Quantitative profiling of the *UGT* transcriptome in human drug metabolizing tissues

Alan Tourancheau[1*], Michèle Rouleau[1*], Sandra Guauque-Olarte[1Ŧ], Lyne Villeneuve[1], Isabelle Gilbert[1], Arnaud Droit[2] and Chantal Guillemette[1¥]

[1]Pharmacogenomics Laboratory, Centre Hospitalier Universitaire (CHU) de Québec Research Center (CRCHU), Québec, QC, Canada and Faculty of Pharmacy, Laval University, Québec, Canada

[2]CRCHU de Québec and Faculty of Medicine, Laval University, Québec, Canada

*Co-first authors

[Ŧ]Current address: Grupo GIOD, Facultad de Odontología, Universidad Cooperativa de Colombia, Pasto, Colombia

[¥]Canada Research Chair in Pharmacogenomics

Corresponding Author:

Chantal Guillemette, Ph.D.

Canada Research Chair in Pharmacogenomics

Pharmacogenomics Laboratory

CHU de Québec Research Center

2705 Boul. Laurier, R4720

Québec, Canada, G1V 4G2

Tel. (418) 654-2296

E-mail: Chantal.Guillemette@crchudequebec.ulaval.ca

Short title: Quantitative *UGT* transcriptome profiles

Abstract: 150

Introduction: 420

Material and Methods: 298

Results: 2066

Discussion: 1370

Body text: 4171

References: 33

NUMBER OF:

Tables: 1

Figures: 10

Supplementary Tables: 2

Supplementary Figures: 1

The RNA sequencing data generated by this project have been submitted to NCBI and have been given the accession number GSE82292. They can be accessed with this private link:
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=uhyvgmaezlsffqz&acc=GSE82292

Abstract

Alternative splicing as a mean to control gene expression and diversify function is suspected to considerably influence drug response and clearance. We report the quantitative expression profiles of the human *UGT* genes including alternatively spliced variants not previously annotated established by deep RNA-sequencing in tissues of pharmacological importance. We reveal a comprehensive quantification of the alternative *UGT* transcriptome that differ across tissues and among individuals. Alternative transcripts that comprise novel in-frame sequences associated or not with truncations of the 5' and/or 3' termini, significantly contribute to the total expression levels of each *UGT1* and *UGT2* gene averaging 21% in normal tissues, with expression of *UGT2* variants surpassing those of *UGT1*. Quantitative data expose preferential tissue expression patterns and remodelling in favour of alternative variants upon tumorigenesis. These complex alternative splicing programs have the strong potential to contribute to interindividual variability in drug metabolism in addition to diversify the UGT proteome.

Introduction

Expansion of the human genome coding potential by alternative splicing (AS) provides multiple means to modulate gene expression and functions. At pharmacogene loci, AS is suspected to considerably influence drug response and clearance, and thus contribute to interindividual variability.[1-3] The human UDP glycosyltransferase (UGT) superfamily involves four families of enzymes (UGT1, UGT2, UGT3 and UGT8) that catalyze the addition of UDP-sugar residues to small lipophilic chemicals including many therapeutic drugs. UGT1 and UGT2 enzymes catalyze glucuronidation using UDP-glucuronic acid as a co-substrate and generally impair molecules' bioactivity. They play a critical role in regulating bioavailability of 55% of most prescribed drugs as well as important endogenous molecules including bilirubin, steroid hormones and bile acids.[4, 5] In contrast, two members of the UGT3 family, UGT3A1 and UGT3A2, use respectively UDP-N-acetylglucosamine and UDP-glucose/UDP-xylose to conjugate bile acids, steroids, and bioflavones.[6] The single UGT8 family member UGT8A1 uses UDP-galactose to galactosidate ceramide as well as bile acids.[7]

