmacology, Institute of Biomedicine, University of Turku, Turku, Finland

<sup>4</sup>Unit of Clinical Pharmacology, Turku University Hospital, Turku, Finland

<sup>5</sup>Department of Clinical Pharmacology, HUS Diagnostic Center, Helsinki University Hospital, Helsinki, Finland

<sup>6</sup>School of Pharmacy, Faculty of Health Science, University of Eastern Finland, Kuopio, Finland

<sup>7</sup>LC-MS Metabolomics Center, Biocenter Kuopio, Kuopio, Finland

## Correspondence

Mikko Niemi, Department of Clinical Pharmacology, University of Helsinki, PO Box 20, 00014 Helsinki, Finland.  
Email: [mikko.niemi@helsinki.fi](mailto:mikko.niemi@helsinki.fi)

## Funding information

European Research Council, Grant/Award Number: 282106; Sigrid Jusélius Foundation

## Abstract

Our aim was to evaluate biomarkers for organic anion transporting polypeptide 1B1 (OATP1B1) function using a hypothesis-free metabolomics approach. We analyzed fasting plasma samples from 356 healthy volunteers using non-targeted metabolite profiling by liquid chromatography high-resolution mass spectrometry. Based on *SLCO1B1* genotypes, we stratified the volunteers to poor, decreased, normal, increased, and highly increased OATP1B1 function groups. Linear regression analysis, and random forest (RF) and gradient boosted decision tree (GBDT) regressors were used to investigate associations of plasma metabolite features with OATP1B1 function. Of the 9152 molecular features found, 39 associated with OATP1B1 function either in the linear regression analysis ( $p < 10^{-5}$ ) or the RF or GBDT regressors (Gini impurity decrease  $> 0.01$ ). Linear regression analysis showed the strongest associations with two features identified as glycodeoxycholate 3-O-glucuronide (GDCA-3G;  $p = 1.2 \times 10^{-20}$  for negative and  $p = 1.7 \times 10^{-19}$  for positive electrospray ionization) and one identified as glycochenodeoxycholate 3-O-glucuronide (GCDCA-3G;  $p = 2.7 \times 10^{-16}$ ). In both the RF and GBDT models, the GCDCA-3G feature showed the strongest association with OATP1B1 function, with Gini impurity decreases of 0.40 and 0.17. In RF, this was followed by one GDCA-3G feature, an unidentified feature with a molecular weight of 809.3521, and the second GDCA-3G feature. In GBDT, the second and third strongest associations were observed with the GDCA-3G features. Of the other associated features, we identified with confidence two representing lysophosphatidylethanolamine 22:5. In addition, one feature was putatively identified as pregnanolone sulfate and one as pregnenolone sulfate. These results confirm GCDCA-3G and GDCA-3G as robust OATP1B1 biomarkers in human plasma.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *Clinical and Translational Science* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

## Study Highlights

### WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Many endogenous compounds are substrates of organic anion transporting polypeptide 1B1 (OATP1B1) and potential biomarkers for OATP1B1-mediated drug–drug interaction risk assessment. Previous studies have found glycochenodeoxycholate 3-O-glucuronide (GCDCA-3G) and glycodeoxycholate 3-O-glucuronide (GDCA-3G) to be highly sensitive and specific OATP1B1 biomarkers.

### WHAT QUESTION DID THIS STUDY ADDRESS?

Can novel OATP1B1 biomarker candidates be found using a hypothesis-free, non-targeted metabolomics approach? Which compounds are the most robust OATP1B1 biomarkers found in non-targeted metabolomics analysis of human plasma?

### WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

In addition to the previously identified GCDCA-3G and GDCA-3G, we found several metabolite features associated with OATP1B1 function, including lysophosphatidylethanolamine (LPE) 22:5 and putatively pregnanolone sulfate and pregnenolone sulfate.

### HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

These data confirm that GCDCA-3G and GDCA-3G are robust OATP1B1 biomarkers in human plasma. Further studies are required to investigate whether LPE 22:5, pregnanolone sulfate, and pregnenolone sulfate are OATP1B1 substrates and potential biomarkers.

## INTRODUCTION

Organic anion transporting polypeptide 1B1 (OATP1B1; encoded by *SLCO1B1*) is an influx transporter expressed on the sinusoidal membrane of human hepatocytes.<sup>1</sup> OATP1B1 transports multiple drugs, including 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors (statins), repaglinide, the angiotensin II receptor antagonists olmesartan and valsartan, as well as methotrexate.<sup>1</sup> Substantial variability exists in OATP1B1 function due to genetic variants in the *SLCO1B1* gene.<sup>1,2</sup> The no function c.521T>C (p.Val174Ala, rs4149056) single nucleotide variant (SNV) significantly increases the plasma concentrations of many OATP1B1 substrate drugs, such as statins.<sup>1,3–8</sup> This increases the risk of statin-associated musculoskeletal symptoms.<sup>8–10</sup> Furthermore, increased function *SLCO1B1* alleles have been associated with decreased plasma concentrations of OATP1B1 substrate drugs, such as rosuvastatin, simvastatin acid, and methotrexate.<sup>6,7,11</sup> Moreover, inhibition of OATP1B1 can impair the hepatic uptake of OATP1B1 substrates and increase their plasma concentrations.<sup>1,12–14</sup>

Endogenous substrates of drug transporters can potentially serve as biomarkers to aid in elucidating transporter-mediated drug–drug interactions in early phases of clinical drug development. In addition, biomarkers may enable the estimation of transporter function and aid in personalized drug dosing.<sup>15</sup> A previous non-targeted metabolomics study showed that the *SLCO1B1* c.521T>C SNV associates with increased plasma levels of several endogenous metabolites.<sup>16</sup> The strongest association was observed with a compound suggested to be a glucuronide conjugate of the bile acid glycochenodeoxycholate (GCDCA), but the structure of the compound was not fully confirmed. Using a targeted approach with authentic reference compounds, GCDCA 3-O-glucuronide (GCDCA-3G) and glycodeoxycholate 3-O-glucuronide (GDCA-3G) were subsequently identified as very sensitive and specific biomarkers for OATP1B1.<sup>17</sup> In addition to these molecules, coproporphyrins I (CPI) and III (CPIII) have also been suggested as specific biomarkers for OATP1B1.<sup>18,19</sup> A recent study, however, showed that GCDCA-3G and GDCA-3G perform better than CPI and especially CPIII in assessing OATP1B1 function.<sup>20</sup>

Although previous non-targeted and targeted metabolomics studies have identified several potential OATP1B1 biomarkers,<sup>16–20</sup> it is possible that all useful OATP1B1 biomarkers have not yet been discovered. Therefore, our aim was to identify novel OATP1B1 biomarkers and to evaluate the performance of already identified biomarkers using non-targeted metabolomics data from healthy volunteers with genotype-predicted OATP1B1 phenotypes.

## METHODS

### Study participants and samples

A total of 356 healthy, non-smoking unrelated White Finnish volunteers (183 women, 173 men) without continuous medication participated in the study. The mean  $\pm$  standard deviation (SD) age of the participants was  $24 \pm 4$  years, weight  $69.7 \pm 12.1$  kg, and body mass index  $22.9 \pm 2.7$  kg/m<sup>2</sup>. Each participant provided written informed consent. Blood samples for measuring plasma metabolome were collected as part of two previously published single-dose pharmacokinetic studies (Trial 1, European Union Drug Regulating Authorities Clinical Trials Database, EudraCT, number 2011-004645-40 and Trial 2, EudraCT number 2015-000540-41).<sup>6,21</sup> Trial 1 took place between January 2012 and October 2014 and Trial 2 between May 2015 and October 2017. Following an overnight fast at 7–8 a.m. and before the study drug administration, a 10 mL blood sample was collected from each participant into a light-protected ethylenediamine-tetraacetic acid-containing tube. The tubes were placed on ice immediately and plasma was separated within 30 min from sample collection. The plasma samples were stored at  $-80^{\circ}\text{C}$  until analysis. The study protocols were approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (Helsinki, Finland) and the Finnish Medicines Agency Fimea.

