http://hdl.handle.net/1765/125549

Erasmus MC
Universitair Medisch Centrum Rotterdam

Alternative splicing of the *SLCO1B1* gene: an exploratory analysis of isoform diversity in pediatric liver

Bianca D van Groen, Chengpeng Bi, Roger Gaedigk, Vincent S Staggs, Dick Tibboel, Saskia N de Wildt, J Steven Leeder

Clin Transl Sci. 2020 Jan 9. DOI: 10.1111/cts.12733.

Erasmus University Rotterdam

ABSTRACT

The hepatic influx transporter OATP1B1 (SLCO1B1) plays an important role in the disposition of endogenous substrates and drugs prescribed to children. Alternative splicing increases the diversity of protein products from >90% of human genes and may be triggered by developmental signals. As concentrations of several endogenous OATP1B1 substrates change during growth and development, with this exploratory study we investigated age-dependent alternative splicing of SLCO1B1 mRNA in 97 post-mortem livers (fetus-adolescents). Twenty-seven splice variants were detected; ten were confirmed by additional bioinformatic analyses and verified by qPCR, and selected for detailed analysis based on relative abundance, association with age and overlap with an adjacent gene. Two splice variants code for reference OATP1B1 protein, and eight code for truncated proteins. The expression of eight isoforms was associated with age. We conclude that alternative splicing of SLCO1B1 occurs frequently in children; although the functional consequences remain unknown, the data raise the possibility of a regulatory role for alternative splicing in mediating developmental changes in drug disposition.

INTRODUCTION

Transporters are membrane-bound proteins that are present in the apical and basolateral membranes of organs, such as the liver.[1] Their biological role is the trafficking of substrates across membranes, making them critical determinants of tissue and cellular substrate disposition. Moreover, they act in concert with drug-metabolizing enzymes (DMEs) to maintain homeostatic balance for endogenous substrates and to facilitate the detoxification and elimination of exogenous substrates, such as drugs and food toxins.[2]

This latter is important for newborns, as after birth they become dependent on exogenous food sources for nutrition, and the diet expands as they grow into infanthood. During all these changes in food exposure, the child must defend itself against potentially toxic dietary constituents, recruiting pathways not or differentially expressed during fetal life. Hence ontogeny of DMEs and transporters occurs, influencing the disposition of their endogenous and exogenous substrates over age. Moreover, ontogeny may well be driven by developmental homeostatic changes in endogenous substrates.[3]

A classic example of age-related changes in a DME that plays an important physiological role is hepatic cytochrome P450 (CYP) 3A7. CYP3A7 is highly expressed in fetal liver but steadily declines throughout the last trimester of pregnancy and first year of postnatal life to low levels characteristic of adult liver.[4,5] From a functional perspective, CYP3A7 catalyzes the 16α-hydroxylation of dehydroepiandosteron 3-sulfate (DHEA-S), to form 16α-DHEA-S, a precursor for estriol synthesis by placental syncytiotrophoblasts.[6] DHEA-S concentrations are high during the fetal and neonatal periods and decline postnatally.[7,8] DHEA-S has been reported to activate CYP3A7 activity, which explains the high expression of CYP3A7 in fetal liver.[9] DHEA-S also provides an example of the interplay between DMEs and transporters in a developmental context as prior to biotransformation by CYP3A7 in fetal liver, DHEA-S needs to cross the hepatocyte membrane, using the solute carrier organic anion transporter (gene name *SLCO1B1*, protein name OATP1B1) located on the basolateral membrane.[10] Consistent with CYP3A7, the OATP1B1 expression also has been demonstrated to decline directly after birth[11], followed by age-dependent increases in mRNA expression throughout childhood.[12] Data are conflicting regarding developmental patterns of OATP1B1 protein, with expression increased around age eight years, compared to younger children in one study, using immunoblotting techniques[13], and no apparent statistically significant relationship with age, using a quantitative proteomic approach.[14]

Whereas the contribution of CYP3A7 to drug clearance postnatally is relatively minor, OATP1B1 plays an important role in the clearance of potentially toxic endogenous

molecules. One example illustrating the importance of transporter function early after birth involves bilirubin; an association has been demonstrated between the *SLCO1B1* 388G>A allele, a variant associated with reduced function of the transporter, and unconjugated hyperbilirubinemia in newborns, which is associated with neurotoxicity.[15,16] Moreover, OATP1B1 is not only involved in the disposition of endogenous substrates but also of drugs that are used in pediatrics, such as statins, methotrexate and bosentan.[10] Malfunctioning of the OATP1B1 transporter may put children at risk of toxic or sub therapeutic effects of these drugs. Thus, understanding the regulatory mechanisms of the gene *SLCO1B1* in response to developmental signals is critical to understand physiological changes in endogenous substrates and to provide safe and effective drug therapy in children.

To date, ontogeny studies have generally focused on mRNA expression and, more recently, have expanded to protein abundance targeting specific regions of the reference gene and/or protein sequence. Recently, it has become increasingly apparent that alternative splicing, a process that increases the diversity of products from a single gene, may have functional consequences. Due to alternative splicing events, more than 90% of our genes give rise to more than one mRNA transcript, varying with respect to numbers of exons, different length of exons, and varying lengths of untranslated regions.[17] Not all products of alternative splicing necessarily result in the production of functional proteins. Alternative splicing may be the result of developmental signals expressed during the course of growth and development.[17-19] For example, developmentally regulated alternative splicing has been demonstrated for neuronal sodium channels genes *SCN1A, SCN2A, SCN3A, SCN8A* and *SCN9A* in brain tissue; the alternative exon 5N predominates in the neonatal period whereas 5A predominates in the adult.[20-25] In the case of *SCN1A,* a gene implicated in the pathogenesis of febrile seizures in newborns[26], an allelic variant *SCN1A* IVS5–91 G>A disrupts the 5′ splice donor site of exon 5N and potentially influences the relative expression of exons 5N and 5A. Although the *SCN1A* IVS5–91 G>A variant does not appear to be associated with febrile seizures *per se*[27], studies suggest that presence of the variant allele may affect dose requirements for phenytoin and carbamazepine.[25,28] Thus, alternative splicing and genetic variants affecting alternative splicing, may have therapeutic consequences.

The *SLCO1B1* gene, consisting of 14 coding and one non-coding exons, codes for the protein OATP1B1 that is composed of 691 amino acids, and consists of 12 transmembrane (TM) regions.[10] It is part of the *SLCO1B* family, for which splice variants have been described. For example, five mRNA transcripts for *SLCO1B3* have been deposited in Ensembl, of which four represent full-length or truncated protein-coding sequences.[29] Furthermore, the splice variant LST-3TM12 is a hybrid transcript with sequences derived

from *SLCO1B3* and *SLCO1B7*, and has functional transporter properties.[30] In contrast, for *SLCO1B1* there is only one reported mRNA transcript, ENST00000256958.2, encoding the functional 691 amino acid OATP1B1 protein; referred to hereafter as the "reference isoform".[10]

Given these considerations, the purpose of this exploratory study was to investigate alternative splicing of *SLCO1B1* in post-mortem pediatric liver tissue over a wide age range from fetal to adolescent ages. Using RNA sequencing (RNA-seq) data, we created a process involving computational software integrating our bioinformatics pipelines and an in-house developed RNA-seq database query system to perform a structured analysis of the RNA-seq data *in silico*. Using this data analysis pipeline, we aimed (1) to predict splice variants for *SLCO1B1*, (2) to identify potential hybrid splice variants overlapping *SLCO1B1* and *SCLO1B7* (another member of the *SLCO1B* family), and (3) to study age-related changes in expression of *SLCO1B1* splice variants.

