

Supporting Information

Integrative Data Mining, Scaffold Analysis, and Sequential Binary Classification Models for Exploring Ligand Profiles of Hepatic Organic Anion Transporting Polypeptides (OATPs)

Alžběta Türková,[†] Sankalp Jain,[†] Barbara Zdrzil^{†}*

[†]University of Vienna, Department of Pharmaceutical Chemistry, Division of Drug Design and Medicinal Chemistry, Althanstraße 14, A-1090 Vienna, Austria.

*Corresponding author: barbara.zdrzil@univie.ac.at

Table of Contents

Table S1.....	p3
Table S2.....	p7
Table S3.....	p10
Table S4.....	p10
Table S5.....	p11
Table S6.....	p12
Table S7.....	p13
Table S8.....	p14
Table S9.....	p15
Table S10.....	p15
Table S11.....	p16
Table S12.....	p17
Table S13.....	p20
Table S14.....	p21
Table S15.....	p22
Table S16.....	p23
Table S17.....	p24

Figure S1.....	p25
Figure S2.....	p26
Figure S3.....	p27
Figure S4.....	p28
Figure S5.....	p30
Figure S6.....	p31
Figure S7.....	p32
Figure S8.....	p33
Figure S9.....	p34
Figure S10.....	p35
Description of Supplementary Data Files.....	p36

Table S1. Data sets annotated with bioactivity end point “inhibition” from ChEMBL: Reference to the original manuscript, PMID, compound concentration used in the experiment, number of unique compounds (after standardization procedure) and recommended activity thresholds are given.

Reference	PMID	c [uM]	Cmpd. number	Recommended threshold
Marada, V.V.; Flörl, S.; Kühne, A.; Burckhardt, G.; Hagos, Y. Interaction of human organic anion transporter polypeptides 1B1 and 1B3 with antineoplastic compounds. <i>Eur. J. Med. Chem.</i> 2015 , <i>92</i> , 723-731.	25618019	100	8	> 60% is active
De Bruyn, T.; van Westen, G.J.P., IJzerman, A.P., Stieger, B., de Witte, P., Augustijns, P.F., Annaert, P.P. Structure-Based Identification of OATP1B1/3 Inhibitors. <i>Mol. Pharmacol.</i> 2013 , <i>83</i> (6), 1257-1267.	23571415	10	1367	> 50% is active
Karlgren, M.; Vildhede, A.; Norinder, U.; Wisniewski, J. R.; Kimoto, E.; Lai, Y.; Haglund, U.; Artursson, P. Classification of Inhibitors of Hepatic Organic Anion Transporting Polypeptides (OATPs): Influence of Protein Expression on Drug–Drug Interactions. <i>J. Med. Chem.</i> 2012 , <i>55</i> (10), 4740–4763.	22541068	20	221	> 50% is active
Kobayashi, D.; Nozawa, T.; Imai, K.; Nezu, J., Tsuji, A.; Tamai, I. Involvement of human organic anion transporting polypeptide OATP-B (SLC21A9) in pH-dependent transport across intestinal apical membrane. <i>J. Pharmacol. Exp. Ther.</i> 2003 , <i>306</i> (2), 703-708.	12724351	1000 10000	2	REMOVED

<p>St-Pierre, M. V.; Hagenbuch, B.; Ugele, B.; Meier, P. J.; Stallmach, T. Characterization of an organic anion-transporting polypeptide (OATP-B) in human placenta. <i>J. Clin. Endocrinol. Metab.</i> 2002, 87(4), 1856-1863</p>	<p>11932330</p>	<p>100</p>	<p>1</p>	<p>> 60% is active (based on the threshold from Marada, V.V.; Flörl, S.; Kühne, A.; Burckhardt, G.; Hagos, Y. Interaction of human organic anion transporter polypeptides 1B1 and 1B3 with antineoplastic compounds. <i>Eur. J. Med. Chem.</i> 2015, 92, 723-731.)</p>
<p>Ishiguro, N.; Shimizu, H.; Kishimoto, W.; Ebner, T.; Schaefer, O. Evaluation and prediction of potential drug-drug interactions of linagliptin using in vitro cell culture methods. <i>Drug Metab. Dispos.</i> 2013, 41(1), 149-158</p>	<p>23073734</p>	<p>100</p>	<p>1</p>	<p>> 60% is active (based on the threshold from Marada, V.V.; Flörl, S.; Kühne, A.; Burckhardt, G.; Hagos, Y. Interaction of human organic anion transporter polypeptides 1B1 and 1B3 with antineoplastic compounds. <i>Eur. J. Med. Chem.</i> 2015, 92, 723-731.)</p>

<p>Sandhu, P.; Lee, W.; Xu, X.; Leake, B. F.; Yamazaki, M.; Stone, J. A.; Lin, J. H.; Pearson, P. G.; Kim, R. B. Hepatic uptake of the novel antifungal agent caspofungin. <i>Drug Metab. Dispos.</i> 2005, 33(1), 676</p>	<p>15716364</p>	<p>100 10</p>	<p>2</p>	<p>> 60% is active (for c=100 uM)</p> <p>(based on the threshold from Marada, V.V.; Flörl, S.; Kühne, A.; Burckhardt, G.; Hagos, Y. Interaction of human organic anion transporter polypeptides 1B1 and 1B3 with antineoplastic compounds. <i>Eur. J. Med. Chem.</i> 2015, 92, 723-731.)</p> <p>> 50% is active (for c=10 uM)</p> <p>(based on the threshold from De Bruyn, T.; van Westen, G.J.P., IJzerman, A.P., Stieger, B., de Witte, P., Augustijns, P.F., Annaert, P.P. Structure-Based Identification of OATP1B1/3 Inhibitors. <i>Mol. Pharmacol.</i> 2013, 83 (6), 1257-1267.)</p>
<p>Nozawa, T.; Minami, H.; Sugiura, S.; Tsuji, A.; Tamai, I. Role of organic anion transporter OATP1B1 (OATP-C) in hepatic uptake of irinotecan and its active metabolite, 7-ethyl-10-hydroxycamptothecin: in vitro evidence and effect of single nucleotide polymorphisms. <i>Drug Metab. Dispos.</i>, 2005, 33(3), 434-439</p>	<p>15608127</p>	<p>10</p>	<p>2</p>	<p>> 50% is active</p> <p>(based on the threshold from De Bruyn, T.; van Westen, G.J.P., IJzerman, A.P., Stieger, B., de Witte, P., Augustijns, P.F., Annaert, P.P. Structure-Based Identification of OATP1B1/3 Inhibitors. <i>Mol. Pharmacol.</i> 2013, 83 (6), 1257-1267.)</p>

Nozawa, T.; Sugiura, S.; Nakajima, M.; Goto, A.; Yokoi, T.; Nezu, J. I.; Tsuji, A.; Tamai, I. Involvement of organic anion transporting polypeptides in the transport of troglitazone sulfate: implications for understanding troglitazone hepatotoxicity. <i>Drug Metab. Dispos.</i> 2004 , 32(3), 291-294	14977862	1 10	6	> 50% is active (based on the threshold from De Bruyn, T.; van Westen, G.J.P., IJzerman, A.P., Stieger, B., de Witte, P., Augustijns, P.F., Annaert, P.P. Structure-Based Identification of OATP1B1/3 Inhibitors. <i>Mol. Pharmacol.</i> 2013 , 83 (6), 1257-1267.)
Nozawa, T.; Tamai, I.; Sai, Y.; Nezu, J. I.; Tsuji, A. Contribution of organic anion transporting polypeptide OATP-C to hepatic elimination of the opioid pentapeptide analogue [d-Ala 2, d-Leu5]-enkephalin. <i>J. Pharm. Pharmacol.</i> 2003 , 55(7), 1013-1020	12906759	1000 5000	3	REMOVED
König, J.; Cui, Y.; Nies, A. T.; Keppler, D. A novel human organic anion transporting polypeptide localized to the basolateral hepatocyte membrane. <i>Am. J. Physiol. Gastrointest. Liver Physiol.</i> 2000 , 278(1): G156-G164	10644574	50 1000	2	REMOVED
Satoh, H.; Yamashita, F.; Tsujimoto, M.; Murakami, H.; Koyabu, N.; Ohtani, H.; Sawada, Y. Citrus juices inhibit the function of human organic anion transporting polypeptide OATP-B. Drug metabolism and disposition. <i>Drug Metab. Dispos.</i> 2005 , 33(1): 518	15640378	1 10	4	> 50% is active (based on the threshold from De Bruyn, T.; van Westen, G.J.P., IJzerman, A.P., Stieger, B., de Witte, P., Augustijns, P.F., Annaert, P.P. Structure-Based Identification of OATP1B1/3 Inhibitors. <i>Mol. Pharmacol.</i> 2013 , 83 (6), 1257-1267.)

Table S2. Data sets annotated with bioactivity end point “inhibition” from Metrabase¹: Reference to the original manuscript, PMID, compound concentration used in the experiment, number of unique compounds (after standardization procedure) and recommended activity thresholds are given.