At present, AS is best characterized for the ten genes encoding UGT1 and UGT2 enzymes, from which as many as 180 transcripts may be derived.[8] This collection of mRNAs was reconstructed following parallel sequencing, which provided long reads that encompass several exon-exon junctions. It allowed a precise assignment of reads to the appropriate *UGT* loci as well as transcript reconstruction but this approach did not permit quantification of *UGT* transcripts. Over 90% of the newly discovered *UGT* transcripts display an open reading frame (ORF) potentially leading to alternative UGT proteins. Because this large collection of alternatively spliced *UGT* transcripts was previously

unannotated, they were overlooked by qPCR and microarray analysis as well as in publicly available deep sequencing RNA-Seq data. These previous transcriptomics studies might thus inaccurately predict *UGT* expression profiles. A comprehensive quantification of the alternative *UGT* transcriptome remains to be achieved to fully appreciate the impact of variant expression on glucuronidation activity and drug metabolism.

Here, using high throughput RNA-Seq applied in tissues of pharmacological importance, we provide an exhaustive quantitative expression portrait of the four families of UGT enzymes. This study reveals high levels of expression of multiple variants in normal hepatic, renal and intestinal/colon tissues, a tissue-preferential expression of some variants, and suggests a remodelling of the *UGT* transcriptome profiles in kidney and intestine/colon tumors. We further report interindividual variability in the expression of alternative *UGT* variants based on the analysis of an independent RNA-Seq dataset derived from 18 different individuals mapped on our complete *UGT* transcriptome.

Materials and Methods

RNA samples, sequencing libraries

RNA expression was measured in normal livers, kidneys, intestines and colons as well as in tumoral kidneys, and intestines/colons of mixed gender origin. For each organ, 9 individual RNA samples were used besides for normal intestine/colon RNA samples that were from 15 individuals. RNA samples and preparation of sequencing libraries was described recently.[8] Libraries were sequenced on an Illumina HiSeq 2500 system (McGill University and Génome Québec Innovation Center, Montreal, QC, Canada). Sequencing data are available in the GEO database (GSE82292).

Processing and analysis of sequencing data

FastQC (v0.11.2) was used for assessment of quality and basic metrics of paired-end datasets (2 x 100 bp), Bowtie[9] (v2.2.3) for mapping, Trimmomatic[10] for read trimming (v0.36) and TopHat[11] (v2.0.13; parameters: --no-novel-juncs) for alignment of trimmed reads on UCSC hg19 reference genome with annotations by Illumina iGenome and with the complete human *UGT* annotated loci.[8] Normalized expression was determined from mapped read files (.bam) with Cufflinks suite of tools (v2.2.1, parameters: --max-bundle-frags 10000000) and is given in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Cuffquant and Cuffdiff analysis were performed with the annotation used for the mapping step with default library normalization and cross-replicate dispersion estimation methods. Normalized expression levels were extracted using R (v3.1.2) and cummeRbund (v2.8.2).[12, 13] Sequencing depth averaged 39.9M reads per sample. The average normalized expression of replicates is provided. Expression of alternatively spliced *UGT* variants including those not previously

annotated was also measured in 18 normal liver and 18 kidney samples sequenced by the Pharmacogenomics Research Network (PGRN) (GSE70503),[2] by alignments of reads on our complete human *UGT* transcriptome and using Cufflinks for quantifications as described above.

Statistical correlations

Statistical correlations scores were determined by a Spearman correlation test using GraphPad Prism software (La Jolla, CA, USA).

Results

In our previous study, we used the pyrosequencing technology that produces longer reads combined to cDNA libraries enriched for *UGT* sequences to reconstruct the most complete *UGT* transcriptome landscape in human tissues[8]. However, this approach did not allow the quantification of *UGT* variant expression levels. In this report, global RNA-Seq was applied to ensure sufficient sequencing depth and fully characterize the quantitative expression pattern of 13 human *UGT* genes in relevant metabolic tissues of multiple pools of donors (at least 9 per tissue) based on the mapping of reads on the previously assembled *UGT* transcriptome. A first observation confirmed the expression and relative abundance of *UGT* transcripts previously unannotated using cDNA libraries enriched for *UGT* sequences (data not shown) and highlighted variability in expression of some *UGT* genes despite pooling of individual RNA samples (**Fig.1**). Data also revealed that several *UGT*s were among the most expressed genes (> 100 FPKM) in each organ**,** along with other important pharmacogenes (CYPs and transporters) (**Fig.2**). For instance, *UGT2B7* is among the high abundance genes expressed in the liver along with *CYP3A4* and *CYP2C9* in contrast to *UGT2B17* and *ABC* transporters that display lower hepatic expression (< 100 FPKM). Quantitative expression data of *UGTs*, *CYPs* and transporters genes are provided in supplemental Tables S1 and S2.