MATERIALS AND METHODS

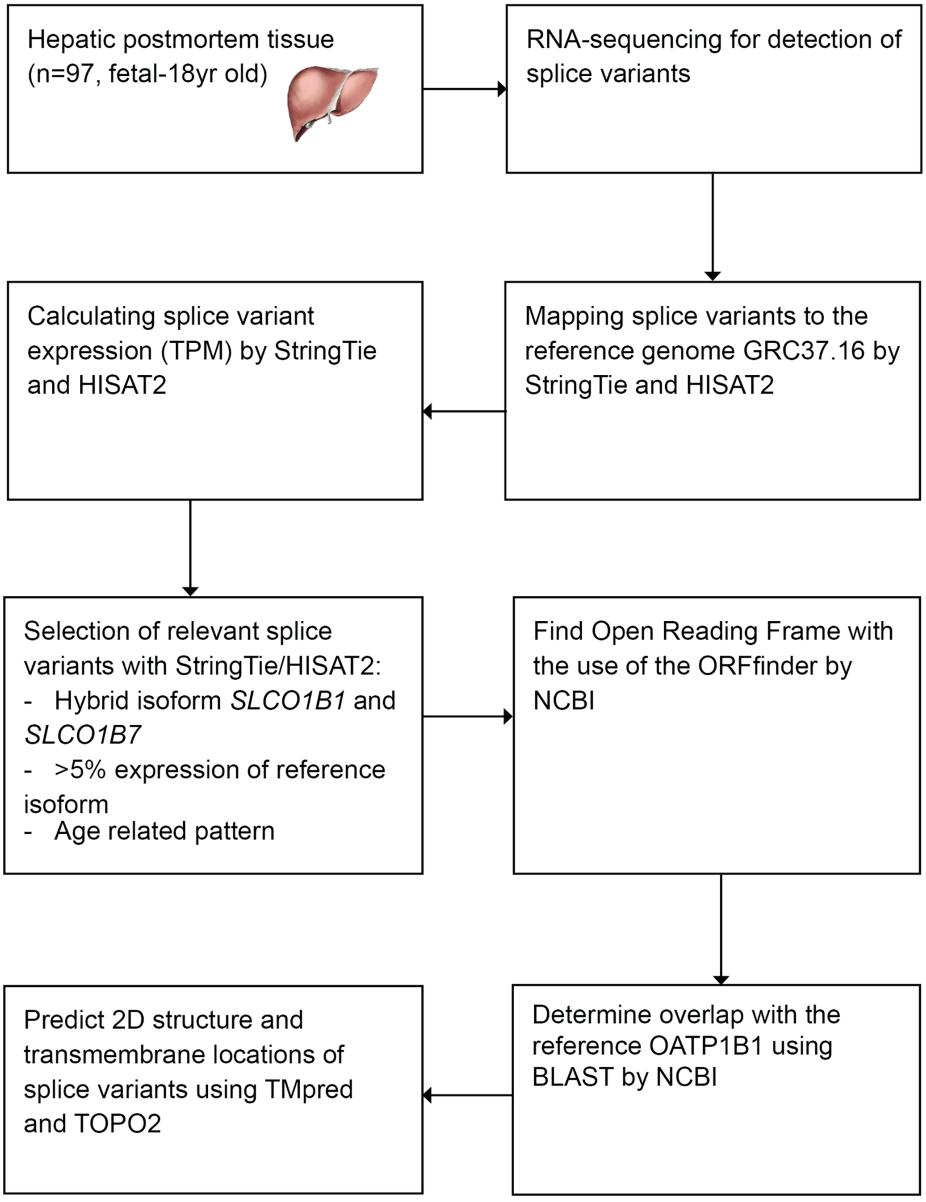
See Figure 1 for the workflow of the methods, and the explanation underneath.

Tissue samples

Post-mortem liver tissue samples from autopsies of fetuses (therapeutic abortions or stillbirths) and infants were provided by the Erasmus MC Tissue Bank (Rotterdam, NL) and the repository of the Division of Clinical Pharmacology, Toxicology, and Therapeutic Innovation at Children's Mercy Kansas City (Kansas City, MO, USA). Tissue was procured at the time of autopsy within 24 hours after death and snap-frozen at -80^{0}C for later research use. For the tissues provided by the Erasmus MC tissue bank, the Erasmus MC Research Ethics Board waived the need for formal ethics approval according to the Dutch Law on Medical Research in Humans. Tissue was collected when parental written informed consent for both autopsy and the explicit use of the tissue for research was present. The samples were selected when the clinical diagnosis of the patient was not related to hepatic problems and the tissue was histologically normal (as estimated by a pathologist based on hematoxylin and eosin staining). Postmortem pediatric liver tissues in the repository of the Division of Clinical Pharmacology, Toxicology, and Therapeutic Innovation at Children's Mercy Kansas City (Kansas City, MO, USA) were obtained from multiple sources, including the Brain and Tissue Bank for Developmental Disorders at the University of Maryland (Baltimore, MD), the Liver Tissue Cell Distribution System (LTCDS; University of Pittsburgh and University of Minnesota), University of Washington Center for Birth Defects Research (Seattle, WA) and XenoTech LLC (Lenexa, KS). The use of

these tissues was declared nonhuman subject research by the Children's Mercy Hospital Pediatric Institutional Review Board.

Figure 1 Flow of methods predicting splice variants of SLCO1B1



RNA sequencing

mRNA expression of SLCO1B1 was determined using RNA-seq. RNA was isolated from hepatic tissue according to the manufacturer's instructions using the RNeasy Mini kit (Qiagen, Valencia, CA). Samples with an RNA integrity number of <5 were excluded. The RNA-seq experiments were performed according to the Illumina RNA-seq protocol (San Diego, CA). In brief, a population of poly(A)$^+$ mRNA was selected and converted to a library of cDNA fragments (220–450 bp) with adaptors attached to both ends, using an Illumina mRNA-Seq sample preparation kit. The quality of the library preparation was confirmed by analysis on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The cDNA fragments were then sequenced on an Illumina HiSeq 2000 to obtain 100-bp sequences from both ends (paired end). The resulting reads were mapped by Bowtie2 and StringTie[31-33] to the transcriptome constructed through reference genes/transcripts according to the reference human genome GRCh37.61/hg19.[34] The mapped reads were then assigned to transcripts from which the abundance of the reference transcript is estimated by RSEM[35] and for the splice variants with HISAT2.[36] The counts of RNA-seq fragments were used to indicate the amount of identified mRNA transcripts, presented in transcripts per million transcripts (TPM).[35]

Validation dataset

To validate the RNA-seq results, the presence of the reference transcript was confirmed (Ensembl transcript ID ENST00000256958.3). Moreover, to further validate the results, the presence of the alternatively spliced transcript LST-3TM12[30] was detected using the algorithm RSEM combined with Bowtie2 and the assembly GRCh37.

Protein prediction

Sequence alignment and overlap of splice variants with consensus coding sequences (CCDS)[37] for *SLCO1B1* and the adjacent gene *SLCO1B7* was performed using Basic Local Alignment Search Tool (BLAST).[38] Splice variants were prioritized for further investigation when one of the following criteria was met and when the presence of the splice variant was verified with RT-PCR:

- the expression of the splice variant was >5% of the expression of the reference isoform,
- the splice variant was a *SLCO1B7* and *SLCO1B1* hybrid transcript, or
- the expression of the splice variant was associated with age (see 'Data and Statistical analysis')

Next, the open reading frame (ORF) of >600 nt of the relevant splice variants was predicted with ORF-Finder by NCBI.[39] Prediction of transmembrane (TM) regions and orientation was done with TMpred based on the TMbase database.[40] Two dimensional protein structures were generated with TOPO2.[41]

To provide additional bioinformatic confirmation that candidate novel alternatively spliced products represent coding transcripts, sequencing data were analyzed using two additional tools: the Coding Potential Calculator Algorithm (CPC2) and the Coding-Potential Assessment Tool (CPAT).[42,43] These algorithms both use logistic regression to distinguish between coding and non-coding transcripts based on four intrinsic features: the Fickett testcode score (both), ORF length (both), ORF integrity (CPC2), isoelectric point (CPC2), ORF coverage defined as the ratio of ORF to transcript lengths (CPAT) and hexamer usage bias (CPAT).

Verification of splice variants by RT-PCR and sequencing

The presence of the splice variants selected for further investigation was verified by RT-PCR and sequencing. Primers were designed to be specific for each splice variant (Figure S1 and Table S1). In addition, a universal primer pair was designed to amplify all splice variants and to function as a positive control. Due to the low abundance of some of the variants, a nested forward primer was also designed.