### Genotyping and OATP1B1 function grouping

All the participants were genotyped using a genome-wide microchip and targeted methods based on TaqMan chemistry as described previously.<sup>17</sup> We computed the *SLCO1B1* alleles from the c.388A>G (p.Asn130Asp, rs2306283), c.463C>A (p.Pro155Thr, rs11045819), c.521T>C, and c.1929A>C (p.Leu643Phe, rs34671512) SNVs with PHASE version 2.1.1.,<sup>22–24</sup> and defined them according to the Pharmacogene Variation Consortium.<sup>2</sup> The participants were classified into five OATP1B1 function phenotype groups based on the no function (\*5 and \*15), normal

function (\*1 and \*37), and increased function (\*14 and \*20) *SLCO1B1* alleles. Individuals homozygous or compound heterozygous for a no function allele were classified as poor OATP1B1 function group, those heterozygous for a no function allele as decreased OATP1B1 function group, those homozygous for a normal function allele as normal OATP1B1 function group, those compound heterozygous for a normal and an increased function allele as increased OATP1B1 function group, and those homozygous or compound heterozygous for an increased function allele as highly increased OATP1B1 function group. For data analysis, these function groups were encoded with a number between 0 and 4, corresponding to the above classification from poor (0) to highly increased function (4). The numeric grouping was treated as a continuous variable.

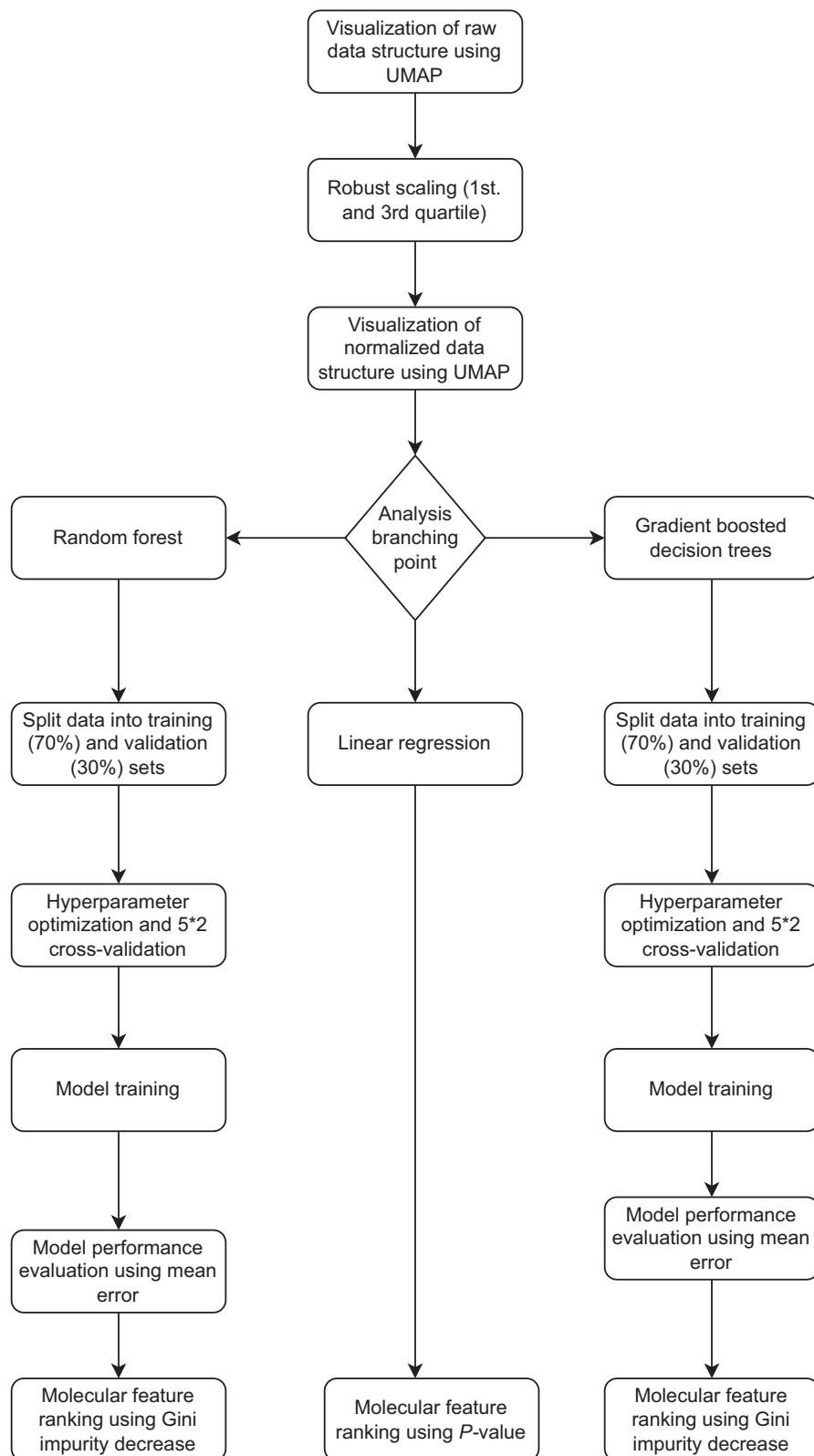
### Non-targeted metabolomic analysis

Non-targeted metabolomic profiling was performed at the Biocenter Kuopio LC–MS Metabolomics Center (University of Eastern Finland, Finland). The analysis was carried out using an ultra-high performance liquid chromatography (LC, Vanquish Flex UHPLC system, Thermo Scientific, Bremen, Germany) coupled online to a high-resolution mass spectrometer (MS, Q Exactive Focus, Thermo Scientific). All samples were analyzed using two different chromatographic techniques: reversed phase (RP) and hydrophilic interaction chromatography (HILIC). Data were acquired in both electrospray ionization (ESI) polarities: ESI positive (ESI+) and ESI negative (ESI–). Data-dependent product ion spectrums were acquired from pooled quality control (QC) samples at the beginning and end of the analysis for each mode. QC samples were injected at the beginning of the analysis and after every 12 samples. The LC–MS instrument setups and data acquisition parameters have been described previously.<sup>25</sup>

### Data preprocessing

Uniform Manifold Approximation and Projection (UMAP)<sup>26</sup> dimension reduction technique was used to analyze the structure of the raw molecular feature peak area data and to visually identify any distinct clusters. To minimize confounding effects of the data structure, the data were normalized by removing the median and scaling the data to the interquartile range using the robust scaler procedure from sklearn-library.<sup>27</sup> The data analysis pipeline is shown in Figure 1. After the normalization, visualization with UMAP was performed again to ensure successful scaling.

**FIGURE 1** Overview of the analysis pipeline for analyzing non-targeted metabolomic data using machine learning methods and linear regression analysis to discover molecular features associated with organic anion transporting polypeptide 1B1 (OATP1B1) function. UMAP, Uniform Manifold Approximation and Projection.