Seventeen of the 22 human *UGT* transcripts encoding canonical and functional enzymes (*referred to as v1 transcripts according to Human Genome Variation Society HGVS nomenclature*)[14] were substantially expressed (> 2 FPKM) in at least one of the three tissue examined (**Fig.3; Table S1**). Highest expression was measured in the normal liver and especially for *UGT2B* genes (> 400 FPKM for *UGT2B4*, *UGT2B7*,

UGT2B10 and *UGT2B15*). Intestine/colon expressed the broadest range of *UGTs* (n=13) whereas expression in the kidney was limited to a relatively small subset (n=6). *UGT2A3* was the most abundant gene of the *UGT2A* subfamily expressed, especially in intestine/colon and kidney tissues. *UGT3A1* was detected in all three tissues whereas *UGT8* was absent from the liver at levels below those of most *UGT1* and *UGT2* genes.

Despite the limited number of samples studied (on average 9 per tissue), another observation is the globally reduced *UGT1* and *UGT2* total expression in tumors relative to normal tissues, revealing a shift in the relative expression of canonical v1 and alternative variants often in favor of the latter (**Fig.3**). Alternative transcripts contributed significantly to the total expression levels of each *UGT1* (ranging from 5% to 61% in normal tissues and 8% to 88% in tumors) and *UGT2* gene (ranging from 8% to 100% in normal tissues and 24% to 96% in tumors) (**Fig.3; Table 1, S1**). Expression of *UGT2* alternative variants exceeded that of *UGT1* alternates, with *UGT2B7*, *UGT2B10* and *UGT2B15* being the most exposed to AS, in terms of alternative transcript abundance. Several variants were substantially expressed, sometimes at levels of the canonical variant *v1* particularly in the kidney and intestine/colon tissues (**Fig.3**).

To gain additional information on the interindividual expression levels, we analyzed expression of *UGT* variants using an independent set of 18 normal liver and 18 normal kidney samples (GSE70503).[2] This analysis consisted of the assignment of sequencing reads to the appropriate *UGT* loci based on our assembled complete *UGT* transcriptome including *UGT* variants not previously annotated. Data revealed relative abundances similar to those of the pools and indicated a wide range of expression of alternate *UGT* variants between individuals (coefficient of variation (CV) ranging from 77-358% for selected alternative variants) (**Fig.4**). It also exposed a lack of correlation between some

canonical and alternate expression, supporting the notion of differentially regulated AS events (**Fig.5**). In support, some alternative variants also displayed tissue- or tumor-preferential expression (detailed below).

A summarized portrait of expressed alternative variants derived from individual *UGT* genes is provided below. The newly discovered *UGT* transcripts that were experimentally validated are named with the prefix 'v' (*UGT_v#*), while the remaining *UGT* transcripts are designated with the prefix 'n' (*UGT_n#*)[8].

The UGT1 gene

A single *UGT1* gene encodes all nine UGT1A enzymes by mean of alternative promoters and exons 1 associated to four common exons in canonical variants 1. The *UGT1* gene was well expressed in all tissues, with preferential tissue expression of specific transcripts (**Fig.3, S1**). The liver expressed highest levels of *UGT1A1, 1A3, 1A4, 1A6,* and *1A9*, but *1A5, 1A7, 1A8 and 1A10* were undetected. In the kidney, *UGT1A6* and *UGT1A9* were the unique *UGT1As* measured. In intestine/colon, *UGT1A1* and *UGT1A10* were by far the most abundant although all *UGT1A* but *UGT1A7* and *UGT1A9* were detected above 2 FPKM. The main AS event included the alternate terminal exon 5b that generates two types of transcripts *v2* and *v3* per *UGT1A* (18 alternative *v2/v3* transcripts in total) (**Fig.6, S1**). They represented 8%, 7% and 13% of all *UGT1* transcripts in the normal liver, kidney and intestine/colon tissues, respectively (**Fig.S1**). In tumor tissues, reduced overall *UGT1A* expression relative to normal tissues was observed, particularly in intestine/colon where expression was nearly shut down (**Fig.3, S1**). However, in proportion to total *UGT1* expression, *v2/v3* expression rose to 16% in kidney and to 22% in intestine/colon tumors (**Fig.6**). Other *UGT1* variants were detected, namely *UGT1A6_n1*, an exon 1-truncated variant transcribed from an