RNA was extracted from frozen liver tissue, utilizing the Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA). One µg of total RNA was DNase treated and reverse transcribed with the Maxima H Minus First Strand cDNA Synthesis Kit, following the manufacturer's instructions (Thermo Scientific, Waltham, MA). The cDNA equivalent to 10 ng of total RNA was used per PCR reaction (2G Fast ReadyMix, KAPA Biosystems, Wilmington, MA). The cycling conditions were: 94°C, 3 min, followed by 40 cycles of 94°C for 15 sec, 60°C for 15 sec and 72°C for 5 sec. The primary PCR amplicons were diluted 1:4000 and a nested PCR was performed with the same KAPA mix and the same cycling conditions. The PCR reactions were column purified up with the QIAquick PCR Purification Kit (Qiagen). One ng was used in subsequent sequencing reactions with BigDye v.3.1 and run on a 3730xl DNA Analyzer (Thermo Fisher). The results were analyzed with Sequencher software (Gene Codes, Ann Arbor, MI).

Data and Statistical analysis

Because of non-normal distribution, the data are presented as median (range). First, the relative abundance of the expression of each transcript compared to the reference isoform was calculated. Also, the relationship of age with expression levels (TPM) was studied by comparing expression levels between age groups. Samples were assigned to one of five age groups: fetal, 0-1.5 years, 1.5-6 years, 6-12 years and 12-18 years. Kruskal-Wallis test with Dunn's post-hoc test were used for multiple comparisons of expression levels between age groups. For Dunn's post-hoc test for multiple comparisons the adjusted p-values are reported, in which a correction for multiple testing for age groups is applied. Second, Spearman correlations between age (on a continuous scale)

and expression levels of splice variants were examined. To control for the number of correlations tested (54), p-values were considered statistically significant only if their corresponding q-value was less than .05 after Benjamini-Hochberg adjustment to control the false discovery rate. Statistical analyses were performed using IBM SPSS Statistics software (SPSS Statistics for Windows, version 21.0; IBM, Armonk, NY) and graphical exploration was done with GraphPad Prism. We explored negative binomial and zero-inflated negative binomial models in SAS 9.4, but the former did not fit well, and we were unable to identify predictors of excess zeros for the latter.

RESULTS

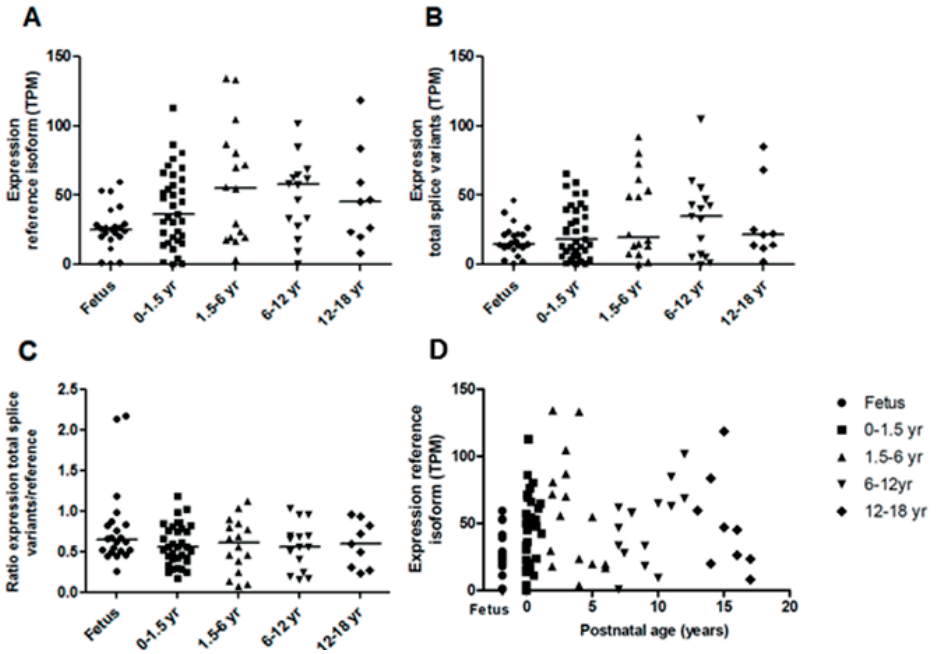
Descriptive results

mRNA expression of the reference isoform and splice variants of SLCO1B1 was quantified in 97 post-mortem liver tissues of humans of various ages, of which the age distribution can be found in Table 1. The reference isoform of SLCO1B1 was detected in all but one sample with a median expression of 33.4 (range 0-134.2) TPM. The transcript consisted of 2791 nucleotides (nt) of which 95 nt comprise the 5'-UTR and 621 nt the '3-UTR, resulting in a protein of 691 amino acids. This is in accordance with Niemi *et al*.[10] In Figure 2A the expression of this transcript in various age groups is presented and did not show any age-related changes when binned in age groups. On the other hand, on a continuous scale, postnatal age was related to transcript expression (ρ=0.316, p=0.002) (Figure 2D). Twenty-seven splice variants of SLCO1B1 were identified using RNA-seq, representing a total expression of 18.8 (0.2-105.0) TPM (Figure 2B). These are numbered randomly between 21 and 55. The ratio of the total expression of the splice variants versus the reference isoform is presented in Figure 2C and was not different between age groups. Thirteen splice variants met the selection criteria for further analysis, and ten of these were subsequently verified by RT-PCR (Table 2), as described below.

Table 1 Median (range) age by group for post-mortem liver samples

Age groups	Number of samples	Gestational age (weeks)	Postnatal age (years)
Fetus	22	16.4 (14.7-41.3)	-
0 – 1.5 yr	35	-	0.1 (0-1.2)
1.5 – 6 yr	16	-	3 (1.8-6)
6 – 12 yr	15	-	9 (7-12)
12 – 18 yr	9	-	15 (13-17)
Total	97		

Figure 2 TPM expression of (A) the reference isoform of SLCO1B1 (B) the total TPM values of splice variants and (C) the ratio of total TPM values of splice variants to TPM values of the reference isoform in various age groups; and (D) the relationship of the reference isoform of SLCO1B1 with postnatal age (ρ=0.316, p=0.002).



Verification splice variants by RT-PCR and sequencing

The presence of the splice variants meeting one or more criteria of 1) expression level >5% of the expression of the reference isoform, 2) a hybrid *SLCO1B7* and *SLCO1B1* transcript, or 3) expression was associated with age, were verified for 10/13 samples; splice variants 21, 37 and 48 could not be verified by RT-PCR (see Figure S2). Splice variants 21 and 37 had sizes different than expected. Splice variant 48 could not be amplified. Further analysis of variants 21, 37 and 48 by sequencing was also unable to confirm the presence of the 21, 37 and 48 splice variants, and thus these transcripts were excluded from further analysis. The splice junctions of variant 46 were not sufficiently unique to design primers that would amplify only this variant or to generate a product with a size that could resolved on an agarose gel from amplicons generated from other transcripts as templates. The results for splice variant 46 should therefore be interpreted with caution.

Validation dataset

To further validate our RNA-seq results, we confirmed the presence of the LST-3TM12 transcript[30] in our samples. The transcript emerged in 3 of the 97 samples with a low abundance of 0.11, 0.18 and 0.30 TPM.