## Machine learning methods

Two decision tree-based machine learning algorithms, random forests (RF)<sup>28</sup> and gradient boosted decision trees (GBDT),<sup>29</sup> as well as a linear regression analysis

were employed in parallel to identify metabolite features associated with OATP1B1 function. The RFs and GBDTs were implemented with Python version 3.8.3 using sklearn-library (a library for machine learning in Python).<sup>30</sup> Initial testing indicated better performance

using regressors than classifiers, and therefore RFs and GBDTs were employed as regressors operating with continuous values. Normalized metabolite feature data were used as the input and numeric OATP1B1 phenotype groups as target values. The optimal hyperparameters were searched by sampling from specified ranges and using randomized search cross-validation for both RFs and GBDTs. To minimize overfitting, cross-validation was carried out with 5\*2-fold nested cross-validation. For both models, performances were evaluated using mean error between the target value and the predicted value. The strength of association of each metabolite feature with the OATP1B1 phenotype was quantified by calculating the average decrease in Gini impurity, a measurement of likelihood of an incorrect classification. A Gini impurity decrease of above 0.01 was considered as a potentially significant contribution to the classification. Linear regression analysis was implemented in R version 4.0.5.<sup>31</sup> The analysis was carried out for each metabolite feature using the normalized metabolite data as independent and numerically encoded OATP1B1 phenotype classes as dependent variables. A  $p$ -value  $<10^{-5}$  was considered statistically significant.

## Metabolite identification

Metabolite identification was focused on molecular features with a  $p$ -value  $<10^{-5}$  in the linear regression analysis or Gini impurity decrease value over 0.01 in the machine learning models. Metabolite identification was done using the open-source software MS-DIAL version 4.36.<sup>32</sup> Metabolite identifications were ranked according to previously published guidelines.<sup>33</sup> Metabolites in the level of identification (LI) 1 had the same exact molecular weight, retention time, and mass fragmentation as a pure reference compound. LI 2 includes metabolites with matching exact molecular weight and spectra from public libraries (METLIN, Lipidmaps and Human Metabolome DataBase were used) or in the case of lipids, the built-in MS-DIAL library version 4.00. In LI 3, only the chemical group of the compound, but not the exact compound, could be identified. In addition, the known OATP1B1 substrates or biomarkers CPI (654.27 g/mol), CPIII (654.27 g/mol), hexadecanedioate (284.20 g/mol), tetradecanedioate (256.17 g/mol), glycochenodeoxycholate 3-sulfate (GCDCA-3S) (529.27 g/mol), glycodeoxycholate 3-sulfate (GDCA-3S) (529.27 g/mol), bilirubin monoglucuronide (760.30 g/mol), bilirubin diglucuronide (936.33 g/mol), and bilirubin (584.26 g/mol) were searched from the entire molecular feature data based on molecular weight.<sup>1,13,16,18,34</sup>

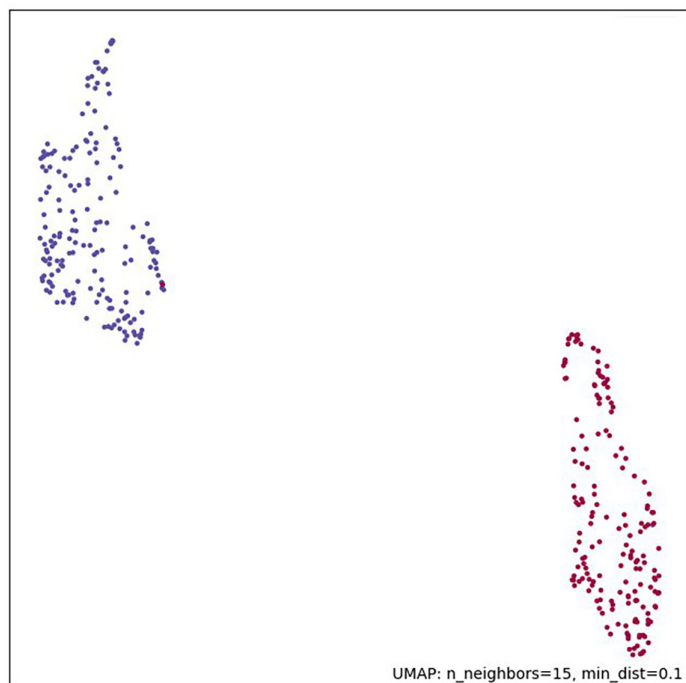
## RESULTS

Analysis of the raw metabolite data with UMAP revealed two distinct clusters (Figure 2), nearly completely differentiating the two clinical trials in which the samples were collected. Following the scaling and normalization procedure, no distinct clusters could be observed. Using all features, the RF and GBDT models were able to robustly identify different OATP1B1 function groups (Table 1, Figure 3).

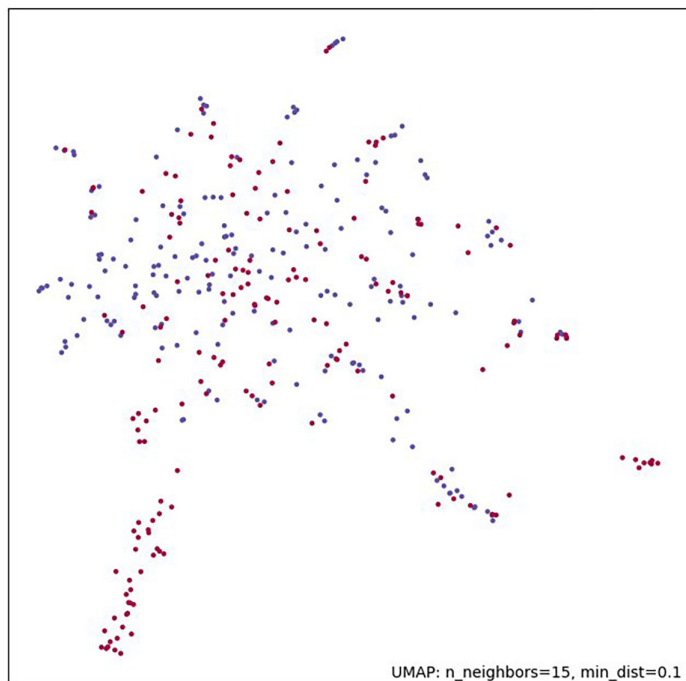
We found altogether 9152 molecular features in the non-targeted metabolomic analysis, of which 39 features associated with the OATP1B1 function (Table 2, Figure 4). A total of 28 features were significantly associated in the linear regression analysis ( $p < 10^{-5}$ ). In addition to these, RF and GBDT models discovered 11 molecular features with Gini impurity decreases of over 0.01. The mean  $\pm$  SD molecular weight of all the molecular features discovered in the non-targeted analysis was  $488.136 \pm 370.403$ . The 39 features associated with OATP1B1 function had a mean  $\pm$  SD molecular weight of  $596.490 \pm 317.194$  ( $p = 0.04$  compared to all other features with a mean  $\pm$  SD molecular weight of  $487.727 \pm 370.457$ ).

Of the 39 molecular features associated with OATP1B1 function, 17 had collision-induced dissociation data from the LC-MS analyses. Of these, seven could be confidently or putatively identified (Table 2). The top three molecular features in the linear regression analysis, which were also among the top four in the RF and GBDT models, had matching molecular weights to GCDCA-3G and GDCA-3G. Based on product ion spectrums and retention times of pure reference compounds, these features were identified as GDCA-3G (retention time 8.825 for negative ionization and 8.834 min for positive ionization) and GCDCA-3G (retention time 8.72 min). When using only the molecular feature representing GCDCA-3G in the RF and GBDT models, the model performances in identifying OATP1B1 function groups were slightly better than when using the entire metabolome dataset (Table 1, Figure 3).