alternative promoter that is well expressed in each tissue but most abundant in the kidney, *UGT1A1_n1* and *UGT1A1_n3*, produced by intronization of 141 nt in exon 1 and 31 nt in exon 4 respectively, as well as *UGT1A4_n2*, devoid of exon 2 (**Fig.S1**).

The UGT2B4 gene

UGT2B4 gene was one of the most abundant *UGT* in the liver, representing 20% of total *UGT* expression, but marginally expressed (< 2 FPKM) in intestine/colon and kidney tissues (**Fig.3, 7A**). Multiple alternative *UGT2B4* variants were well expressed, and comprised 16% of the *UGT2B4* liver transcriptome. Variants *v9*, *v10* and *v11* represented one abundant group (**Fig.7A**), characterized by exon 1 truncations resulting from intronization of parts of this exon, as well as variants *v5, n2* and *n3* that present intronization of parts or all of exon 5. Two other expressed variants were *v4* and *n4*. Each carry an alternative 5' exon 6 (6c and 6d) that encodes respectively short 8- and 6-amino acid (aa) C-termini rather than the 91 aa encoded by the canonical exon 6.

The UGT2B7 gene

Transcripts encoding UGT2B7 were by far the most abundant group of *UGT* transcripts, representing respectively 29%, 49% and 13% of the *UGT* transcriptome in normal liver, kidney and intestine/colon tissues (**Fig.3**). The *UGT2B7* transcriptome was also the most complex based on a previous study, with 44 distinct transcripts detected[8] and 16 of those expressed at $\geq$ 2 FPKM in at least one tissue. The alternative *UGT2B7* transcriptome collectively constituted 8%, 33% and 9% of the canonical *UGT2B7_v1* expression in the normal liver, kidney and intestine/colon (**Table 1**). This proportion of variants rose considerably in tumor tissues, especially in the kidney (**Table 1**; **Fig.8**). Of all alternative *UGT2B7* variants detected, a largely expressed class (collectively termed

UGT2B7_Δexon1) were produced by skipping exon 1 owing to intronization events with or without the use of an alternative promoter (**Fig.8**). This class represented 1.5%, 16%, 2.1% of the *UGT2B7* transcriptome in the normal liver, kidney and intestine/colon, respectively. In kidney and intestine/colon tumors, expression rose to 25% and 21% respectively (**Fig.8**). The alternate UGT2B7 isoform encoded by these variants would be devoid of most of the substrate-binding domain. One further class of variants are derived from an alternative *UGT2B7* promoter but are devoid of the canonical exons 1-6. Expression levels were remarkably high in the kidney, representing 9.8% and 67% of *UGT2B7* transcripts in normal and tumoral kidney respectively. One other abundant class of *UGT2B7* variants have altered 3' terminal ends. *UGT2B7_v6* includes an additional exon 6b, whereas *UGT2B7_v7* lacks exon 5 due to exon skipping (**Fig.8**). In normal tissues, these variants represented nearly 5% of all *UGT2B7* transcripts and were drastically reduced in tumor tissues. Lastly, a group of *UGT2B7* variants (*n4*, *n5* and *n6*) that include a novel in-frame sequence encoded by exon 2b were expressed in normal livers and kidneys.