Table 2 Relevant splice variants of SLCO1B1 in 97 pediatric liver samples for which the presence in the samples is confirmed by RT-PCR

Splice variant	Abundance of all novel isoforms (%)	Abundance compared to reference isoform (%)	Found in number of samples	Number of exons	Length (nt)	ORF (n AA)	Overlapping number of AA with locus:					Number of TM helices
							ORF SLCO1B1 (% of reference SLCO1B1)	Intron SLCO1B1	ORF SLCO1B7	In between SLCO1B1 and SLCO1A2†	SLCO-1A2†	
46‡*	26.55	16.48	63	10	4638	284	274 (40%)	10	-	-	-	6
50‡	14.26	8.85	93	17	34388	453	444 (65%)	9	-	-	-	10
34‡	9.24	5.73	46	18	4156	625	622 (90%)	-	-	3	-	11
24◊*	0.32	0.20	77	25	3151	659	622 (90%)	-	0	-	37	11
26◊*	1.80	1.12	81	24	13158	691	691 (100%)	-	0	-	-	12
28◊*	1.21	0.75	96	20	14577	691	691 (100%)	-	0	-	-	12
						210	-	-	-	-	210	0
						881	-	-	-	-	881	1
30◊*	0.34	0.21	54	20	5684	484	453 (66%)	-	0	-	31	8
38*	0.52	0.33	95	15	22310	453	444 (65%)	9	-	-	-	10
44*	7.86	4.88	65	3	7364	98	98 (14%)	-	-	-	-	2
51*	0.73	0.46	76	17	15407	522	522 (75%)	-	-	-	-	8

*AA=amino acid nt=nucleotides TM=transmembrane, †SLCO1A2 is on the reverse strand ‡Abundant splice variant ◊ Hybrid SLCO1B1 and -1B7 *Age related changes in expression*

Splice variants meeting the abundance criterion

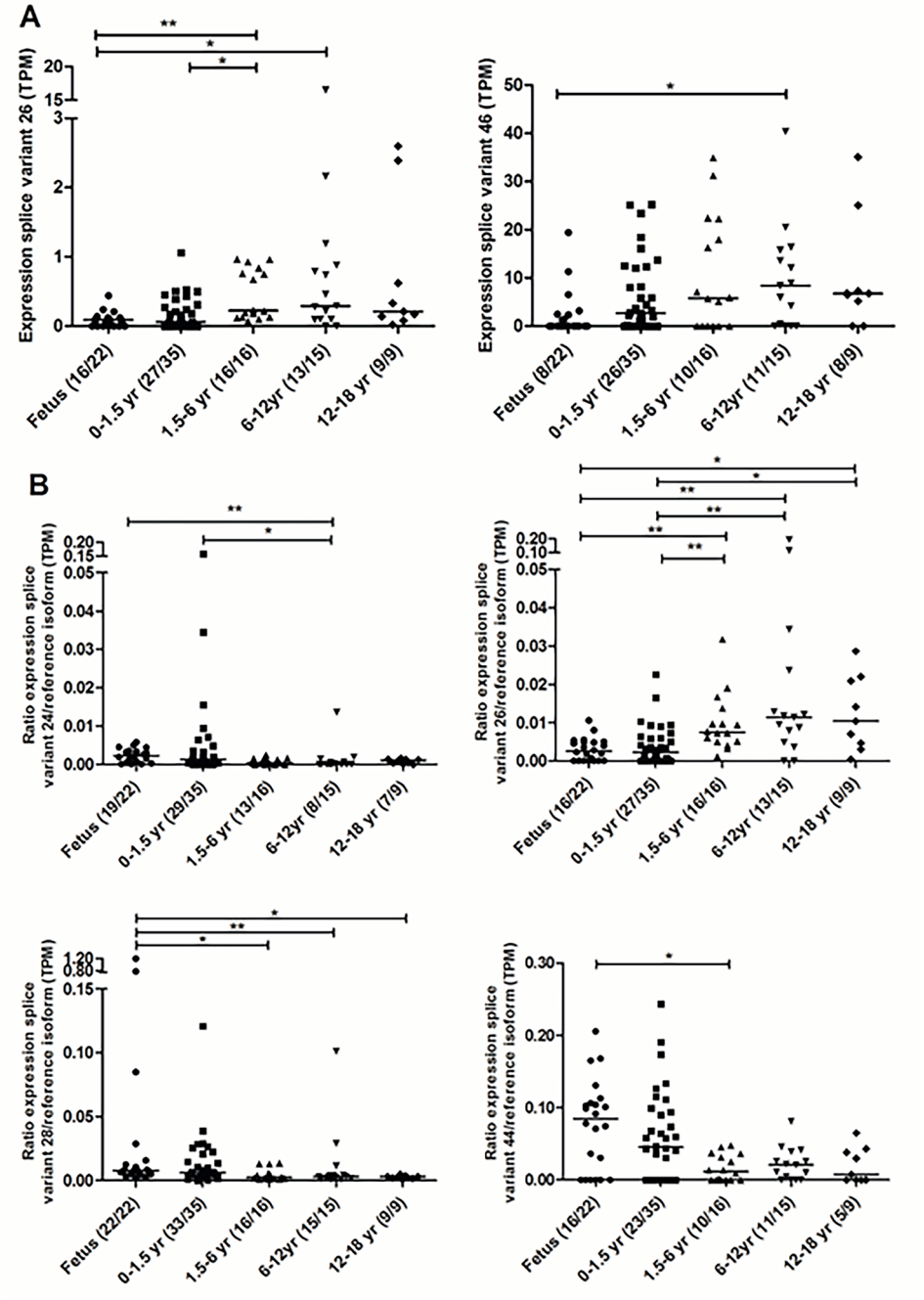
Three splice variants had an abundance of >5% of the expression of the reference isoform (Table 2). They had 40-90% overlap with the ORF from the reference amino acid sequence for OATP1B1, resulting in a prediction of a number of TM helices ranging from 6 to 11. These three splice variants are therefore predicted to result in truncated versions of OATP1B1.

Splice variants overlapping with SLCO1B7 and SLCO1B1

Four splice variants were identified with exons overlapping the *SLCO1B1* as well as the *SLCO1B7* gene region (Table 2). They all had an ORF overlapping >40% of the amino acid sequence of OATP1B1. However, the ORF of none of them was overlapping the *SLCO1B7* region. Two of these isoforms are predicted to translate into similar protein versions of OATP1B1, as they have 100% overlap with the reference isoform. All four splice variants have longer untranslated regions than the reference isoform.

One isoform, sv28, overlapped with *SLCO1A2*. This hybrid isoform contains an intron-less complete coding sequence, which is officially located in intron 13 of *SLCO1A2*.

Figure 3 Expression (A) and expression in relation to the reference isoform (B) of developmentally regulated splice variants of SLCO1B1 in various age groups. Counts of tissues with isoform expressed out of total counts by age group are provided in parentheses. *p<0.05; **p<0.01



Splice variants with age-related expression

Age groups

The splice variants 26 and 46 showed age-related changes in their expression with lower expression in fetal tissue than in tissue from older children (see Figure 3A for specific changes and Table 2 for splice variant information). When analyzed as a ratio to the expression of the reference isoform, for one splice variant (26) the ratio variant/reference isoform increased with age, while three splice variants decreased with age: isoforms 24, 28 and 44 (see Figure 3B for specific changes and Table 2 for splice variant information). This latter observation reflects that for the individual samples either the expression of the splice variant was lower, or the expression of reference isoform was higher. As the expression of the reference isoform was shown to be similar when binned in age groups (Figure 3A), it is therefore likely that the expression of the splice variants that decreased with age was lower.