Of the other associated features with collision-induced dissociation data we confidently identified lysophosphatidylethanolamine (LPE) 22:5 (LI 1) in both the positive and negative ionization modes. In addition, one feature was putatively identified as pregnanolone sulfate (LI 2) and one as pregnenolone sulfate (LI 3). Six of the remaining nine features with collision-induced dissociation data showed a fragment characteristic of a sulfate group ( $m/z$  96.96 in the negative ionization mode), but the compounds could not be identified further (molecular weights [g/mol] and retention times [min] of 809.35214 and 6.414, 428.22377 and 7.778, 396.19689 and 0.491, 289.62754 and 8.163, 511.26122 and 8.63, and 255.63058 and 8.613).



(a) Before scaling and normalization



(b) After scaling and normalization

■ Trial 1  
■ Trial 2

**FIGURE 2** Clustering of the samples based on entire metabolome dataset, with dimension reduction using the Uniform Manifold Approximation and Projection (UMAP) technique before (a) and after (b) data normalization by removing the median and scaling the data to the interquartile range using the robust scaling procedure. Individual samples are depicted with red (Trial 1) and blue (Trial 2) dots.

Two of the 39 features associated with OATP1B1 function had the same molecular weight as GCDCA-3S and GDCA-3S (529.27 g/mol), but no data-dependently collected product ion spectrum information was available. In addition, three other features with an identical molecular weight were found in the dataset. Retention times of pure reference compounds for GCDCA-3S (8.825 min) or

GDCA-3S (8.834 min) did not match with the retention times of any of these features.

Molecular features matching the molecular weights of CPI, CPIII, hexadecanedioate, tetradecanedioate, bilirubin monoglucuronide, or bilirubin diglucuronide were not found in the metabolome dataset. Several features not associated with OATP1B1 function had a

**TABLE 1** Comparison of the mean errors of the random forest and gradient boosted decision tree models for identification of organic anion transporting polypeptide 1B1 (OATP1B1) function groups, encoded as continuous values between 0 and 4, using the whole metabolome dataset or the top metabolite molecular feature representing glycochenodeoxycholate 3-O-glucuronide (GCDCA-3G).

OATP1B1 phenotype ( <i>n</i> )	Encoded value	Full model mean error		GCDCA-3G mean error	
		RF	GBDT	RF	GBDT
Poor (13)	0	0.366	0.467	0.255	0.242
Decreased (112)	1	0.241	0.357	0.208	0.268
Normal (170)	2	0.159	0.176	0.099	0.123
Increased (65)	3	0.245	0.363	0.191	0.200
Highly increased (5)	4	0.620	0.721	0.562	0.619
Overall (356)		0.213	0.279	0.159	0.191

Abbreviations: GBDT, gradient boosted decision tree; GCDCA-3G, glycochenodeoxycholate 3-O-glucuronide; OATP1B1, organic anion transporting polypeptide 1B1; RF, random forest.

molecular weight (584.26 g/mol) matching unconjugated bilirubin.

## DISCUSSION

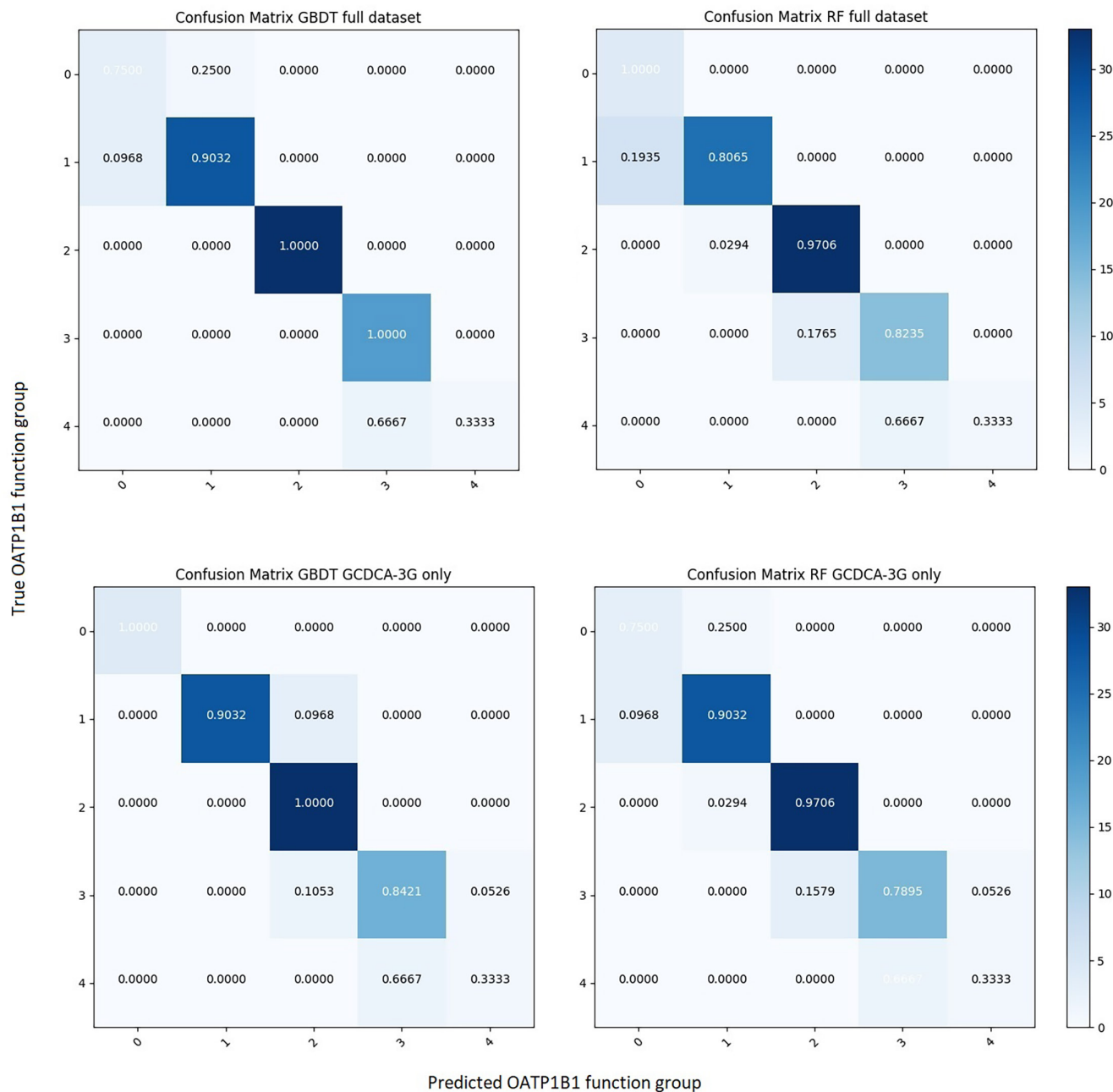
In this study, we investigated the effects of decreased and increased function *SLCO1B1* variants on plasma metabolite levels in healthy volunteers. Using a non-targeted approach, 9152 molecular features were found, of which 39 associated with OATP1B1 function in either linear regression analysis, RF, or GBDT. The RF and GBDT decision tree models were able to robustly separate the samples into the OATP1B1 function groups based on whole metabolome data. In the linear regression analysis, GDCA-3G showed the strongest association for OATP1B1 function, whereas RF and GBDT indicated strongest association for GCDCA-3G. Of the other features associated with OATP1B1 function, we were able to confidently identify LPE 22:5, and putatively pregnanolone sulfate and pregnenolone sulfate. In addition, several associated features appeared to include sulfate groups. Overall, the findings of this study support the feasibility of GCDCA-3G and GDCA-3G as robust endogenous OATP1B1 biomarkers.