The UGT2B10 gene

Similar to *UGT2B4*, *UGT2B10* expression was high in the liver but barely detectable (< 2 FPKM) in kidney and intestine/colon (**Fig.3**). The 10 alternative *UGT2B10* variants recently described[8] were highly expressed in the liver, making up 62% of total *UGT2B10* expression (**Table 1, Fig.7B**). The most abundant variants (*n9/n10*) lack a large part of the canonical exon 6 due to an intronization event but comprise a novel exon 6c that extends the open reading frame. The C-terminal 19 aa of the UGT2B10 enzyme would be replaced by 10 or 65 aa sequences unique to the novel isoforms n9 and n10, respectively. They represented 47% of the total hepatic *UGT2B10* expression. A well-

expressed novel variant, *UGT2B10_n8,* arises from usage of an in-frame alternative exon 6b encoding an isoform lacking the C-terminal 92 aa replaced by a unique 21-aa sequence. Expression of *UGT2B10_n1-n5*, with intronization of parts of exons 1 or 2, was also appreciable. These transcripts encode isoforms with a truncated substrate-binding domain.

The UGT2B15 gene

Highest expression of the *UGT2B15* gene was measured in the liver (**Fig.3**). Alternative variants, which represented 6% of *UGT2B15* hepatic expression, are produced by intronization of exon 1 (*n7/n8*), exon 4 (*n5*) or exons 4/5 (*n3*) (**Fig.9A**). One remarkable observation was the absence of canonical *UGT2B15_v1* transcripts in normal intestines/colons, and the nearly exclusive expression of two alternative variants (*n2* and *n7*). In contrast to *n7* that lacks exon 1, *n2* lacks canonical exon 6 that was replaced by the novel exon 6b encoding an alternative 35 aa C-terminus in place of the canonical 92 aa. In intestine/colon tumors, *UGT2B15* expression was almost totally repressed.

The UGT2B17 gene

UGT2B17 gene was the most expressed of all *UGT* genes in the intestine/colon but low in the liver (6 FPKM) and undetected in the kidney (**Fig.3**). Expression of variants (**Table 1**) was largely contributed by the *n9* variant (23% of total *UGT2B17* transcriptome). In this variant, exons 3-6 are skipped and replaced by a novel exon 2b (unrelated to *UGT2B7* exon 2b described above) with an in-frame coding sequence that introduces a unique 13 aa hydrophobic tail to the substrate-binding domain of this C-terminally truncated isoform (**Fig.9B**). Other well expressed variants were *n6*, created by skipping of exon 3, and *n5* with partial intronization of exon 1, and constituted 3.3% and 0.9% of the intestinal *UGT2B17* transcriptome, respectively. Tumoral expression of *UGT2B17*

was globally repressed by 94%. However, downregulation of alternative variants was less pronounced, therefore representing 26% of *UGT2B17* gene expression in tumors.

The UGT2A3 gene

The *UGT2A3* gene is expressed at significant levels in all three surveyed tissues, with highest expression in the intestine/colon (**Fig.3,10**). This 6 exon-gene is widely exposed to AS, with 20 variants.[8] However, *n2*/*n9, n3, n17 n18* and *n20* transcripts predominated the *UGT2A3* alternative transcriptome, representing over 10% of intestine/colon transcripts. Shifts in acceptor splice sites in exon 2 characterize *n2/n9*, whereas *n17* includes a novel exon 2e but lacks the canonical exons 3-6. The *n3* variant is created by a splice donor shift that extends exon 2 by 18 nt thus introducing a short novel internal in-frame sequence. *UGT2A3_n18/n20* variants are devoid of 5' terminal exons 1 and 2 (**Fig.10**). Expression of *UGT2A3* gene was significantly reduced in intestine/colon tumors (-96%). By contrast, the relative expression of alternative transcripts in tumor kidney rose to 29%, despite the reduced *UGT2A3_v1* expression in tumors (-44%) (**Table S1**). In the liver, total *UGT2A3* expression was mostly of the canonical *v1* variant, with alternative variants (all < 2 FPKM) representing 14% of total *UGT2A3* expression.