Reference	PMID	c [uM]	Cmpd. number	Recommended threshold
Karlgren, M.; Vildhede, A.; Norinder, U.; Wisniewski, J. R.; Kimoto, E.; Lai, Y.; Haglund, U.; Artursson, P. Classification of Inhibitors of Hepatic Organic Anion Transporting Polypeptides (OATPs): Influence of Protein Expression on Drug–Drug Interactions. <i>J. Med. Chem.</i> 2012 , <i>55</i> (10), 4740–4763.	22541068	20	5	> 50% is active
Sai, Y.; Kaneko, Y.; Ito, S.; Mitsuoka, K.; Kato, Y.; Tamai, I.; Artursson, P.; Tsuji, A. Predominant contribution of organic anion transporting polypeptide OATP-B (OATP2B1) to apical uptake of estrone-3-sulfate by human intestinal Caco-2 cells. <i>Drug Metab Dispos.</i> 2006 , <i>34</i> (8):1423-31	16714376	1000 10000	2	REMOVED

¹ This table does not include ~150 data sets from Metrabase where there were additional sources available (e.g. from ChEMBL).

Fuchikami, H.; Satoh, H.; Tsujimoto, M.; Ohdo, S.; Ohtani, H.; Sawada, Y. Effects of Herbal Extracts on the Function of Human Organic Anion Transporting Polypeptide, OATP-B. <i>Drug Metab Dispos.</i> 2006 , 34(4):577-82	16415120	1 10 100	4	> 60% is active (based on the threshold from Marada, V.V.; Flörl, S.; Kühne, A.; Burckhardt, G.; Hagos, Y. Interaction of human organic anion transporter polypeptides 1B1 and 1B3 with antineoplastic compounds. <i>Eur. J. Med. Chem.</i> 2015 , 92, 723-731.)
Kis, O.; Zastre, J. A.; Ramaswamy, M.; Bendayan, R. pH dependence of organic anion-transporting polypeptide 2B1 in Caco-2 cells: potential role in antiretroviral drug oral bioavailability and drug–drug interactions. <i>J Pharmacol Exp Ther.</i> 2010 , 334(3):1009-22	20507927	100	4	> 60% is active (based on the threshold from Marada, V.V.; Flörl, S.; Kühne, A.; Burckhardt, G.; Hagos, Y. Interaction of human organic anion transporter polypeptides 1B1 and 1B3 with antineoplastic compounds. <i>Eur. J. Med. Chem.</i> 2015 , 92, 723-731.)
Grube, M.; Köck, K.; Oswald, S.; Draber, K.; Meissner, K.; Eckel, L.; Böhm, M.; Felix, S. B.; Vogelgesang, S.; Jedlitschky, G.; Siegmund, W; Warzok, R.; Kroemer, H. K. Organic anion transporting polypeptide 2B1 is a high-affinity transporter for atorvastatin and is expressed in the human heart. <i>Clin Pharmacol Ther.</i> 2006 , 80(6):607-20	17178262	10	1	> 50% is active

<p>Grube, M.; Kock, K.; Karner, S.; Reuther, S.; Ritter, C. A.; Jedlitschky, G.; Kroemer, H. K. Modification of OATP2B1 mediated transport by steroid hormones. <i>Molecular pharmacology. Mol Pharmacol.</i> 2006, 70(5):1735-41.</p>	16908597	10	2	<p>> 50% is active</p> <p>(based on the threshold from De Bruyn, T.; van Westen, G.J.P., IJzerman, A.P., Stieger, B., de Witte, P., Augustijns, P.F., Annaert, P.P. Structure-Based Identification of OATP1B1/3 Inhibitors. <i>Mol. Pharmacol.</i> 2013, 83 (6), 1257-1267.)</p>
<p>Reyes, M.; Benet, L. Z. Effects of uremic toxins on transport and metabolism of different biopharmaceutics drug disposition classification system xenobiotics. <i>J Pharm Sci.</i> 2011, 100(9):3831-42</p>	21618544	400	2	REMOVED

Table S3. Removed substrates from OATP1B1 substrate data set with conflicting annotations (median value equals to 0.5). Compound names and InChIKeys are provided.

Name	InChIKey
Methylaminopterin	FBOZXECLQNJBKD-UHFFFAOYSA-N
No name	HMBKEXWSKGGBBT-WCHIAOBISA-N
3,7-Dihydroxycholan-24-oic acid	RUDATBOHQWOJDD-UHFFFAOYSA-N
(8S,9S,13S,14S)-17-ethynyl-17-hydroxy-13-methyl-3-sulfoxy-7,8,9,11,12,14,15,16-octahydro-6H-cyclopenta[a]phenanthrene	WLGIWVFFGMPRLM-UHFFFAOYSA-N

Table S4. Removed substrates from OATP1B3 substrate data set with conflicting annotations (median value equals to 0.5). Compound names and InChIKeys are provided.

Name	InChIKey
Caloxetic acid	AQOXEJNYXXLRQQ-UHFFFAOYSA-N

Table S5. Removed substrates from OATP2B1 substrate data set with conflicting annotations (median value equals to 0.5). Compound names and InChIKeys are provided.

Name	InChIKey
Breviscapine	DJSISFGPUUYILV- UHFFFAOYSA-N
Ethyldiisopropylamine	FUWQJNSUHNKFNP- UHFFFAOYSA-N
Bosentan	GJPICJJJRGTNOD- UHFFFAOYSA-N
6- {[5,7-dihydroxy-2-(4-hydroxyphenyl)-4-oxo-4H-chromen-6-yl]oxy}-3,4,5-trihydroxyoxane-2-carboxylic acid	HBLWMMBFOKSEKW- UHFFFAOYSA-N
Talinolol	MXFWWQICDIZSOA- UHFFFAOYSA-N
No name	QQCSUWGQBREWRO- CYVLTUHYSAN
2,2',4,4'-Tetrabromodiphenyl ether	XYBSIYMGXVUVGY- UHFFFAOYSA-N

Table S6. Removed inhibitors from OATP1B1 inhibitor data set with conflicting annotations (median value equals to 0.5). Compound names and InChIKeys are provided.

Name	InChIKey
Lanosteryl acetate	BQPPJGMMIYJVBR-UHFFFAOYSA-N
Capsazepine	DRCMAZOSEIMCHM-UHFFFAOYSA-N
Zopiclone	GBBSUAFBMRNDJC-UHFFFAOYSA-N
3-Episarsasapogenin	GMBQZIIUCVWOCU-UHFFFAOYSA-N
Quinine	LOUPRKONTZGTKE-UHFFFAOYSA-N
Spirolactone	LXMSZDCAJNLERA-UHFFFAOYSA-N
Estradiol-3-sulfate	QZIGLSSUDXBTLJ-UHFFFAOYSA-N
Taurocholic acid	WBWWGRHZICKQGZ-UHFFFAOYSA-N
10-acetyloxy-1,2,6a,6b,9,9,12a-heptamethyl-13-oxo-1,2,3,4,5,6,6a,7,8,8a,10,11,12,14b-tetradecahydronicene-4a-carboxylic acid	XDHCWTUZCOFKRH-UHFFFAOYSA-N
Eltrombopag	XDXWLKQMMKQXPV-QYQHSDDTDSA-N
Cephalothin	XIURVHNZVLADCM-UHFFFAOYSA-N

Table S7. Removed inhibitors from OATP1B3 inhibitor data set with conflicting annotations (median value equals to 0.5). Compound names and InChIKeys are provided.

Name	InChIKey
Lanosteryl acetate	BQPPJGMMIYJVBR-UHFFFAOYSA-N
Capsazepine	DRCMAZOSEIMCHM-UHFFFAOYSA-N
Zopiclone	GBBSUAFBMRNDJC-UHFFFAOYSA-N
17beta-estradiol 17beta-D-glucuronide	MTKNDAQYHASLID-UHFFFAOYSA-N
Vincristine	OGWKCGZFUXNPDA-UHFFFAOYSA-N
Hoechst 33342	PRDFBSVERLRMY-UHFFFAOYSA-N
(1-{{(1-{{3-[(E)-2-(7-chloroquinolin-2-yl)ethenyl]phenyl}}-3-[2-(2-hydroxypropan-2-yl)phenyl]propyl)sulfanyl)methyl}cyclopropyl)acetic acid	UCHDWCPVSPXUMX-XNTDXEJSSA-N
No name	UWNQSONNONTGTF-UHFFFAOYSA-N
Nefazodone	VRBKIVRKKCLPHA-UHFFFAOYSA-N
No name	WTDQOIVETOYZMZ-UHFFFAOYSA-N
Antamanide	WTINJQXJTHUFRF-UHFFFAOYSA-N
10-acetyloxy-1,2,6a,6b,9,9,12a-heptamethyl-13-oxo-1,2,3,4,5,6,6a,7,8,8a,10,11,12,14b-tetradecahydronicene-4a-carboxylic acid	XDHCWTUZCOFKRH-UHFFFAOYSA-N
Levothyroxine	XUIIKFGFIJCVMT-UHFFFAOYSA-N

Table S8. Removed inhibitors from OATP1B3 inhibitor data set with conflicting annotations (median value equals to 0.5). Compound names and InChIKeys are provided.