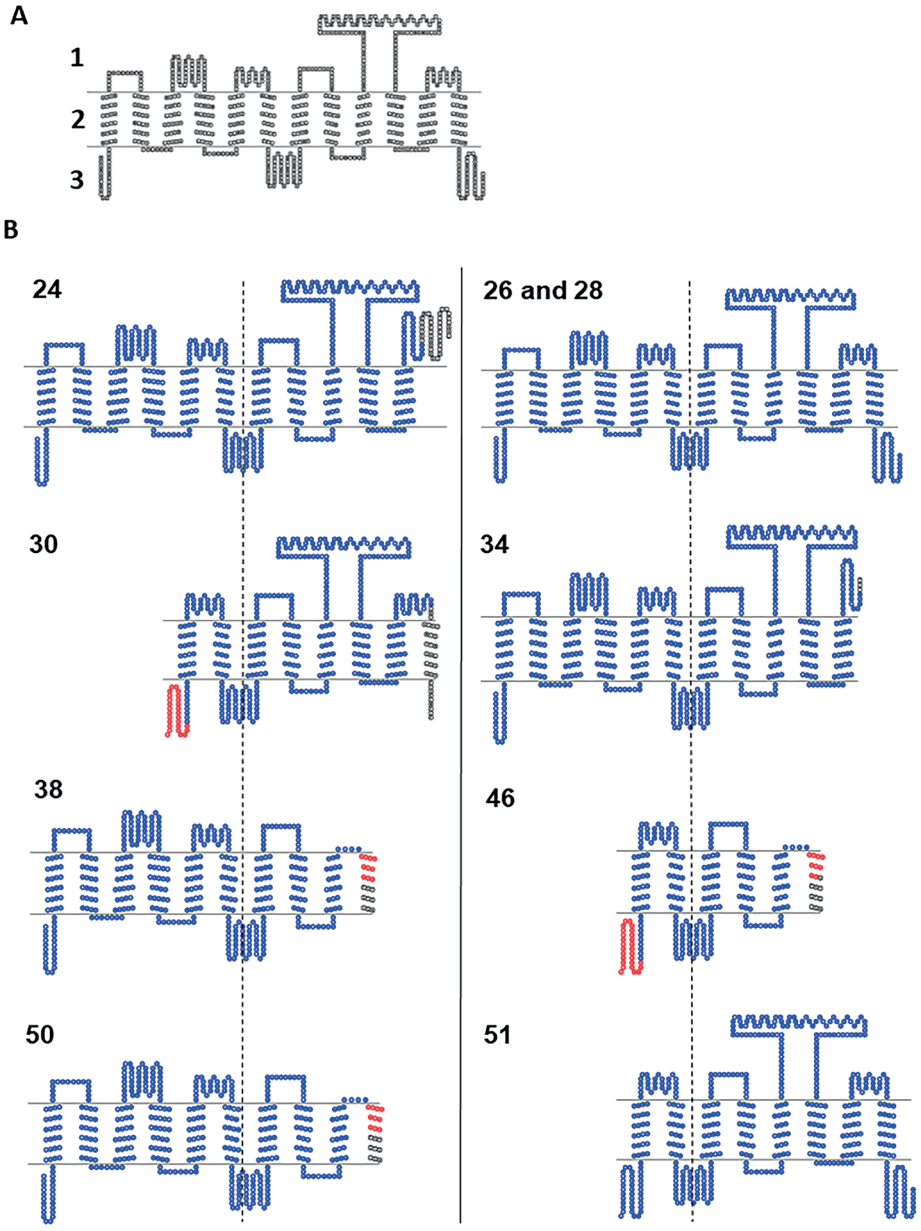
Age on continuous scale

The expression of four of the five hybrid SLCO1B7 and SLCO1B1 splice variants (24, 26, 28, 30) and the abundant splice variant 46 are significantly correlated with age (see Table 3). More specifically, the expression of the splice variants 24 and 28 decreased with increasing age, and the expression of 26, 30 and 46 increased with increasing age. When splice variant expression is analyzed as a ratio to the expression of the reference isoform, correlation with age was found for the same and for four additional splice variants (see Table 3). The expression of splice variant 48 was correlated with age, but was excluded for further analysis as their presence was not verified by RT-PCR (see 5.2).

Predicted protein structure

In Table S3 the coding-potential prediction results using the CPAT and CPC2 tool are depicted. Splice variant 44 has a low coding probability, hence will likely not result in a protein product. This can be explained by the fact that the ORF length was small compared to the size of the splice variant or because of a high hexamer-score.[42,43] The hexamer-score is a feature dependent on adjacent amino acids in proteins and is based on a log-likelihood ratio to measure differential hexamer usage between coding and noncoding sequences.[44] All other splice variants have a very high probability to be translated into a protein product. In Figure 4A the 2D structure of OATP1B1 is presented, consisting of 12 TM helices. In Figure 4B the predicted 2D structure of the splice variants with an ORF with high probability to be translated in a protein product are presented.

Figure 4 (A) Predicted 2D structure of reference OATP1B1 (1: extracellular, 2: transmembrane, 3: intracellular) and (B) the predicted 2D structure of splice variants of OATP1B1, centered on the 4th intracellular loop (dashed line) of the reference structure for OATP1B1.



The number of the splice variant is presented in the upper left corner of each structure. Red and blue: overlapping amino acid sequence with OATP1B1. Blue: overlapping structure OATP1B1.

Table 3 Spearman correlations expression splice variant vs. postnatal age

Splice variant	Expression splice variant (TPM) vs. postnatal age (weeks)		Ratio expression splice variant/reference isoform vs. postnatal age (weeks)	
	r_s	p value	r_s	p value
21	0.231	0.023	0.194	0.057
24	-0.330	0.001*	-0.392	0.000*
26	0.489	0.000*	0.518	0.000*
28	-0.263	0.009*	-0.433	0.000*
30	0.290	0.004*	0.296	0.003*
33	0.142	0.166	0.143	0.163
34	0.133	0.193	0.110	0.285
35	-0.035	0.734	-0.222	0.029
36	0.063	0.539	0.058	0.575
37	0.069	0.501	0.006	0.955
38	-0.070	0.496	-0.271	0.007*
39	-0.141	0.167	-0.188	0.065
40	-0.017	0.869	-0.092	0.371
41	0.003	0.977	-0.001	0.990
42	0.055	0.591	-0.059	0.566
43	0.065	0.528	0.063	0.542
44	-0.121	0.240	-0.259	0.010*
45	-0.017	0.867	-0.020	0.843
46	0.386	0.000*	0.343	0.001*
47	-0.023	0.825	-0.106	0.299
48†	-0.243	0.016	-0.463	0.000*
49	-0.117	0.253	-0.146	0.152
50	0.204	0.045	-0.202	0.047
51	-0.177	0.083	-0.308	0.002*
53	-0.029	0.779	-0.043	0.678
54	-0.155	0.129	-0.169	0.097
55	-0.173	0.091	-0.203	0.046

** Significant after adjustment to control false discovery rate at .05 †excluded from further analysis because the presence was not confirmed by RT-PCR*

DISCUSSION

In the current study we have developed a data analysis pipeline that allowed a structured analysis of a large amount of RNA-seq data generated from pediatric liver samples and used this to investigate alternative splicing of the *SLCO1B1* gene that could potentially translate into functional OATP1B1 proteins. More specifically, we report three major findings from the ten relevant splice variants that we identified: (1) two splice variants are predicted to translate into the same amino acid sequence as the reference isoform for OATP1B1; (2) eight splice variants may translate into truncated versions of the OATP1B1 protein because of an altered length of amino acid sequence, and (3) the expression of eight splice variants was associated with age. None of the splice variants had an ORF that covered the *SLCO1B7* region.

Our results show that the *SLCO1B1* gene locus is subject to alternative splicing, as supported by the major findings presented above. More specifically, the fact that eight splice variants of SLCO1B1 showed a developmental pattern is consistent with developmentally regulated alternative splicing as a mechanism for altered SLCO1B1/OATP1B1 expression during growth and maturation. This finding may have implications for the functionality of the transporter in children, and with that the disposition of its endogenous and exogenous substrates, as most of these splice variants were predicted to result in truncated OATP1B1 isoforms with fewer TM regions compared to the reference OATP1B1 protein. Available evidence suggests that *SLCO1B7* is a pseudogene resulting in a non-functional protein product with only 11 TM regions, whereas the *SLCO1B1* gene, the *SLCO1B3* gene and the hybrid transcript LST-3TM12 give rise to at least one mRNA transcript that translates into functional transporters with 12 TM regions.[30] Moreover, the truncated proteins encoded by the transcripts we observed may lack one or more of the N-glycosylation sites Asn134, Asn503 and Asn516, located at the extracellular loop 2 and 5 of OATP1B1.[45] Glycosylation is a post-translational modification that is suggested to be essential for the proper function of OATP1B1. Disruption of all these sites led to lower protein stability with reduced total protein levels, and non-glycosylated OATP1B1 was retained within the endoplasmic reticulum, e.g. was not present on the cell membrane.[45]