The UGT3 and UGT8 genes

Expression of the *UGT3A1* gene was appreciable in normal livers and kidneys but low in intestines/colons and repressed in tumors whereas expression of *UGT3A2* was not detected in any of the three tissues (**Fig.3**). Based on NCBI annotations (release 107), reported alternative *UGT3A1* variants were below 2 FPKM in all tissues (**Table S1**).

UGT8 was appreciably expressed in normal kidneys and intestines/colons but undetected in livers (**Fig.3**). The alternative *UGT8* transcript with a truncated non-coding exon 1 but extended exon 2 (NM_003360) that encodes nonetheless the UGT8 enzyme,

was significantly expressed along with the canonical variant. *UGT8* expression was modestly perturbed in tumors. It remained unchanged in kidney tumors, and was partially reduced in intestines/colons such that *UGT8* was the second most expressed *UGT* in intestine/colon tumors.

Discussion

This work describes quantitative and tissue-specific expression of alternatively spliced human *UGT* transcripts by global RNA-seq in pharmacogenetically important tissues, including the liver, kidney and intestine/colon, which are primarily involved in drug metabolism, secretion and reabsorption.

Strength of this study is the mapping of RNA sequencing short reads on a comprehensive human *UGT1* and *UGT2* transcriptome comprising previously unannotated *UGT* alternative transcripts, which enabled a precise assignment to the appropriate *UGT* loci despite high sequence similarity as well as quantification of expression. It revealed that several *UGT1* and *UGT2* genes are among the top 5% most expressed genes, such as *UGT2B7*, highly expressed in all three normal metabolic tissues along with some *CYPs* and *ABC* transporters. It is worth noting that expression of *UGT3A1* in liver and kidney as well as renal expression of *UGT8* were well in the range of multiple *UGT1* and *UGT2* genes, supporting an important contribution of these enzymes to the conjugation of lipophilic substrates in these tissues.[7, 15, 16]

The limited number of individuals initially sequenced from normal tissues (3 pools of at least 3 RNA samples) is a significant limitation of our study. However, data were consistent with those of an independent RNA-seq dataset of 18 individual normal liver and kidney tissues. Moreover, genotypes were not available for these samples and may account for under or over representation of some *UGT* transcripts such as for *UGT2B17* and *UGT2B28* that are among most common deleted genes.[17] Still, expression data are consistent with lower expression of *UGTs* in kidney and colon tissues as previously reported using other approaches, at the mRNA and protein levels.[18-22] Our preliminary

observations also suggest a shift in the transcript profiles in kidney and intestine/colon tumors that requires additional investigations using a larger set of well-characterized tumor samples.

Alternative *UGT* mRNAs constitute a substantial proportion of the total *UGT* transcriptome, averaging 21% of all *UGT* transcripts in normal tissues with the vast majority retaining an ORF in frame with the *UGT* sequences. Expression levels of many alternative variants are in the moderate to high expression range (> 20 FPKM). These observations support that *UGT* AS programs are constitutively ongoing in normal tissues and imply that they may be associated with physiological processes. Moreover, low expression may not indicate marginal expression but possibly prominent expression in a specific tissue substructure, which is not perceived by RNA-Seq on whole tissues[23]. A recent RNA-Seq study of normal tissues exposed that, like *UGTs*, the vast majority (72%) of 389 pharmacogenes displayed clear evidence of being alternatively spliced[2]. An assignment of reads produced by this other RNA-Seq project onto our comprehensive *UGT* transcriptome map that includes *UGT* transcripts not previously annotated, supports our conclusions on the extensive expression of spliced variants in normal liver and kidney tissues (**Fig.4**). This analysis further allowed establishing a wide range of expression of alternate *UGT* variants between individuals, suggesting a meaningful contribution of AS to the variability in drug metabolism. It also exposed a lack of correlation between some canonical and alternate expression, supporting the notion of differentially regulated AS events.