Name	InChIKey
Dehydroepiandrosterone sulfate	CZWCKYRVOZZJNM-UHFFFAOYSA-N
Estrone-d4	DNXHEGUUPJUMQT-UHFFFAOYSA-N
Repaglinide	FAEKWTJYAYMJKF-UHFFFAOYSA-N
6',7'-dihydroxy Bergamottin	IXZUPBUEKFXTSD-MDWZMJQESA-N
Epicatechin-3-gallate	LSHVYAFMTMFKBA-UHFFFAOYSA-N
Nobiletin	MRIAQLRQZPPODS-UHFFFAOYSA-N
Nelfinavir	QAGYKUNXZHKKMR-UHFFFAOYSA-N
Quercetin	REFJWTPEDEVJJIY-UHFFFAOYSA-N
Tangeretin	ULSUXBXHSYSGDT-UHFFFAOYSA-N
Efavirenz	XPOQHMRABVBWPR-UHFFFAOYSA-N
Rosiglitazone	YASAKCUCGLMORW-UHFFFAOYSA-N

Table S9. Percentage of conflicting compound activities based on the comparison of data from ChEMBL and Metrabase.

Activity	OATP1B1	OATP1B3	OATP2B1
(non-)substrates	40%	36%	60%
(non-)inhibitors	64%	74%	10%

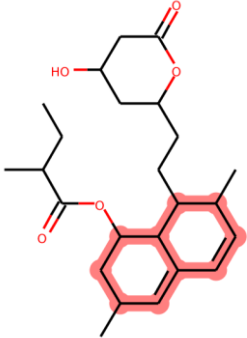
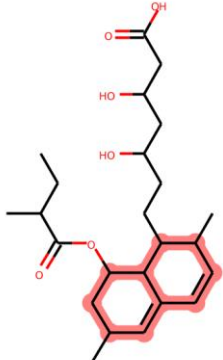
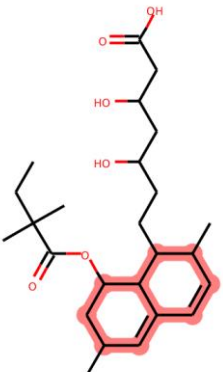
Table S10. “Dense dataset” for hepatic OATP substrates.

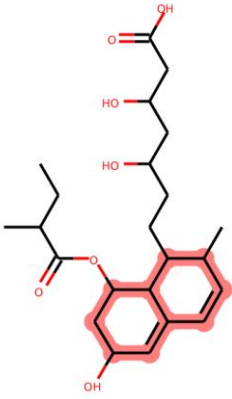
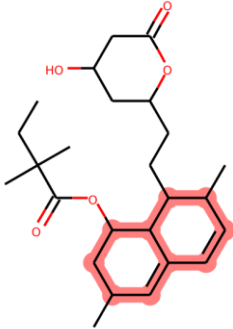
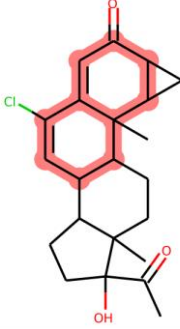
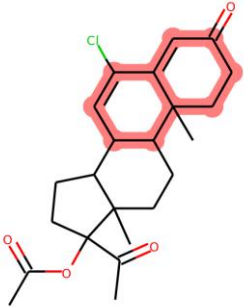
Class	Nr.	Description
0	6	General substrates
1	0	1B1+1B3 overlapping substrates
2	2	1B1+2B1 overlapping substrates
3	0	1B3+2B1 overlapping substrates
4	0	Selective 1B1 substrates
5	2	Selective 1B3 substrates
6	2	Selective 2B1 substrates
7	1	General non-substrates

Table S11. “Dense dataset” for hepatic OATP inhibitors.

Class	Nr.	Description
0	26	General inhibitors
1	14	1B1+1B3 overlapping inhibitors
2	10	1B1+2B1 overlapping inhibitors
3	1	1B3+2B1 overlapping inhibitors
4	12	Selective 1B1 inhibitors
5	2	Selective 1B3 inhibitors
6	10	Selective 2B1 inhibitors
7	88	General non- inhibitors

Table S12. Ten detected compounds with hexahydronaphthalene-associated scaffold (highlighted in red) with pharmacological profiles included: “1”....active; “0”....inactive; “?”missing annotation.

Name	Structure	OATP1B1	OATP1B3	OATP2B1
Mevinolin		1	0	0
Lovastatin Acid		1	?	?
Simvastatin Acid		1	0	?

<p>Pravastatin</p>	 <p>The structure shows a statin core consisting of a naphthalene ring system with a hydroxyl group at the 3-position and a decalin ring system at the 4-position. The decalin ring has a hydroxyl group at the 1-position and a side chain at the 2-position. The side chain is a 3,5-dihydroxyhexanoic acid derivative, with a hydroxyl group at the 3-position and a carboxylic acid group at the 6-position.</p>	<p>1</p>	<p>0</p>	<p>0</p>
<p>Simvastatin</p>	 <p>The structure shows a statin core consisting of a naphthalene ring system with a hydroxyl group at the 3-position and a decalin ring system at the 4-position. The decalin ring has a hydroxyl group at the 1-position and a side chain at the 2-position. The side chain is a 2-methylbutanoic acid derivative, with a methyl group at the 2-position and a carboxylic acid group at the 1-position.</p>	<p>1</p>	<p>0</p>	<p>0</p>
<p>Cyprotenone</p>	 <p>The structure shows a complex polycyclic system with a naphthalene ring system fused to a decalin ring system. The naphthalene ring has a chlorine atom at the 1-position and a carbonyl group at the 2-position. The decalin ring has a hydroxyl group at the 1-position and a carbonyl group at the 2-position.</p>	<p>1</p>	<p>?</p>	<p>?</p>
<p>Chlormadinone Acetate</p>	 <p>The structure shows a complex polycyclic system with a naphthalene ring system fused to a decalin ring system. The naphthalene ring has a chlorine atom at the 1-position and a carbonyl group at the 2-position. The decalin ring has a hydroxyl group at the 1-position and a carbonyl group at the 2-position. The hydroxyl group is esterified with an acetate group.</p>	<p>1</p>	<p>1</p>	<p>?</p>

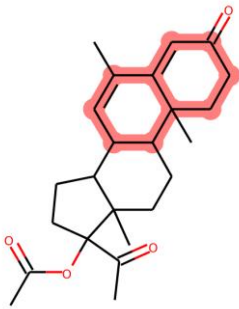
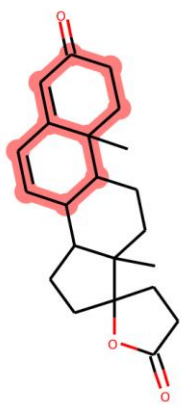
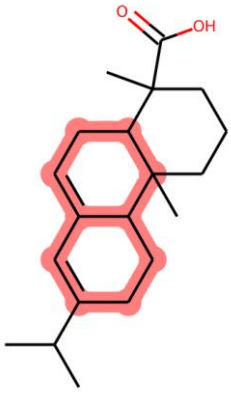
Megestrol Acetate	 <p>The structure shows a steroid nucleus with a phenyl ring at C-3, a methyl group at C-10, and an acetate ester group at C-20. The phenyl ring and its substituents are highlighted in red.</p>	0	0	?
Canrenone	 <p>The structure shows a steroid nucleus with a phenyl ring at C-3, a methyl group at C-10, and a lactone ring at C-20. The phenyl ring and its substituents are highlighted in red.</p>	1	0	?
Abietic Acid	 <p>The structure shows a tricyclic diterpene skeleton with a carboxylic acid group at C-18 and a methyl group at C-19. The aromatic rings and their substituents are highlighted in red.</p>	1	1	?

Table S13: Results on Level 1 (All inhibitors + general non-inhibitors) for all calculated statistics metrics: Sensitivity, Specificity, Balanced Accuracy and MCC. The performance is given for both 10-fold cross-validation and on the external test set. With bold font are depicted those models that gave the best results.

Model	Validation	Sensitivity	Specificity	Balanced Accuracy	MCC	Descriptor (% involvement in subset selection of attributes)
RandomTree	Training set 10fold-CV	0.893	0.339	0.616	0.255	SlogP, AMW
RandomTree	Test Set	0.805	0.346	0.576	0.142	
CostSensitive Classifier	Training set 10fold-CV	0.889	0.355	0.622	0.264	
CostSensitive Classifier	Test Set	0.788	0.462	0.625	0.222	
Stratified bagging	Training set	0.760	0.790	0.775	0.455	
Stratified bagging	Test Set	0.796	0.769	0.783	0.477	

Level 1: All inhibitors + general non-inhibitors (Training set)

CostSensitive Classifier: Cost matrix [0.0, 4.0; 1.0, 0.0]

Stratified Bagging with 64 bags

Table S14: Results on OPATP1B1 inhibition data set for all calculated statistics metrics: Sensitivity, Specificity, Balanced Accuracy and MCC. The performance is given for both 10-fold cross-validation and on the external test set. With bold font are depicted those models that gave the best results.