Some of the alternative proteins of OATP1B1 reported in this study therefore are likely to result in non-functional protein products incapable of cellular transport, but could possibly possess alternative functional properties, such as regulating the activity of the functional OATP1B1 transporter protein. A precedent for this type of regulatory role is illustrated by the DME *UGT1A*, a complex gene with 3 major mRNA variants created by splicing events involving an alternative 5a or 5b exon. The classic variant (i1) with exon 5a has transferase activity, whereas the alternative proteins (i2), with either exon 5b or with both exon 5a and

5b, lack transferase activity.[46] The relative glucuronidation of SN-38, a substrate for UGT1A, was decreased in the presence of i2 proteins despite the same amount of i1 enzyme.[47] This phenomenon is explained by oligomerization of UGT1A enzyme; i2 proteins can form dimers with i1 enzymes, inhibiting the activity of i1 enzymes.[48] Interestingly, OATP1B transporters not only form homo-oligomers, but can also form hetero-oligomers, even with transporters from another family, *e.g.* with Na+/taurocholate co-transporting polypeptide (NTCP).[49,50] It has been demonstrated that a non-functional unit of OATP1B3, containing a lysine at position 41 in place of the wild-type cysteine in the homodimer did not affect normal substrate transport by the functional, cysteine-containing OATP1B3 component of the homodimer, suggesting that each unit within the dimer works as an independent functional unit.[49] However, each splice variant may have its own function and so we hypothesize that those resulting in truncated proteins may influence the transporter activity of the reference protein OATP1B1 and other transporters.

The specific *SLCO1B1* region is part of the wider *SLCO1B*-family region, which gives rise to the SLCO1B3/SLCO1B7 hybrid transcript LST-3TM12 that results in a functional transporter.[30] The four SLCO1B1 splice variants found in this study that contained exons covering the region of *SLCO1B7* did not contain the start codon from the *SLCO1B7* locus, thus we conclude that the *SLCO1B1* gene is not subject to hybridization with adjacent genes. However, the length of the untranslated region (UTR) of these and other transcripts could well be influencing the function of the transporter, even when the ORF of the splice variant is the same as the reference sequence. We note that a transcript of CYP3A4 with a shorter 3'-UTR than the canonical transcript due to an additional polyadenylation site was more stable and generated more protein[51] than an alternative transcript with a longer 3'-UTR. Interestingly, this shorter transcript showed developmental regulation as it was higher expressed in adult livers than in pediatric livers. Nevertheless, it remains to be seen whether this is also the case for the splice variants presented in this study.

Another consequence of these truncated versions of OATP1B1 is that they may interfere with the estimation of OATP1B1 content using LC-MS/MS-based proteomic methods. This quantitative proteomic approach utilizes short peptides to target the protein of interest. All truncated versions of OATP1B1 presented in this manuscript contained the amino acid sequences NVTGFFQSFK[14] and of LNTVGIAK[11] that have been used in studies presenting results on expression of OATP1B1 protein in pediatric liver tissue. The latter refers to our previous study, where a poor correlation was seen between total mRNA expression of SLCO1B1 as measured with RNA-seq and protein abundance of OATP1B1 in a subset of the samples presented in the current study.[11] This lack of correlation can be explained by the fact that not all mRNA transcripts translate into protein. Moreover, potential translation of splice variants into non-functional proteins that are nevertheless

detected by the peptide sequences used to quantitate OATP1B1 content, could also result in poor correlations between abundance and activity. Thus, care may be needed when extrapolating mRNA expression to protein abundance, protein abundance to actual activity, and ultimately the prediction of disposition of transporter substrates.

We recognize that a limitation of our study is that our results are based on *in silico* predictions, and the presence of the corresponding truncated proteins must be confirmed by protein abundance studies before any of the implications we propose above can be assessed, including investigations of a dominant-negative regulatory role analogous to the UGT1A situation. Developmental regulation of alternative splicing is a commonly recognized phenomenon during tissue development and cell differentiation. In fact, level of expression, localization within the cell, mRNA stability, translation efficiency and splicing of specific RNA binding proteins (RBPs) is finely regulated. RBPs bind to cis-elements and promote or inhibit splice site recognition, and therefore RBP expression coordinates alternative splicing networks during development.[18] Further work is necessary to characterize the specific developmental signals responsible for the observed changes in expression of the SLCO1B1 splice variants.

These exploratory data imply that the complexity of processes involved with growth and development throughout childhood may have influences on transporter expression and subsequent substrate disposition, as yet unrecognized. The observed age-related changes in expression of splice variants in the context of age-related changes in concentrations of endogenous OATP1B1 substrates, such as DHEA-S and 16alpha-hydroxylated metabolites, makes it tempting to speculate that additional regulatory mechanisms may be in play, with implications for the disposition of exogenous substrates used in pediatrics. Most importantly, the data analysis pipeline we have developed allows the analyses described in this manuscript for SLCO1B1 to be applied to any other gene of interest and will be repeated for other transporters and genes involved in drug disposition or growth of children in the future.

In conclusion, we have shown that SLCO1B1 splice variants with an ORF could potentially translate into proteins with unknown function; they are unlikely to code for functional, transporters, but may have other roles, such as regulatory activity. Moreover, as the expression of particular SLCO1B1 splice variants showed age-related changes, the data raise the possibility of a regulatory role for alternative splicing in mediating developmental changes of SLCO1B1/OATP1B1 in drug disposition. These data can contribute to improved understanding of age-related changes in expression of SLCO1B1, and possibly other enzymes and transporters involved in the disposition of endogenous and exogenous substrates throughout growth and development.

REFERENCES

1. Brouwer KL, Aleksunes LM, Brandys B, et al. Human ontogeny of drug transporters: review and recommendations of the pediatric transporter working group. Clin Pharmacol Ther 2015;98(3):266-287.
2. Nigam SK. What do drug transporters really do? Nat Rev Drug Discov 2015;14(1):29-44.
3. Mooij MG, Nies AT, Knibbe CA, et al. Development of human membrane transporters: drug disposition and pharmacogenetics. Clin Pharmacokinet 2016;55(5):507-524.
4. Stevens JC, Hines RN, Gu C, et al. Developmental expression of the major human hepatic CYP3A enzymes. J Pharmacol Exp Ther 2003;307(2):573-582.
5. Leeder JS, Gaedigk R, Marcucci KA, et al. Variability of CYP3A7 expression in human fetal liver. J Pharmacol Exp Ther 2005;314(2):626-635.
6. Oshiro C, Mangravite L, Klein T, Altman R. PharmGKB very important pharmacogene: SLCO1B1. Pharmacogenet Genomics 2010;20(3):211-216.
7. Kojima S, Yanaihara T, Nakayama T. Serum steroid levels in children at birth and in early neonatal period. Am J Obstet Gynecol 1981;140(8):961-965.
8. France JT. Levels of 16-alpha-hydroxydehydroepiandrosterone, dehydroepiandrosterone and pregnenolone in cord plasma of human normal and anencephalic fetuses. Steroids 1971;17(6):697-719.
9. Nakamura H, Torimoto N, Ishii I, et al. CYP3A4 and CYP3A7-mediated carbamazepine 10,11-epoxidation are activated by differential endogenous steroids. Drug Metab Dispos 2003;31(4):432-438.
10. 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(1):157-181.
11. van Groen BD, van de Steeg E, Mooij MG, et al. Proteomics of human liver membrane transporters: a focus on fetuses and newborn infants. Eur J Pharm Sci 2018;124:217-227.
12. Mooij MG, Schwarz UI, de Koning BA, et al. Ontogeny of human hepatic and intestinal transporter gene expression during childhood: age matters. Drug Metab Dispos 2014;42(8):1268-1274.
13. Thomson MM, Hines RN, Schuetz EG, Meibohm B. Expression Patterns of Organic Anion Transporting Polypeptides 1B1 and 1B3 Protein in Human Pediatric Liver. Drug Metab Dispos 2016;44(7):999-1004.
14. Prasad B, Gaedigk A, Vrana M, et al. Ontogeny of hepatic drug transporters as quantified by LC-MS/MS proteomics. Clin Pharmacol Ther 2016;100(4):362-370.
15. Dennery PA, Seidman DS, Stevenson DK. Neonatal hyperbilirubinemia. N Engl J Med 2001;344(8):581-590.
16. Huang MJ, Kua KE, Teng HC, Tang KS, Weng HW, Huang CS. Risk factors for severe hyperbilirubinemia in neonates. Pediatr Res 2004;56(5):682-689.
17. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature 2008;456(7221):470-476.
18. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 2017;18(7):437-451.
19. Castle JC, Zhang C, Shah JK, et al. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. Nat Genet 2008;40(12):1416-1425.
20. Plummer NW, Meisler MH. Evolution and diversity of mammalian sodium channel genes. Genomics 1999;57(2):323-331.