The molecular features identified as GCDCA-3G and GDCA-3G were strongly associated with OATP1B1 function in all models. This suggests that they are the best performing OATP1B1 biomarkers found in non-targeted metabolomics analysis of human plasma. In a previous study, the mean plasma concentrations of GCDCA-3G and GDCA-3G were 8.57-fold and 5.76-fold higher in the poor OATP1B1 function group than in the normal OATP1B1 function group.<sup>20</sup> This previous study also showed that especially GCDCA-3G detected poor OATP1B1 function with high sensitivity and specificity. Substantial changes in plasma concentrations of GCDCA-3G and GDCA-3G have been detected in the presence of strong OATP1B1

inhibitors, and their plasma concentrations have been shown to increase even in the presence of weak OATP1B1 inhibition.<sup>35</sup> These data show that the plasma concentrations of GCDCA-3G and GDCA-3G depend strongly on OATP1B1 function.

Of the other features associated with OATP1B1 function, we confidently identified LPE 22:5 in both the positive and negative ionization modes and putatively identified one feature as pregnanolone sulfate and one as pregnenolone sulfate. LPE 22:5 is a phospholipid located in cellular membranes in all human tissues.<sup>36</sup> Pregnenolone sulfate and pregnanolone sulfate are endogenous excitatory neurosteroids (Figure 5).<sup>37</sup> Pregnenolone sulfate is a substrate of OATP2B1 and has been suggested to play a role in OATP2B1 regulation.<sup>38,39</sup> To the best of our knowledge, it is not known whether pregnenolone sulfate or pregnanolone sulfate is an OATP1B1 substrate. Interestingly, also several other features that associated with OATP1B1 function contained a sulfate group. Many sulfated bile acids as well as dehydroepiandrosterone sulfate and estrone sulfate are endogenous substrates of OATP1B1.<sup>1,40,41</sup> The molecular weights of the features associated with the OATP1B1 function were significantly higher than those of the other features. In line with this, previous studies have suggested that OATP substrates in general have a relatively high molecular weight.<sup>41</sup>

In addition to the 3-O-glucuronides, the sulfate conjugates of GDCA and GCDCA have been suggested as OATP1B1 biomarkers. For example, the non-selective OATP1B1 inhibitors cyclosporine A and rifampin (rifampicin) have increased the plasma concentrations of GDCA-3S and GCDCA-3S.<sup>13,14</sup> In a recent study, the plasma concentrations of GDCA-3S but not those of GCDCA-3S were significantly increased in individuals with the *SLCO1B1* no function genotype.<sup>40</sup> Five molecular features with an exact molecular weight identical to that of GDCA-3S and GCDCA-3S were found in the dataset. Two of these features associated with OATP1B1 activity.



**FIGURE 3** Confusion matrix on classification accuracy of the gradient boosted decision tree (GBDT) and random forest (RF) machine learning models for detecting the organic anion transporting polypeptide 1B1 (OATP1B1) function groups using the full metabolome dataset and the feature representing glycochenodeoxycholate 3-O-glucuronide (GCDCA-3G). The genetically determined OATP1B1 function groups are poor function (0), decreased function (1), normal function (2), increased function (3), and highly increased function (4). The values in the matrix are derived from normalizing confusion matrix over the true (rows) values. The sliding color scale indicates the number of observations in each cell.

However, the retention times of these features did not match with the retention time of the reference compounds of either GDCA-3S or GCDCA-3S, suggesting that these features may represent isomers of GDCA-3S or GCDCA-3S, or some other compounds. For example, the bile acid conjugates glycooursodeoxycholate sulfate (GUDCA-S), GCDCA-7S, and GDCA-7S have molecular weights identical to GCDCA-3S and GDCA-3S, but further studies are required to investigate

whether the features represent any of these compounds. Interestingly, the weak OATP1B inhibitor GDC-0810 has raised the peak plasma concentration of GUDCA-S by 57% and that of GCDCA-3G by 58%.<sup>35</sup>

In addition to GCDCA-3G and GDCA-3G, previous studies have identified CPI, CPIII, hexadecanedioate, and tetradecanedioate as potential OATP1B1 biomarkers.<sup>16,18,20,42</sup> Based on the exact molecular weights, CPI,



**TABLE 2** Associations of metabolite features found in non-targeted metabolomics analysis of human plasma with organic anion transporting polypeptide 1B1 (OATP1B1) function groups.

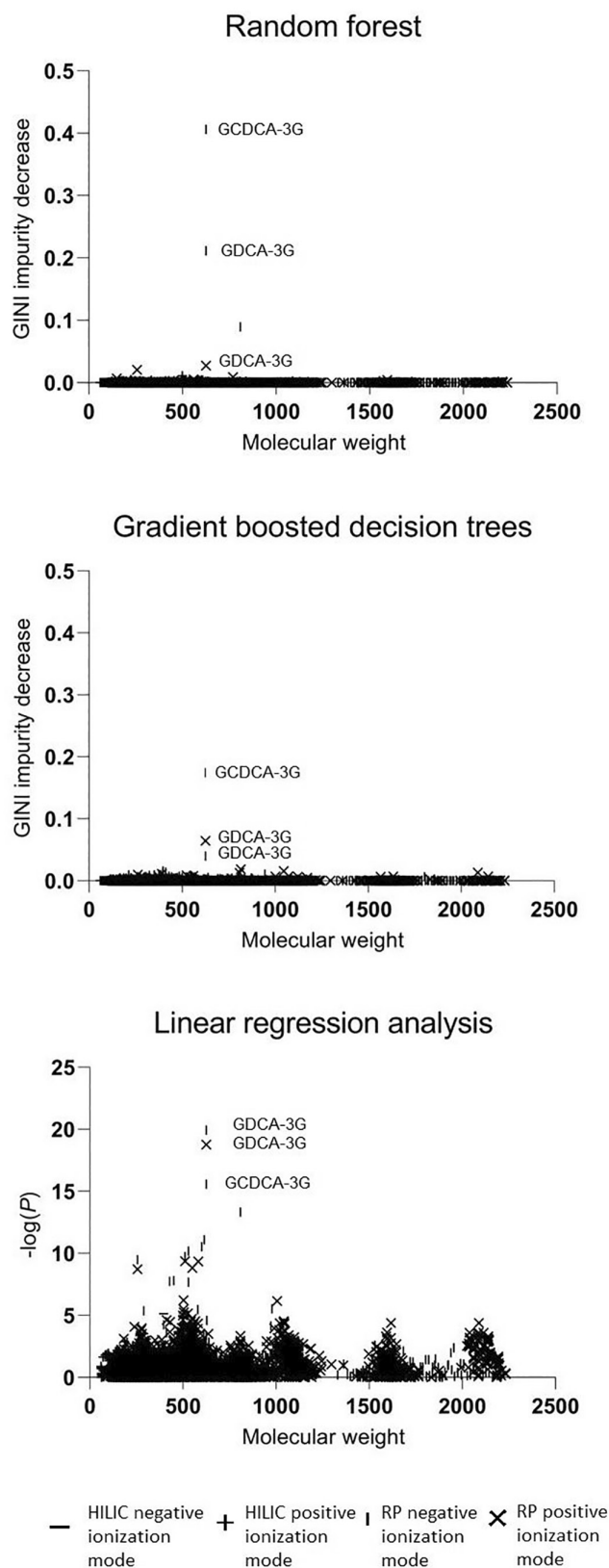
Chromatography	Ionization mode	Linear regression p-value	RF	GBDT	Molecular weight (g/mol)	Retention time (min)	Compound identification
RP	Negative	$1.2 \times 10^{-20}$	0.21	0.040	625.34756	8.825	GDCA-3G (LI 1)
RP	Positive	$1.7 \times 10^{-19}$	0.028	0.064	625.34571	8.834	GDCA-3G (LI 1)
RP	Negative	$2.7 \times 10^{-16}$	0.41	0.17	625.34722	8.72	GDCA-3G (LI 1)
RP	Negative	$4.7 \times 10^{-14}$	0.090	0.0090	809.35214	6.414	
RP	Negative	$7.9 \times 10^{-12}$	$<10^{-7}$	0.0030	614.3678	8.833	
RP	Negative	$2.9 \times 10^{-11}$	0.0043	$<10^{-7}$	601.29326	6.711	
RP	Negative	$6.8 \times 10^{-11}$	$<10^{-7}$	$<10^{-7}$	529.27196	8.984	
RP	Negative	$1.8 \times 10^{-10}$	$<10^{-7}$	$<10^{-7}$	511.26122	8.615	
RP	Negative	$3.2 \times 10^{-10}$	$<10^{-7}$	0.00007	255.63058	8.613	
RP	Positive	$4.2 \times 10^{-10}$	$<10^{-7}$	$<10^{-7}$	511.26022	8.63	
RP	Negative	$4.3 \times 10^{-10}$	$<10^{-7}$	$<10^{-7}$	289.62754	8.163	
RP	Positive	$4.5 \times 10^{-10}$	$<10^{-7}$	$<10^{-7}$	583.18982	8.984	
RP	Positive	$1.5 \times 10^{-9}$	$<10^{-7}$	$<10^{-7}$	550.18636	10.123	
RP	Positive	$1.9 \times 10^{-9}$	$<10^{-7}$	$<10^{-7}$	256.21901	10.125	
RP	Negative	$1.5 \times 10^{-8}$	$<10^{-7}$	$<10^{-7}$	450.2625	10.139	
RP	Negative	$1.8 \times 10^{-8}$	$<10^{-7}$	$<10^{-7}$	428.22377	7.778	
RP	Negative	$2.1 \times 10^{-8}$	0.0016	0.000092	529.27162	9.007	
RP	Positive	$6.1 \times 10^{-7}$	$<10^{-7}$	$<10^{-7}$	503.30107	10.365	
RP	Positive	$6.8 \times 10^{-7}$	$<10^{-7}$	0.000063	1006.60247	10.476	
RP	Negative	$3.0 \times 10^{-6}$	$<10^{-7}$	$<10^{-7}$	376.28314	10.477	
RP	Negative	$3.0 \times 10^{-6}$	$<10^{-7}$	$<10^{-7}$	978.5735	10.29	
RP	Negative	$3.4 \times 10^{-6}$	$<10^{-7}$	0.000004	579.25514	8.159	
RP	Negative	$4.3 \times 10^{-6}$	$<10^{-7}$	$<10^{-7}$	289.62754	8.163	
RP	Positive	$5.7 \times 10^{-6}$	$<10^{-7}$	0.0011	505.31686	10.71	
HILIC	Negative	$7.2 \times 10^{-6}$	$<10^{-7}$	$<10^{-7}$	396.19689	0.491	
RP	Positive	$7.2 \times 10^{-6}$	$<10^{-7}$	$<10^{-7}$	527.30103	10.414	LPE 22:5 (LI 1)
RP	Negative	$9.6 \times 10^{-6}$	$<10^{-7}$	$<10^{-7}$	503.30237	10.36	
RP	Negative	$9.5 \times 10^{-6}$	$<10^{-7}$	$<10^{-7}$	527.302	10.448	LPE 22:5 (LI 1)
RP	Positive	0.9	0.021	$<10^{-7}$	257.07133	7.628	
HILIC	Positive	0.9	0.011	$<10^{-7}$	499.29717	1.382	