Model	Validation	Sensitivity	Specificity	Balanced Accuracy	MCC	Descriptor (% involvement in subset selection of attributes)
RandomTree	Training set 10fold-CV	0.547	0.830	0.689	0.370	SlogP, SMR, AMW, TPSA
RandomTree	Test Set	0.580	0.828	0.704	0.396	
CostSensitive Classifier	Training set 10fold-CV	0.539	0.817	0.678	0.345	
CostSensitive Classifier	Test Set	0.540	0.802	0.671	0.328	
Stratified bagging	Training set	0.703	0.799	0.751	0.462	
Stratified bagging	Test Set	0.730	0.809	0.769	0.497	

CostSensitive Classifier: Cost matrix [0.0, 1.0; 3.0, 0.0]

Stratified Bagging with 64 bags

Table S15: Results on OPATP1B3 inhibition data set for all calculated statistics metrics: Sensitivity, Specificity, Balanced Accuracy and MCC. The performance is given for both 10-fold cross-validation and on the external test set. With bold font are depicted those models that gave the best results.

Model	Validation	Sensitivity	Specificity	Balanced Accuracy	MCC	Descriptor (% involvement in subset selection of attributes)
RandomTree	Training set 10fold-CV	0.496	0.895	0.696	0.384	SlogP, SMR, AMW, TPSA
RandomTree	Test Set	0.373	0.899	0.636	0.282	
CostSensitive Classifier	Training set 10fold-CV	0.504	0.891	0.697	0.383	
CostSensitive Classifier	Test Set	0.424	0.892	0.658	0.316	
Stratified bagging	Training set	0.748	0.834	0.791	0.486	
Stratified bagging	Test Set	0.746	0.829	0.787	0.476	

CostSensitive Classifier: Cost matrix [0.0, 1.0; 5.0, 0.0]

Stratified Bagging with 64 bags

Table S16: Results on OPATP2B1 inhibition data set for all calculated statistics metrics: Sensitivity, Specificity, Balanced Accuracy and MCC. The performance is given for both 10-fold cross-validation and on the external test set. With bold font are depicted those models that gave the best results.

Model	Validation	Sensitivity	Specificity	Balanced Accuracy	MCC	Descriptor (% involvement in subset selection of attributes)
RandomTree	Training set 10fold-CV	0.535	0.856	0.695	0.400	SMR, AMW
RandomTree	Test Set	0.526	0.840	0.683	0.373	
CostSensitive Classifier	Training set 10fold-CV	0.535	0.839	0.687	0.377	
CostSensitive Classifier	Test Set	0.526	0.860	0.693	0.400	
Stratified bagging	Training set	0.698	0.771	0.734	0.434	
Stratified bagging	Test Set	0.632	0.840	0.736	0.464	

CostSensitive Classifier: Cost matrix [0.0, 1.0; 3.0, 0.0]

Stratified Bagging with 64 bags

Table S17: Summary statistics for molecular descriptors calculated for inhibitors of OATP1B1, OATP1B3, and OATP2B1. The median and mean values for lipophilicity (SlogP), molecular refractivity (SMR), the topological polar surface area (TPSA), average molecular weight (AMW), the number of rotatable bonds (RotB), the number of amide bonds (AmideBonds), the number of rings (NumRings), and the number of aromatic carbocycles (Aromatic Carbocycles) are given.

	SlogP(median)	SlogP(mean)	SMR(median)	SMR(mean)	TPSA(median)	TPSA(mean)	AMW(median)	AMW(mean)
OATP1B1 inhibitors	3.99	3.88	123.70	141.42	106.86	126.30	471.64	537.04
OATP1B1 non-inhibitors	2.24	2.20	81.58	86.91	66.76	81.31	300.27	327.92
OATP1B3 inhibitors	4.22	4.21	132.00	148.31	113.08	133.75	504.67	564.47
OATP1B3 non-inhibitors	2.33	2.26	82.39	87.05	68.22	81.78	303.60	330.61
OATP2B1 inhibitors	3.64	3.71	127.47	131.36	109.33	120.63	481.99	503.34
OATP2B1 non-inhibitors	2.48	2.22	95.40	104.36	84.83	99.87	358.16	394.43
	RotB(median)	RotB(mean)	AmideBonds (median)	AmideBonds (mean)	NumRings (median)	NumRings (mean)	AromaticCarbo-cycles(median)	AromaticCarbo-cycles(mean)
OATP1B1 inhibitors	5	6.56	0	1.03	4	4.03	1	1.10
OATP1B1 non-inhibitors	3	3.96	0	0.34	3	2.77	1	0.92
OATP1B3 inhibitors	6	6.90	0	0.80	4	4.37	1	1.21
OATP1B3 non-inhibitors	3	3.87	0	0.35	3	2.79	1	0.92
OATP2B1 inhibitors	7	7.52	0	0.74	4	3.74	2	1.76
OATP2B1 non-inhibitors	4	4.99	0	0.68	3	3.18	1	1.20

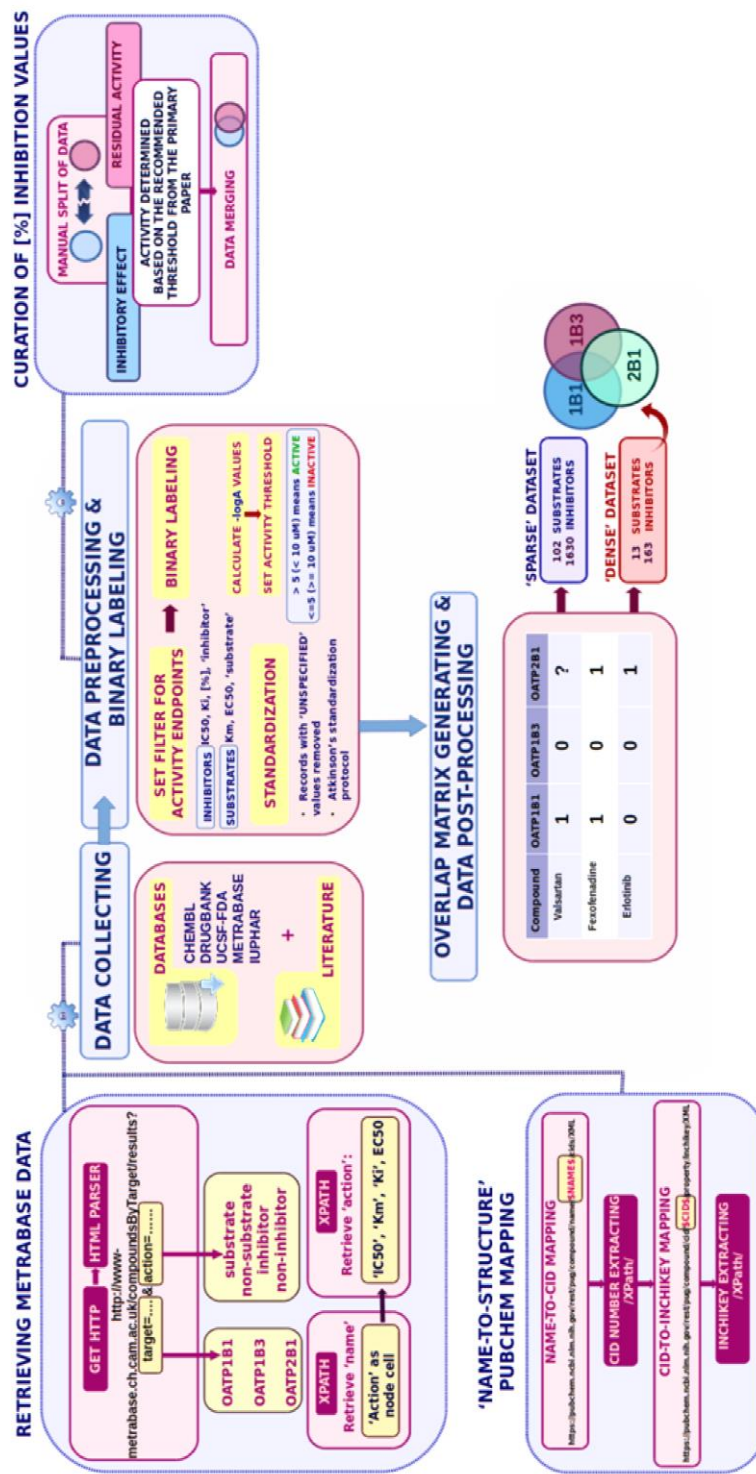


Figure S1: Schematic workflow for integrative data mining and curation.