21. Plummer NW, McBurney MW, Meisler MH. Alternative splicing of the sodium channel SCN8A predicts a truncated two-domain protein in fetal brain and non-neuronal cells. J Biol Chem 1997;272(38):24008-24015.
22. Gustafson TA, Clevinger EC, O'Neill TJ, Yarowsky PJ, Krueger BK. Mutually exclusive exon splicing of type III brain sodium channel alpha subunit RNA generates developmentally regulated isoforms in rat brain. J Biol Chem 1993;268(25):18648-18653.
23. Sarao R, Gupta SK, Auld VJ, Dunn RJ. Developmentally regulated alternative RNA splicing of rat brain sodium channel mRNAs. Nucleic Acids Res 1991;19(20):5673-5679.
24. Belcher SM, Zerillo CA, Levenson R, Ritchie JM, Howe JR. Cloning of a sodium channel alpha subunit from rabbit Schwann cells. Proc Natl Acad Sci U S A 1995;92(24):11034-11038.
25. Tate SK, Depondt C, Sisodiya SM, et al. Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. Proc Natl Acad Sci U S A 2005;102(15):5507-5512.
26. Mulley JC, Scheffer IE, Petrou S, Dibbens LM, Berkovic SF, Harkin LA. SCN1A mutations and epilepsy. Hum Mutat 2005;25(6):535-542.
27. Petrovski S, Scheffer IE, Sisodiya SM, O'Brien TJ, Berkovic SF, Consortium E. Lack of replication of association between scn1a SNP and febrile seizures. Neurology 2009;73(22):1928-1930.
28. Tate SK, Singh R, Hung CC, et al. A common polymorphism in the SCN1A gene associates with phenytoin serum levels at maintenance dose. Pharmacogenet Genomics 2006;16(10):721-726.
29. Hunt SE, McLaren W, Gil L, et al. Ensembl variation resources. Database 2018.
30. Malagnino V, Hussner J, Seibert I, Stolzenburg A, Sager CP, Meyer Zu Schwabedissen HE. LST-3TM12 is a member of the OATP1B family and a functional transporter. Biochem Pharmacol 2018;148:75-87.
31. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 2016;11:1650.
32. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(4):357-359.
33. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33:290.
34. Genome Reference Consortium. Genome Reference Consortium Human Build 37. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/. Accessed May, 2018.
35. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.
36. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nature Methods 2015;12:357.
37. Farrell CM, O'Leary NA, Harte RA, et al. Current status and new features of the Consensus Coding Sequence database. Nucleic Acids Res 2014;42(Database issue):D865-872.
38. Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL. Magic-BLAST, an accurate DNA and RNA-seq aligner for long and short reads. bioRxiv 2018.
39. NCBI. ORFfinder. https://www.ncbi.nlm.nih.gov/orffinder/. Accessed May, 2018.
40. ExPaSy. TMpred. https://embnet.vital-it.ch/software/TMPRED_form.html. Accessed May, 2018.
41. UCSF. TOPO2. http://www.sacs.ucsf.edu/cgi-bin/open-topo2.py. Accessed May, 2018.
42. Kang YJ, Yang DC, Kong L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res 2017;45(W1):W12-W16.
43. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 2013;41(6):e74.

44. Fickett JW, Tung CS. Assessment of protein coding measures. Nucleic Acids Res 1992;20(24):6441-6450.
45. Yao J, Hong W, Huang J, Zhan K, Huang H, Hong M. N-Glycosylation dictates proper processing of organic anion transporting polypeptide 1B1. PLoS One 2012;7(12):e52563.
46. Girard H, Levesque E, Bellemare J, Journault K, Caillier B, Guillemette C. Genetic diversity at the UGT1 locus is amplified by a novel 3' alternative splicing mechanism leading to nine additional UGT1A proteins that act as regulators of glucuronidation activity. Pharmacogenet Genomics 2007;17(12):1077-1089.
47. Rouleau M, Roberge J, Bellemare J, Guillemette C. Dual roles for splice variants of the glucuronidation pathway as regulators of cellular metabolism. Mol Pharmacol 2014;85(1):29-36.
48. Bellemare J, Rouleau M, Harvey M, Guillemette C. Modulation of the human glucuronosyltransferase UGT1A pathway by splice isoform polypeptides is mediated through protein-protein interactions. J Biol Chem 2010;285(6):3600-3607.
49. Zhang Y, Boxberger KH, Hagenbuch B. Organic anion transporting polypeptide 1B3 can form homo- and hetero-oligomers. PLoS One 2017;12(6):e0180257.
50. Ni C, Yu X, Fang Z, Huang J, Hong M. Oligomerization Study of Human Organic Anion Transporting Polypeptide 1B1. Mol Pharm 2017;14(2):359-367.
51. Li D, Gaedigk R, Hart SN, Leeder JS, Zhong XB. The role of CYP3A4 mRNA transcript with shortened 3'-untranslated region in hepatocyte differentiation, liver development, and response to drug induction. Mol Pharmacol 2012;81(1):86-96.

SUPPLEMENTAL INFORMATION

Figure S1 Example for splice variant 34 of the design of primers by the identification of unique location F=Forward, R=Reverse, nest=nested primer

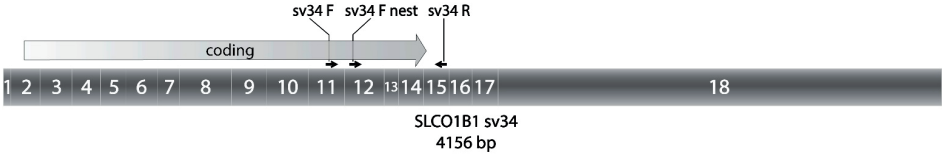
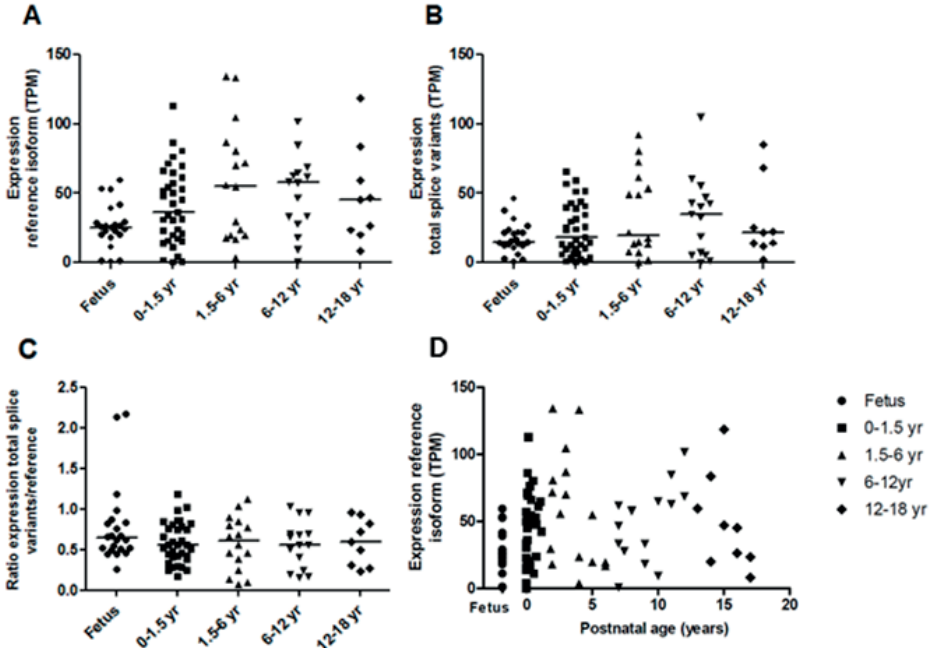


Figure S2 Nested PCR products for different splice variants, analyzed on a 4% agarose gel before subsequent sequencing.