(Continues)

TABLE 2 (Continued)

Chromatography	Ionization mode	Linear regression p-value	RF	GBDT	Molecular weight (g/mol)	Retention time (min)	Compound identification
RP	Positive	0.9	$<10^{-7}$	0.018	815.5847	13.538	
RP	Negative	0.1	$<10^{-7}$	0.016	398.21333	9.289	Pregnanolone sulfate (LI 2)
RP	Positive	0.01	$<10^{-7}$	0.015	1046.72526	10.788	
RP	Negative	0.0002	$<10^{-7}$	0.014	412.21055	9.13	
RP	Positive	0.01	$<10^{-7}$	0.013	2089.16939	10.331	
RP	Positive	0.002	$<10^{-7}$	0.013	809.43437	10.273	
RP	Negative	$5.9 \times 10^{-5}$	$<10^{-7}$	0.012	396.19763	8.65	Pregnenolone sulfate (LI 3)
RP	Positive	0.8	$<10^{-7}$	0.011	391.18396	6.59	
RP	Negative	0.03	$<10^{-7}$	0.001	376.28314	10.813	

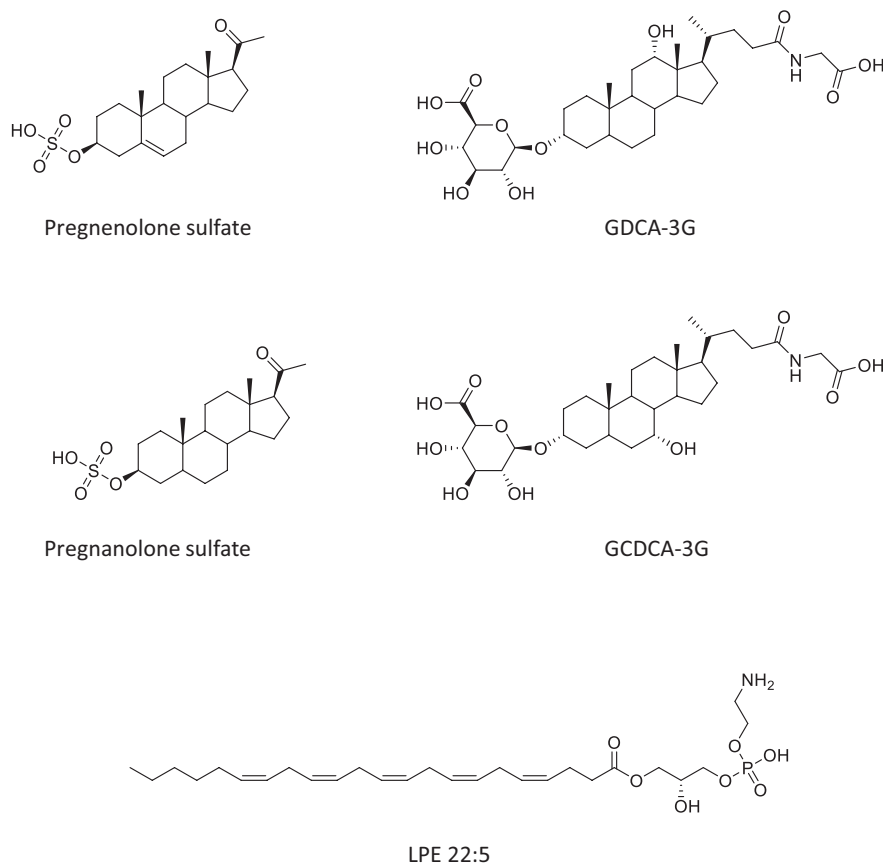
Abbreviations: GBDT, gradient boosted decision tree; GCDCA-3G, glycochenodeoxycholate 3-O-glucuronide; GDCA-3G, glycodeoxycholate 3-O-glucuronide; HILIC, hydrophilic interaction liquid chromatography; LI, level of identification; LPE, lysophosphatidylethanolamine; OATP1B1, organic anion transporting polypeptide 1B1; RF, random forest; RP, reversed phase liquid chromatography.