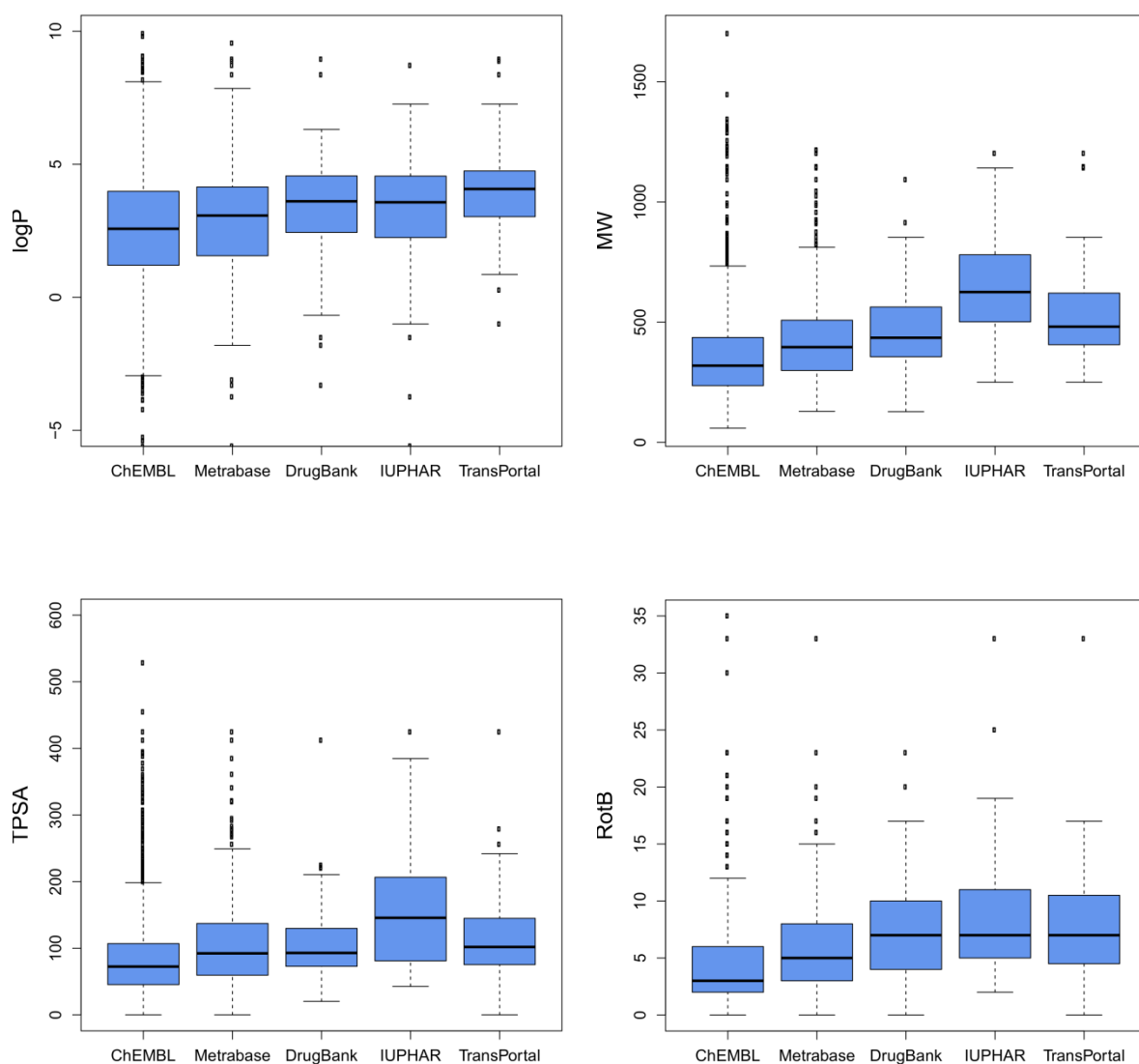


Figure S2. Box- and Whisker-Plots showing the distribution of molecular properties for compounds measured against human OATP1B1, OATP1B3, and OATP2B1 originating from five different data sources (ChEMBL, Metrabase, DrugBank, IUPHAR, TransPortal): partition coefficient (logP), molecular weight (MW), topological surface area (TPSA), number of rotatable bonds (RotB).

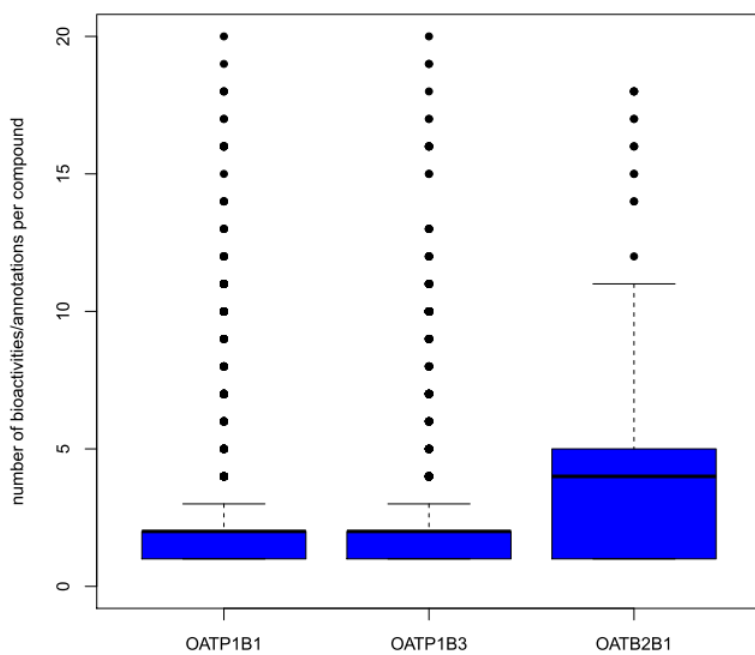
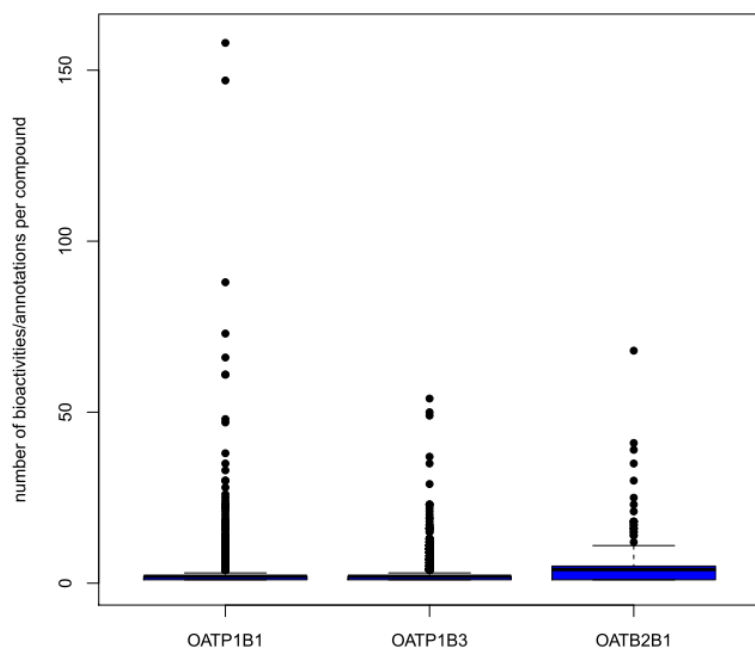
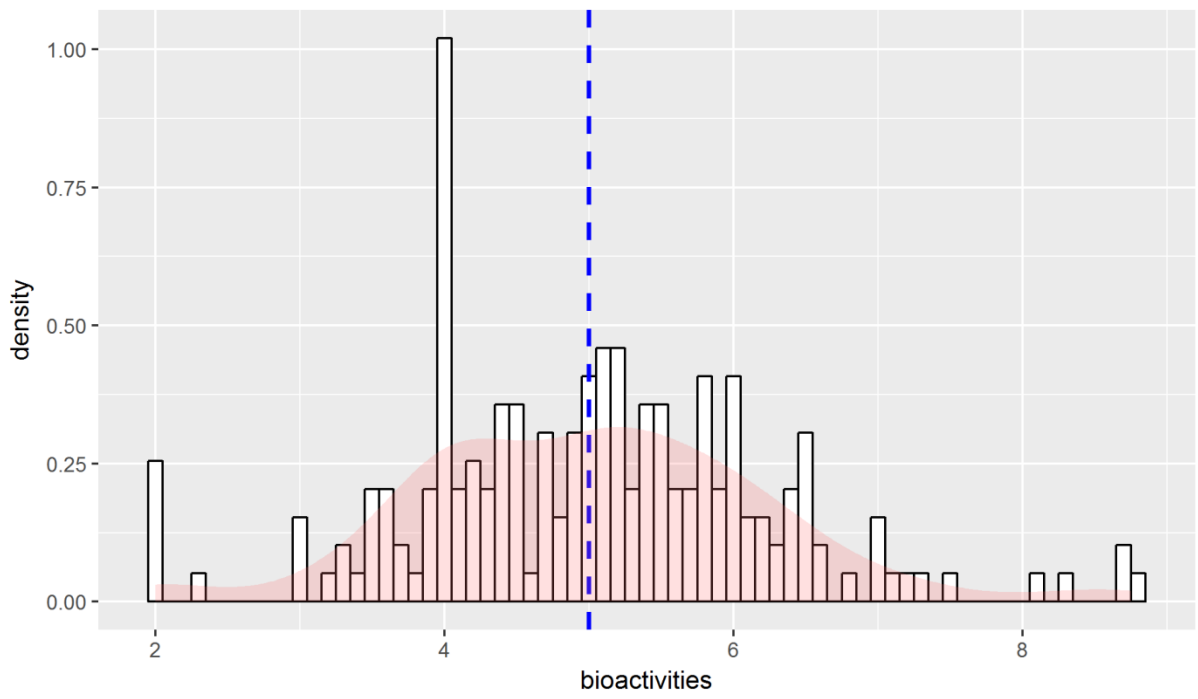
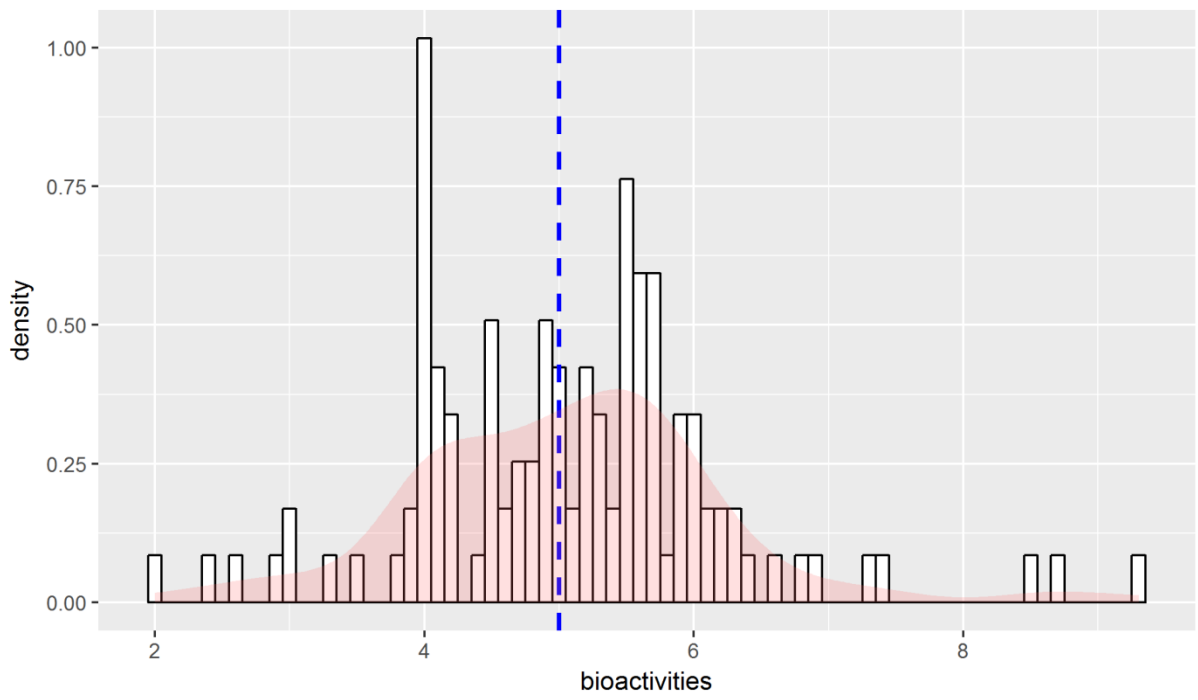