The arrow indicates the band which was excised from the gel; the amplicon was extracted and sequenced. Black indicates correct amplicon size which was confirmed by sequencing. Gray indicated band sizes which were either too large (sv21), too short (sv37) or did not amplify at all. The universal SLCO1B1 primer pair (sv-all) was used for the positive control (+), while an actin primer pair was utilized in the none template control (NTC).

Table S1 Primer sequences to verify the existence of splice variants in our samples by RT-PCR. F=Forward, R=Reverse, nest=nested, sv=splice variant

Primer Name	Sequences (5'-3')	Sv
SLCO1B1 sv21 F	CGGCTTCCATTCAATGATTATG	sv21 and sv30
SLCO1B1 sv21 F nest	GATCGCTAGGAGGTATTCTAGTTCC	sv21
SLCO1B1 sv21 R	CCACTATCTCAGGTGATGCTCTATTG	sv21
SLCO1B1 sv24 F	CAGCTGTGGAGCACGAGG	sv24
SLCO1B1 sv24 F nest	GGCTTGAAGTATCTTCTAGGTATGAGAC	sv24
SLCO1B1 sv24 R	ATCTCCGTTCTATATGAATGATGGAAC	sv24
SLCO1B1 sv26 F	ATGATAGTGGCGTCTGCTCCTA	sv26
SLCO1B1 sv26 F nest	GTGAGAGCAGGATTGTTCAACC	sv26
SLCO1B1 sv26 R	AGCTTTGTTCCAGCCTTAATCATC	sv26
SLCO1B1 sv28 F	TTCAAAATAGCTATTTTGAGGAAACTCATAG	sv28
SLCO1B1 sv28 F nest	GGATAATACCAGAGAACTTCTCAAACTTAGAG	sv28
SLCO1B1 sv28 R	CTGGTATTGATGAAATCCCTCAGTG	sv28
SLCO1B1 sv30 F nest	GAAGAGACATTTTCACCAGTATCTTCTAG	sv30
SLCO1B1 sv30 R	TGATGCTCAGTTTGAAACAATCAC	sv30
SLCO1B1 sv34 F	AACAATGGAATAACTTACATCTCACCC	sv34
SLCO1B1 sv34 F nest	CAGAACAGAAATTACTCAGCCCAT	sv34
SLCO1B1 sv34 R	GATTTAGAACCTACAGCAACTGCAG	sv34 and sv50
SLCO1B1 sv37 F	ATCTAAGGCTAACATCTTATTGGGAGTC	sv37
SLCO1B1 sv37 F nest	ATAACCATACCTATTTTTGCAAGTGG	sv37
SLCO1B1 sv37 R	TGGTACATCTCTATGAGATGTCACTGG	sv37
SLCO1B1 sv38 F	GGGTTTCCACTCAATGGTTATACG	sv38
SLCO1B1 sv38 F nest	GGGCTCTGATTGATACAACGTGTATA	sv38
SLCO1B1 sv38 R	CATCTCTTAAGCCCAGGAAGC	sv38
SLCO1B1 sv44 F nest	CAATGGATTGAAGGAATTTCATAATAC	sv44
SLCO1B1 sv44 R	TGATGATTATGTGTCTTGTGGATGAC	sv44
SLCO1B1 sv48 F	TGCTGTAGGATTCTAAATCCAGGTG	sv48 and sv44
SLCO1B1 sv48 F nest	GAGGCACAACCTTCAGAGAATAAG	sv48
SLCO1B1 sv48 R	TTCCAAATATTGGAGTGAATGTATTCTC	sv48
SLCO1B1 sv50 F	AGCTATTGGGACTGAAGAGACCATAC	sv50
SLCO1B1 sv50 F nest	GGACATAAGAAAGTCTGTTCTAAACTTACAG	sv50
SLCO1B1 sv51 F	CAGCTTTATTGCTAAGACACTAGGTGC	sv51
SLCO1B1 sv51 F nest	GAATTGGAGGTGTTTTGACTGC	sv51
SLCO1B1 sv51 R	TCTTATAGGCAAAGACGTACAGTATATACGTTATAC	sv51
SLCO1B1 sv-all F	CTGGGAAATTGACAGAAAGTACTCTG	all sv
SLCO1B1 sv-all F nest	GGGAAGATAATGGTGCAAATAAAG	all sv
SLCO1B1 sv-all R	CAAAGAAGAATGTCCTTCTTTAGCG	all sv

Table S2 non-relevant splice variants of SLCO1B1 in 97 pediatric liver samples

Splice variant	Abundance of all novel isoforms (%)	Abundance compared to reference isoform (%)	Found in number of samples	Number of exons	Length (nt)	ORF (n AA)	Overlapping number of AA with locus:				Number of TM helices
							ORF SLCO1B1 (% of reference SLCO1B1)	Intron *SLCO1B1*	In between *SLCO1B1* and *SLCO1A2*†	*SLCO-1A2*†	
49	3.81	2.37	24	17	2511	455	443 (64%)	12	-	-	9
43	0.92	0.57	6	10	3826	420	199 (29%)	221	-	-	9
55	0.79	0.49	43	4	10573	107	61 (9%)	42	5	-	2
39	0.64	0.40	61	16	8163	453	443 (64%)	10	-	-	10
54	0.51	0.31	30	2	5898	50	50 (7%)	-	-	-	1
42	3.88	2.41	88	13	13920	347	331 (48%)	16	-	-	6
53	1.50	0.93	55	2	3148	98	98 (14%)	-	-	-	2
47	3.69	2.29	46	16	2394	455	443 (64%)	12	-	-	9
41	0.83	0.51	18	19	13498	691	691 (100%)	-	-	-	12
33	0.67	0.42	18	20	13612	691	691 (100%)	-	-	-	12
35	1.24	0.77	85	18	4697	456	452 (65%)	-	4	-	8
40	1.52	0.94	56	18	3389	659	621 (90%)	-	-	38	12
36	2.11	1.31	16	19	2530	625	621 (90%)	-	4	-	11
45	0.47	0.29	11	16	6316	453	443 (64%)	10	-	-	10
37	8.65	5.37	56	16	2935	455	444 (65%)	11	-	-	9
21	2.56	1.59	53	17	2151	490	452 (65%)	-	-	38	7
48	3.36	2.08	83	3	9877	-	-	-	-	-	-

Table S3 Coding-potential prediction using the Coding-Potential Assessment Tool (CPAT) and Coding Potential Calculator version 2 (CPC2)

	CPAT				CPC2				
Splice variant	**Fickett Score**	**Hexamer Score**	**Coding Probability**	**Coding Label**	**Fickett Score**	**Isoelectric point**	**ORF Integrity**	**Coding Probability**	**Coding Label**
24	0.701	-0.145	0.999	Coding	0.390	8.704	1	1.000	Coding
26	0.659	-0.152	0.999	Coding	0.294	8.852	1	1.000	Coding
28	0.845	0.029	1.000	Coding	0.334	9.696	1	1.000	Coding
30	0.648	-0.146	0.999	Coding	0.324	8.682	1	0.999	Coding
34	0.661	-0.152	0.999	Coding	0.304	8.905	1	1.000	Coding
38	0.841	-0.146	0.999	Coding	0.275	9.185	1	0.999	Coding
44	0.621	-0.284	0.005	Noncoding	0.294	8.864	1	0.089	Noncoding
46	0.845	-0.154	0.924	Coding	0.285	9.026	1	0.972	Coding
50	0.841	-0.146	0.998	Coding	0.271	9.185	1	0.999	Coding
51	0.663	-0.159	0.999	Coding	0.266	8.633	1	0.999	Coding