Our preliminary findings with tumor samples also suggested a reprogramming of *UGT* expression. In tumors derived from the intestine and colon, this was characterized by a globally reduced overall *UGT* expression whereas in kidney tumors, we observed a

considerably enhanced relative expression of alternative variants compared to normal kidney tissues. While these preliminary data are well in line with the reduced UGT1A and UGT2B protein expression and glucuronidation recently reported in tumor kidneys and colons,[3, 18, 20, 24] they also suggest a preferential and regulated expression of alternative variants. These results raise the possibility of a physiological contribution of AS to pathogenic states such as cancer. Our recent studies support that some of these alternative *UGT* variants with new in-frame sequences are functional. They modified cell detoxification activity by positively or negatively regulating UGT enzymes, and further induced a significant rewiring of cell metabolism with impact on cancer cells phenotype.[25, 26] We also reported a switch in alternative splicing at the *UGT2B7* locus toward functional enzyme upon maturation in the kidney and reversal of this process in kidney tumors, considerably modifying tissular glucuronidation of drug.[20, 24] This supports the notion that variation in the expression levels of *UGT* genes induced by changes in alternative splicing has the potential to affect drug response.

Expressed alternative *UGT* variants may be classified in a small number of categories namely 1) truncated/absence of exons encoding substrate-binding domain, 2) truncated/absence of exons encoding co-substrate/transmembrane domains, 3) inclusion of alternative exons and 4) inclusion of novel in-frame sequences.[8] Quantitatively, the first two categories are highly and about equally expressed for the *UGT2B* family members. In turn, *UGT1* and *UGT2A3* alternate variants rather fall predominantly in the co-substrate/transmembrane domains truncated group. Multiple variants carry novel sequences owing to splice shifts extending an exon with adjacent, usually intronic sequence, to partial exon intronization or less frequently to inclusion of a

novel exon. It should be noted that most transcripts are not predicted to undergo non-sense mediated decay according to the 50-bp rule.[27]

The AS program of individual *UGT* genes was differentially regulated in surveyed tissues, with a preferential tissue expression for several variants. For instance, *UGT2B7* AS is far more prominent in the kidney (33%) than in the liver (8%) despite a high hepatic expression of the canonical *UGT2B7* variant. This is exemplified by a group of transcripts lacking exon 1 that was preferentially expressed in the kidney (24 % of v1 expression) with only minor expression in the liver (1.6%). Yet, AS in the liver presents high expression of variants of *UGT2B4* and *UGT2B10*. The regulated expression of variants is further supported by the use of alternative promoters. For instance, *UGT2B7* variants transcribed from the alternative promoter in exon 1a are significantly expressed in normal and tumor kidneys but undetected in the liver.

We also observed recurrent patterns of AS transcripts among *UGT* genes, lending support to the potential production of variant UGT proteins sharing similar primary structure and putative domain. Examples include the experimentally validated *UGT1* variants lacking exon 2, and the corresponding region deleted in *UGT2A3*, *UGT2B15* and *UGT2B17* lacking exon 3,[8,28] as well as the skipping of exon 5 common to variants of *UGT2B4*, *UGT2B7* and *UGT2B10*. In addition, a significant number of novel sequences, in frame with the *UGT* coding sequence and unique to one *UGT*, are introduced by novel exons. This class of variants include two variants of *UGT2B7*, whose modest expression level may reflect a localized expression, and two *UGT2B10* variants, whose hepatic expression is in the same range than the canonical variant *v1*.

Clearly, the next challenge will be to translate sequences into functions whereas current evidence supports that AS at *UGT* loci produces proteins with antagonistic, activating or

even potentially divergent functions.[25, 26] For instance, usage of the alternate terminal exon 5b in *UGT1* transcripts and alternate terminal 6b in *UGT2B7* transcripts create classes of UGT proteins (UGT1A_i2 and UGT2B7_i2) that are lacking glucuronic acid transferase activity for common substrates of these enzymes. Yet, they are able to oligomerize with UGT enzymes to considerably reduce their conjugation activity.[1, 3, 25, 28, 29] Moreover, alternative functions of variant UGT1A_i2 proteins in the regulation of redox and glycolytic enzymes via protein-protein interactions have been reported, supporting roles diverging from conjugation reactions.[25, 30] This is of particular interest in the context of cancer, where a preferential expression of some *UGT* alternates is observed. Many cancer-related genes are regulated by AS, with the encoded proteins often involved in all major aspects of cancer cell biology.[31]

The diversity and abundance of alternative *UGT* variants are suggestive of a physiological role, possibly in the control of coordinate cellular responses induced by small molecules and in determining exposure to xenobiotics. Given that several alternative variants are expressed at “biologically significant” levels (> 20 FPKM)[23] and some in the context of low canonical expression, alternative functions independent of UGT enzymes may be anticipated. There is great interest in discovering the impact of AS on the *UGT* transcriptome complexity in normal functioning of metabolizing organs and in understanding how this regulatory and diversification mechanism contributes to normal physiology, tumorigenesis and drug response.