Figure S3. Number of bioactivities/annotations per unique compound for hepatic OATPs: upper plot....full range of bioactivity values displayed; lower plot....zoomed-in view with max. 20 bioactivities displayed.

Density plot: OATP1B1 activities



Density plot: OATP1B3 activities



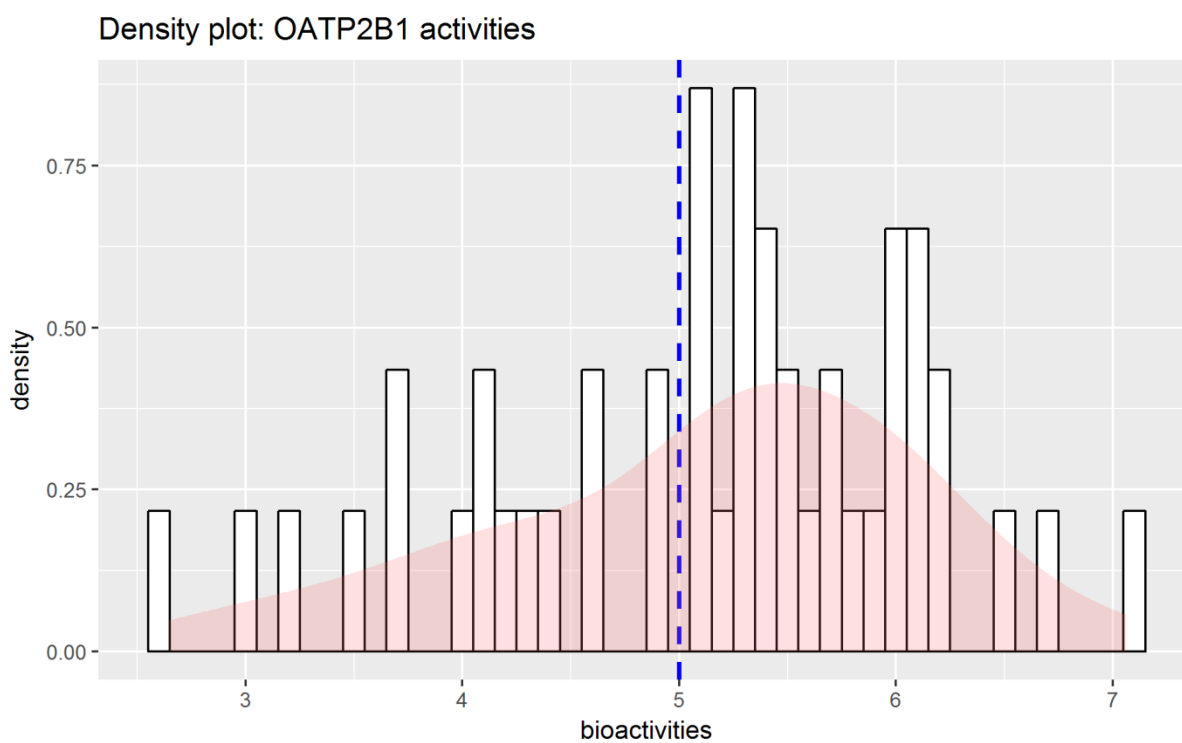


Figure S4. Histograms showing the distribution of median bioactivities [in negative logarithmic representation (molar)] for OATP1B1, OATP1B3, and OATP2B1. The chosen cut off for classifying compounds into actives and inactive ($10\mu\text{M}$; $-\log(\text{activity}[\text{molar}]) = 5$) is shown as a dashed blue line.

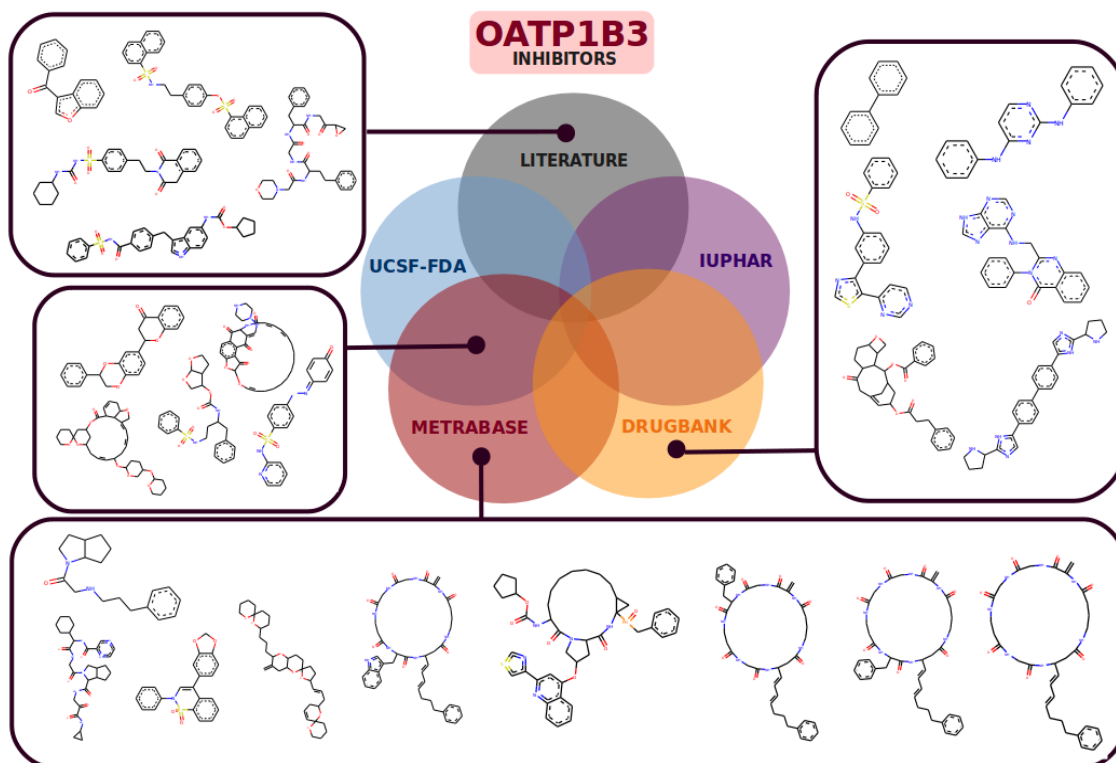


Figure S5. Murcko Scaffolds for OATP1B3 inhibitors retrieved from other databases than CHEMBL.