CPIII, hexadecanedioate, and tetradecanedioate were not found in our non-targeted metabolomic analysis of fasting plasma samples. This may be due to low concentrations in plasma or sensitivity to physical factors

**FIGURE 4** Associations of molecular features found in non-targeted metabolomics analysis of the fasting plasma samples from 356 healthy volunteers with genetically determined organic anion transporting polypeptide 1B1 (OATP1B1) function groups using random forest and gradient boosted decision tree machine learning models, and linear regression analysis. Each feature is represented by a symbol indicating the chromatographic technique and ionization polarity. Significance based on  $p$ -value or Gini impurity decrease is shown on the  $y$ -axis and exact molecular weight on the  $x$ -axis. GCDCA-3G, glycochenodeoxycholate 3-O-glucuronide; GDCA-3G, glycodeoxycholate 3-O-glucuronide; HILIC, hydrophilic interaction liquid chromatography; RP, reversed phase liquid chromatography.

**FIGURE 5** Chemical structures of pregnenolone sulfate, pregnanolone sulfate, glycodeoxycholate 3-O-glucuronide (GDCA-3G), glycochenodeoxycholate 3-O-glucuronide (GCDCA-3G), and lysophosphatidylethanolamine (LPE) 22:5. Only one isomer of LPE 22:5 is shown.



affecting their concentrations during prolonged storage, which makes them less reliable as potential biomarkers. The fasting plasma concentrations of CPI and CPIII in healthy volunteers are below 1 ng/mL, and those of hexadecanedioate and tetradecanedioate are in the range 1–100 ng/mL.<sup>20,42,43</sup>

The participants were stratified to poor, decreased, normal, increased, and highly increased OATP1B1 function groups based on their *SLCO1B1* genotypes as done previously.<sup>20</sup> The current Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline assigns the OATP1B1 phenotypes into four function groups: poor, decreased, normal, and increased function,<sup>8</sup> in which the increased function group is identical to the highly increased function group in our study. In our study, the increased function group consisted of individuals compound heterozygous for a normal and an increased function allele, who belong to the normal function group in the CPIC guideline classification. Previous studies have shown that

the concentrations of GCDCA-3G and CPI are lower in individuals compound heterozygous for a normal and an increased function allele than in those homozygous for a normal function allele,<sup>17,20</sup> supporting our classification system. The RF and GBDT machine learning models using either the whole metabolome dataset or only the GCDCA-3G feature identified the increased OATP1B1 function group with high accuracy, further supporting the idea that individuals compound heterozygous for a normal and an increased function allele have increased OATP1B1 activity as compared with those homozygous for a normal function allele.

The samples used in this study were collected during two separate pharmacokinetic studies.<sup>6,21</sup> Using the UMAP dimension reduction technique, the samples from these two trials formed two distinct clusters. Similarly, a previous study using the same samples found a significant difference in the concentrations of CPI and CPIII between the two studies.<sup>20</sup> The underlying cause for these

differences in the metabolite levels between the studies remains unclear. The only differences between the samples from the two studies were the time that the samples were stored in the deep freezer and the brand of the light-protected polypropylene freezing tubes. Nevertheless, the effects of the differences between the studies on the results should be minimal, as the normalization procedure removed the clustering entirely. This shows that standardization of both the collection and storage of the samples is crucially important in metabolomic biomarker studies.

RFs and GBDTs are machine learning algorithms based on decision trees. RFs combine the decision trees at the end of the process, whereas GBDTs combine the trees throughout the process using gradient boosting to minimize error. Linear regression analysis is simpler to implement than a machine learning approach and produces easily interpretable results, but functions best with linear relationships between the input and target values. We found that analysis using multiple complementary modeling techniques can be useful in identifying metabolite–transporter function associations as the relationships between molecular features and transporter function can be complex. Linear regression analysis and the decision tree-based machine learning models complement each other in this regard.

Our study has some limitations. First, we could not verify the identities of pregnanolone sulfate or pregnenolone sulfate using reference compounds. In addition, there were several metabolite features associated with OATP1B1 activity that could not be identified. Furthermore, a few previously identified OATP1B1 substrates or biomarkers could not be seen in the non-targeted metabolomics analysis. It is possible, that a more sensitive method could have captured these compounds. Moreover, as the use of machine learning methods in biomedical research is still evolving, there are no standardized cutoff values for significant Gini impurity decreases. Lastly, it is important to note that while the effect of genetically poor OATP1B1 function is most likely substrate-independent, the effects of OATP1B1 inhibitors may be substrate-dependent. Therefore, when investigating OATP1B1-mediated drug interactions with endogenous biomarkers, possible substrate-dependent effects should be considered.

To conclude, using non-targeted metabolomics we found that GCDCA-3G and GDCA-3G are robust and specific biomarkers for OATP1B1 function, a finding supported by previous targeted studies. In addition, we discovered possible novel biomarker candidates including LPE 22:5, pregnanolone sulfate, and pregnenolone sulfate. Non-targeted metabolomics analysis of carefully collected plasma samples from a relatively large homogenous group of healthy volunteers and the use

of genetically determined phenotype classification provides a robust method for discovering biomarkers for pharmacokinetic proteins.

## AUTHOR CONTRIBUTIONS

K.H., P.H., and M.Ni. wrote the manuscript. K.H., P.H., and M.Ni. designed the research. K.H., P.H., M.Ne., A.T., J.T.B., M.L., and M.Ni. performed the research. K.H., P.H., and M.Ni. analyzed the data.

## ACKNOWLEDGMENTS

The authors wish to thank Eija Mäkinen-Pulli and Lisbet Partanen for skillful technical assistance. Seppo Auriola is gratefully acknowledged for valuable assistance with non-targeted metabolomics. The School of Pharmacy mass spectrometry laboratory is supported by Biocenter Finland and Biocenter Kuopio.

## FUNDING INFORMATION

This study was supported by grants from European Research Council (Grant agreement 282106), the Sigrid Jusélius Foundation, and state funding for university-level health research (Helsinki, Finland).

## CONFLICT OF INTEREST STATEMENT

All authors declare no competing interests for this work.

## ORCID

Kreetta Hämäläinen  <https://orcid.org/0000-0002-1659-649X>

Päivi Hirvensalo  <https://orcid.org/0000-0002-9125-5716>

Mikko Neuvonen  <https://orcid.org/0000-0002-7627-1638>

Aleksi Tornio  <https://orcid.org/0000-0001-5713-5692>

Janne T. Backman  <https://orcid.org/0000-0002-9577-2788>

Marko Lehtonen  <https://orcid.org/0000-0002-5611-0466>

Mikko Niemi  <https://orcid.org/0000-0003-4550-2189>

<https://orcid.org/0000-0002-9577-2788>

<https://orcid.org/0000-0002-5611-0466>

<https://orcid.org/0000-0002-9577-2788>

<https://orcid.org/0000-0002-5611-0466>

<https://orcid.org/0000-0003-4550-2189>

## REFERENCES

- Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. *Pharmacol Rev.* 2011;63:157-181.
- Ramsey LB, Gong L, Lee SB, et al. PharmVar GeneFocus: SLCO1B1. *Clin Pharmacol Ther.* 2023;113:782-793.
- Niemi M, Pasanen MK, Neuvonen PJ. SLCO1B1 polymorphism and sex affect the pharmacokinetics of pravastatin but not fluvastatin. *Clin Pharmacol Ther.* 2006;80:356-366.
- Pasanen MK, Neuvonen M, Neuvonen PJ, Niemi M. SLCO1B1 polymorphism markedly affects the pharmacokinetics of simvastatin acid. *Pharmacogenet Genomics.* 2006;16:873-879.