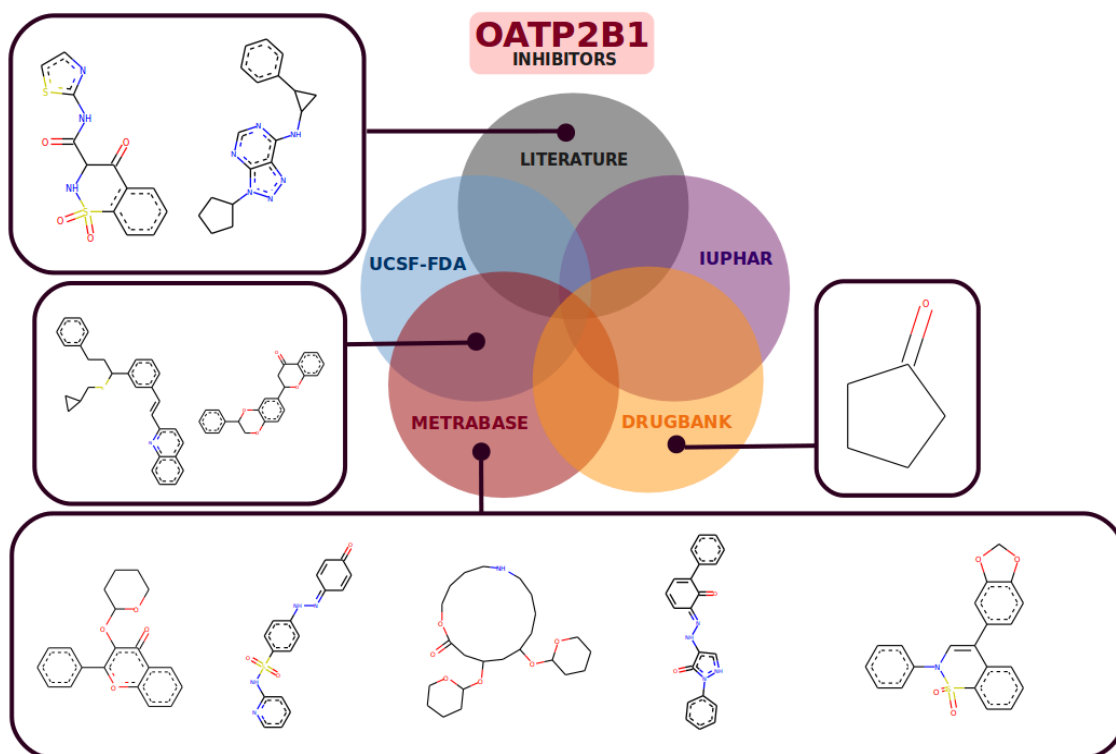


Figure S6. Murcko Scaffolds for OATP2B1 inhibitors retrieved from other databases than CHEMBL.

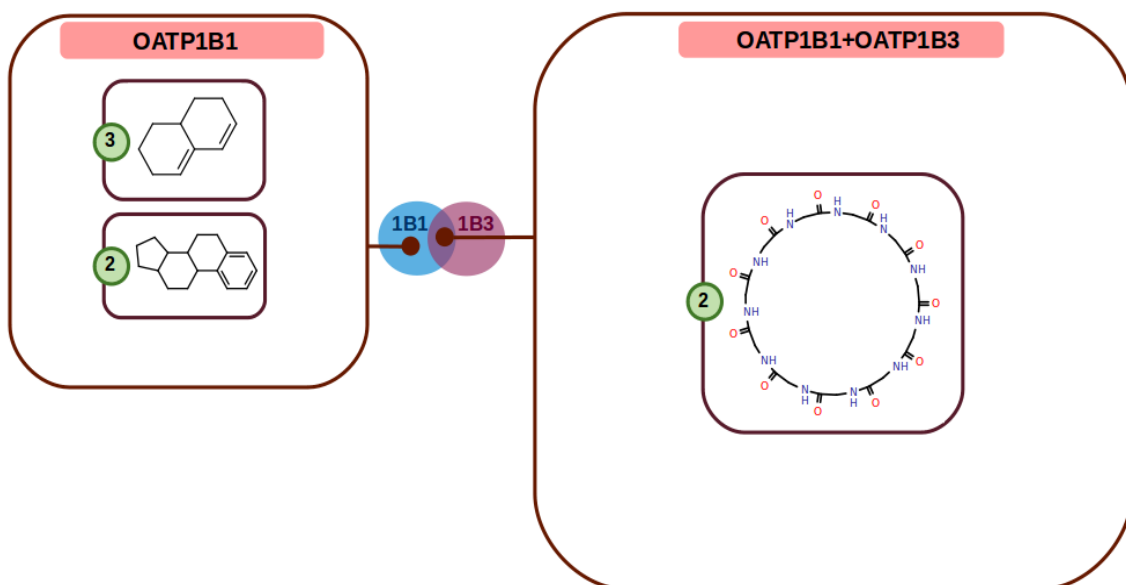


Figure S7. Enriched scaffolds (p -value < 0.05) for hepatic OATP inhibitors considering the dense data set (with complete pharmacological profile). Numbers in green circles correspond to the numbers of associated compounds for a respective scaffold cluster.

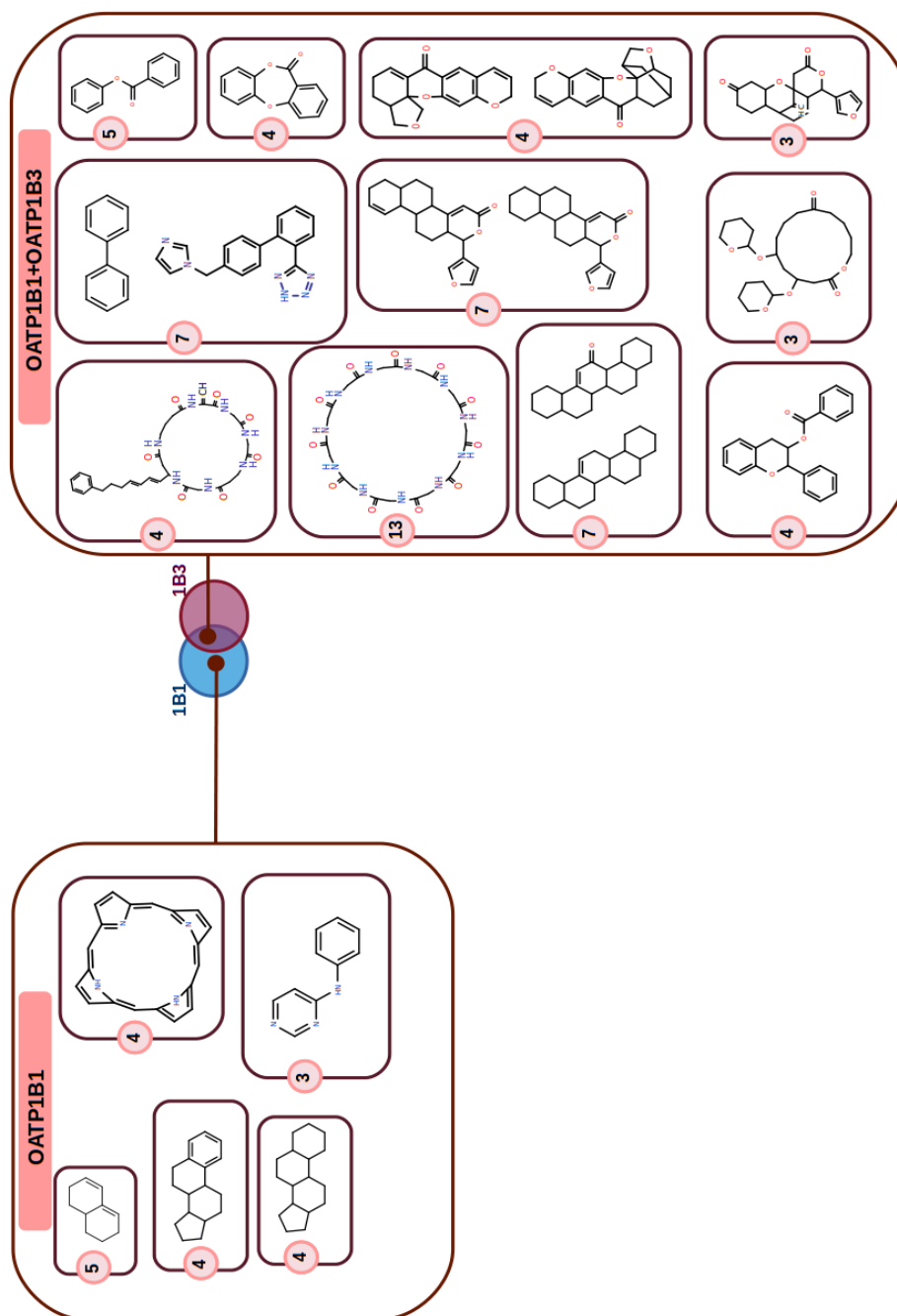


Figure S8. Enriched scaffolds (p -value < 0.05) for hepatic OATP inhibitors excluding data with the activity endpoint “percentage inhibition”. Numbers in pink circles correspond to the number of associated compounds for a respective scaffold cluster.