5. Pasanen MK, Fredrikson H, Neuvonen PJ, Niemi M. Different effects of SLCO1B1 polymorphism on the pharmacokinetics of atorvastatin and rosuvastatin. *Clin Pharmacol Ther.* 2007;82:726-733.
6. Mykkänen AJH, Taskinen S, Neuvonen M, et al. Genomewide association study of simvastatin pharmacokinetics. *Clin Pharmacol Ther.* 2022;112:676-686.
7. Lehtisalo M, Taskinen S, Tarkiainen EK, et al. A comprehensive pharmacogenomic study indicates roles for SLCO1B1, ABCG2 and SLCO2B1 in rosuvastatin pharmacokinetics. *Br J Clin Pharmacol.* 2023;891:242-252.
8. Cooper-DeHoff RM, Niemi M, Ramsey LB, et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1, ABCG2, and CYP2C9 genotypes and statin-associated musculoskeletal symptoms. *Clin Pharmacol Ther.* 2022;111:1007-1021.
9. Link E, Parish S, Armitage J, et al. SLCO1B1 variants and statin-induced myopathy – a genomewide study. *N Engl J Med.* 2008;359:789-799.
10. Lönnberg KI, Tornio A, Hirvensalo P, et al. Real-world pharmacogenetics of statin intolerance: effects of SLCO1B1, ABCG2, and CYP2C9 variants. *Pharmacogenet Genomics.* 2023;33:153-160.
11. Ramsey LB, Bruun GH, Yang W, et al. Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Res.* 2012;22:1-8.
12. Hedman M, Neuvonen PJ, Neuvonen M, Holmberg C, Antikainen M. Pharmacokinetics and pharmacodynamics of pravastatin in pediatric and adolescent cardiac transplant recipients on a regimen of triple immunosuppression. *Clin Pharmacol Ther.* 2004;75:101-109.
13. Mori D, Kimoto E, Rago B, et al. Dose-dependent inhibition of OATP1B by rifampicin in healthy volunteers: comprehensive evaluation of candidate biomarkers and OATP1B probe drugs. *Clin Pharmacol Ther.* 2020;107:1004-1013.
14. Mochizuki T, Zamek-Gliszczynski MJ, Yoshida K, et al. Effect of cyclosporin a and impact of dose staggering on OATP1B1/1B3 endogenous substrates and drug probes for assessing clinical drug interactions. *Clin Pharmacol Ther.* 2022;111:1315-1323.
15. Müller F, Sharma A, König J, Fromm MF. Biomarkers for in vivo assessment of transporter function. *Pharmacol Rev.* 2018;702:246-277.
16. Yee SW, Giacomini MM, Hsueh CH, et al. Metabolomic and genome-wide association studies reveal potential endogenous biomarkers for OATP1B1. *Clin Pharmacol Ther.* 2016;1005:524-536.
17. Neuvonen M, Hirvensalo P, Tornio A, et al. Identification of glycochenodeoxycholate 3-O-glucuronide and glycodeoxycholate 3-O-glucuronide as highly sensitive and specific OATP1B1 biomarkers. *Clin Pharmacol Ther.* 2021;109:646-657.
18. Lai Y, Mandlekar S, Shen H, et al. Coproporphyrins in plasma and urine can be appropriate clinical biomarkers to recapitulate drug-drug interactions mediated by organic anion transporting polypeptide inhibition. *J Pharmacol Exp Ther.* 2016;358:397-404.
19. Kikuchi R, Chothe PP, Chu X, et al. Utilization of OATP1B biomarker coproporphyrin-I to guide drug-drug interaction risk assessment: evaluation by the pharmaceutical industry. *Clin Pharmacol Ther.* 2023;114:1170-1183.
20. Neuvonen M, Tornio A, Hirvensalo P, Backman JT, Niemi M. Performance of plasma coproporphyrin I and III as OATP1B1 biomarkers in humans. *Clin Pharmacol Ther.* 2021;110:1622-1632.
21. Hirvensalo P, Tornio A, Neuvonen M, et al. Enantiospecific pharmacogenomics of fluvastatin. *Clin Pharmacol Ther.* 2019;106:668-680.
22. Stephens M, Smith N, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 2001;68:978-989.
23. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* 2003;73:1162-1169.
24. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Human Gen.* 2005;76:449-462.
25. Driuchina A, Hintikka J, Lehtonen M, et al. Identification of gut microbial lysine and histidine degradation and CYP-dependent metabolites as biomarkers of fatty liver disease. *MBio.* 2023;14:e0266322.
26. McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection for dimension reduction. *J Open Source Softw.* 2018;3:861.
27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825-2830.
28. Breiman L. Random forests. *Mach Learn.* 2001;45:5-32.
29. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 2001;29:1189-1232.
30. Van Rossum G, Drake FL. Python 3 Reference Manual. <https://docs.python.org/3/reference/index.html>. Updated 2009. Accessed November 14, 2023.
31. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>. 2018. Accessed November 14, 2022.
32. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods.* 2015;12:523-526.
33. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics.* 2007;3:211-221.
34. Cui Y, König J, Leier I, Buchholz U, Keppler D. Hepatic uptake of bilirubin and its conjugates by the human organic anion transporter SLC21A6. *J Biol Chem.* 2001;276:9626-9630.
35. Yoshida K, Jaochico A, Mao J, Sangaraju D. Glycochenodeoxycholate and glycodeoxycholate 3-O-glucuronides, but not hexadecanedioate and tetradecanedioate, detected weak inhibition of OATP1B caused by GDC-0810 in humans. *Br J Clin Pharmacol.* 2023;89:1903-1907.
36. Cong L, Wan Z, Li P, et al. Metabolic, genetic, and pharmacokinetic parameters for the prediction of olanzapine efficacy. *Eur J Pharm Sci.* 2022;177:106277.
37. Grube M, Köck K, Oswald S, et al. Organic anion transporting polypeptide 2B1 is a high-affinity transporter for atorvastatin and is expressed in the human heart. *Clin Pharmacol Ther.* 2006;80(6):607-620.
38. Koenen A, Köck K, Keiser M, Siegmund W, Kroemer HK, Grube M. Steroid hormones specifically modify the activity of organic anion transporting polypeptides. *Eur J Pharm Sci.* 2012;47(4):774-780.

39. Tóth B, Jani M, Beéry E, et al. Human OATP1B1 (SLCO1B1) transports sulfated bile acids and bile salts with particular efficiency. *Toxicol in Vitro*. 2018;52:189-194.
40. Orozco CC, Neuvonen M, Bi YA, et al. Characterization of bile acid sulfate conjugates as substrates of human organic anion transporting polypeptides. *Mol Pharm*. 2023;20:3020-3032.
41. Hagenbuch B, Meier PJ. Organic anion transporting polypeptides of the OATP/SLC21 family: phylogenetic classification as OATP/SLCO superfamily, new nomenclature and molecular/functional properties. *Pflugers Arch*. 2004;447:653-665.
42. Shen H, Chen W, Drexler DM, et al. Comparative evaluation of plasma bile acids, dehydroepiandrosterone sulfate, hexadecanedioate, and tetradecanedioate with coproporphyrins I and III as markers of OATP inhibition in healthy subjects. *Drug Metab Dispos*. 2017;45:908-919.
43. Yee SW, Giacomini MM, Shen H, et al. Organic anion transporter polypeptide 1B1 polymorphism modulates the extent of drug-drug interaction and associated biomarker levels in healthy volunteers. *Clin Transl Sci*. 2019;12:388-399.

**How to cite this article:** Hämäläinen K, Hirvensalo P, Neuvonen M, et al. Non-targeted metabolomics for the identification of plasma metabolites associated with organic anion transporting polypeptide 1B1 function. *Clin Transl Sci*. 2024;17:e13773. doi:[10.1111/cts.13773](https://doi.org/10.1111/cts.13773)