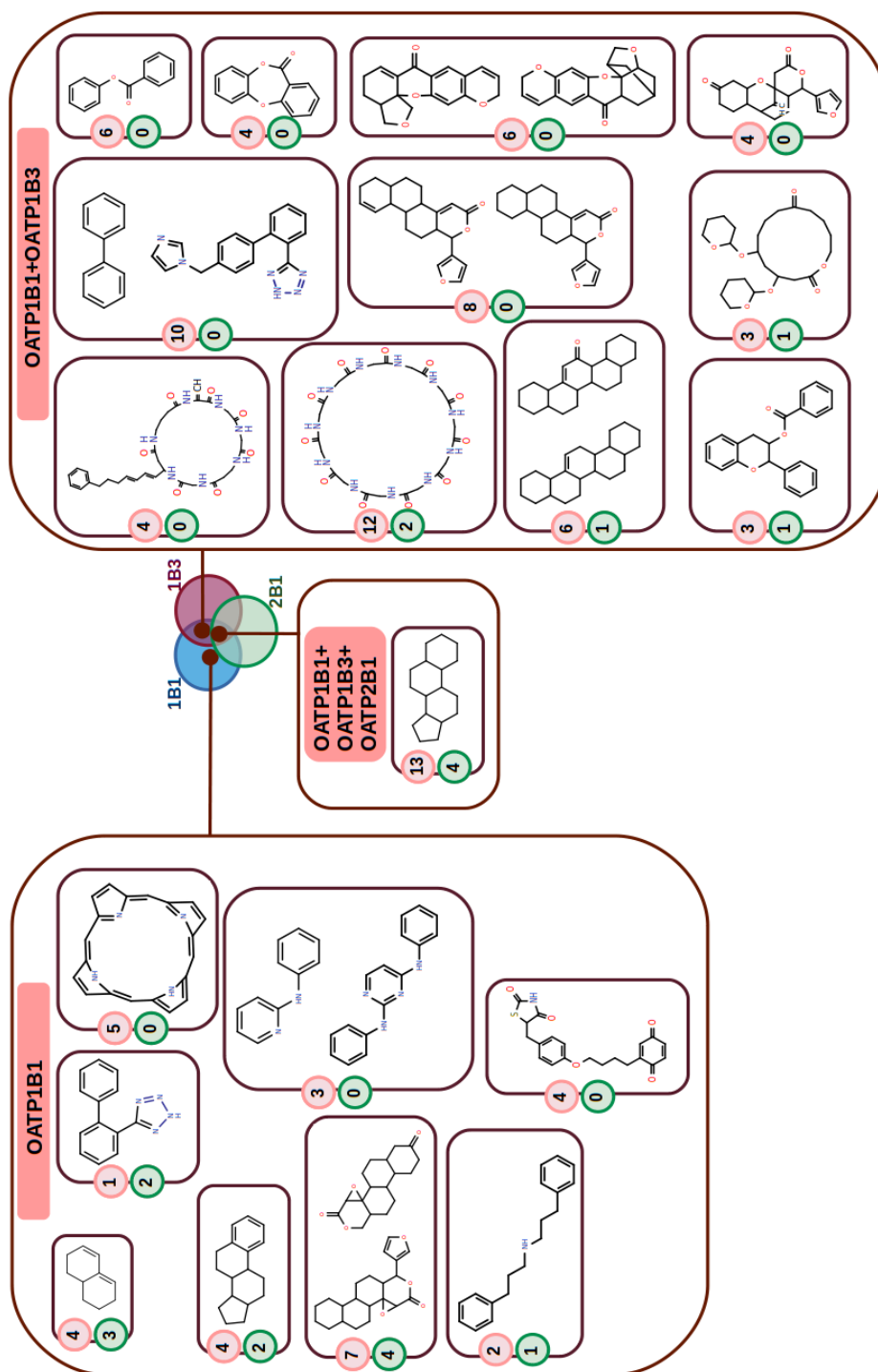


Figure S9. Enriched scaffolds (p -value < 0.1) for hepatic OATP inhibitors. Numbers in pink circles correspond to the number of associated compounds with incomplete pharmacological profile, whereas numbers in green circles correspond to the number of associated compounds with complete pharmacological profile (i.e., compounds from “dense” dataset).

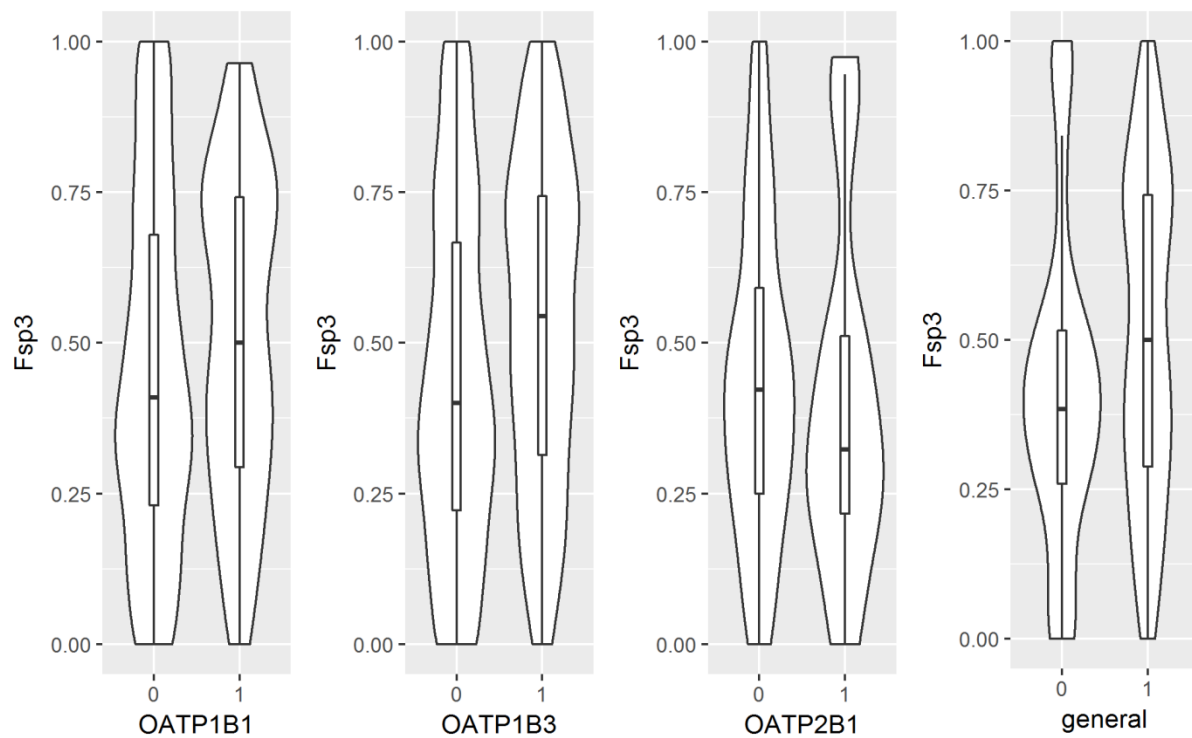


Figure S10. Violin- and boxplots showing the distribution of values for the feature “FractionCSP3” (Fsp3) for inhibitors vs non-inhibitors within four different data sets. Labelling on abscissae: 0....inactives; 1....actives.

Description of Supplementary Data Files:

Data File S1. Csv-file with sparse substrate data set (102 compounds): including CHEMBL_ID (if available), molecule name (if available), InChIKey, canonical SMILES, binary annotations (“1”....active; “0”....inactive; “?”....missing value), and median [-log[activity)] values (if available).

Data File S2. Csv-file with sparse inhibitor data set (1630 compounds): including CHEMBL_ID (if available), molecule name (if available), InChIKey, canonical SMILES, binary annotations (“1”....active; “0”....inactive; “?”....missing value), and median [-log[activity)] values (if available).