Acknowledgements

This work was supported by the Canadian Institutes of Health Research (CIHR) [MOP-42392, to CG]; and the Canada Research Chair in Pharmacogenomics (Tier I) to CG. The authors wish to thank the Genomics Analysis Platform of the *Institut de Biologie Integrative et des Systèmes* (IBIS; Laval University, Québec, Canada; specifically Brian Boyle) for their help with RNA-Seq experiments and Anne-Marie Duperré for help with preparation of figures. We also acknowledge the excellent artwork by France Couture.

Competing Interests

The authors declare no conflict of interest

Supplementary information is available at *The Pharmacogenomics Journal's* website

References

1. 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:** 3600-3607.
2. Chhibber A, French CE, Yee SW, Gamazon ER, Theusch E, Qin X*, et al*. Transcriptomic variation of pharmacogenes in multiple human tissues and lymphoblastoid cell lines. *Pharmacogenomics J* 2016.
3. Menard V, Collin P, Margaillan G, Guillemette C. Modulation of the UGT2B7 enzyme activity by C-terminally truncated proteins derived from alternative splicing. *Drug Metab Dispos* 2013; **41:** 2197-2205.
4. Guillemette C, Levesque E, Rouleau M. Pharmacogenomics of human uridine diphospho-glucuronosyltransferases and clinical implications. *Clin Pharmacol Ther* 2014; **96:** 324-339.
5. Stingl JC, Bartels H, Viviani R, Lehmann ML, Brockmoller J. Relevance of UDP-glucuronosyltransferase polymorphisms for drug dosing: A quantitative systematic review. *Pharmacol Ther* 2014; **141:** 92-116.
6. Meech R, Rogers A, Zhuang L, Lewis BC, Miners JO, Mackenzie PI. Identification of residues that confer sugar selectivity to UDP-glycosyltransferase 3A (UGT3A) enzymes. *J Biol Chem* 2012; **287:** 24122-24130.
7. Meech R, Mubarokah N, Shivasami A, Rogers A, Nair PC, Hu DG*, et al*. A novel function for UDP glycosyltransferase 8: galactosidation of bile acids. *Mol Pharmacol* 2015; **87:** 442-450.

8. Tourancheau A, Margaillan G, Rouleau M, Gilbert I, Villeneuve L, Levesque E*, et al*. Unravelling the transcriptomic landscape of the major phase II UDP-glucuronosyltransferase drug metabolizing pathway using targeted RNA sequencing. *Pharmacogenomics J* 2016; **16:** 60-70.

9. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9:** 357-359.

10. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30:** 2114-2120.

11. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013; **14:** R36.

12. Goff L, Trapnell C, Kelley D. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data.. *R package version 282* 2013.

13. R Development Core Team, R.C. (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; URL: https://www.R-project.org.

14. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J*, et al*. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* 2016; **37:** 564-569.

15. MacKenzie PI, Rogers A, Elliot DJ, Chau N, Hulin JA, Miners JO*, et al*. The novel UDP glycosyltransferase 3A2: cloning, catalytic properties, and tissue distribution. *Mol Pharmacol* 2011; **79:** 472-478.

16. Mackenzie PI, Rogers A, Treloar J, Jorgensen BR, Miners JO, Meech R. Identification of UDP glycosyltransferase 3A1 as a UDP N-acetylglucosaminyltransferase. *J Biol Chem* 2008; **283:** 36205-36210.

17. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC*, et al*. Common deletion polymorphisms in the human genome. *Nat Genet* 2006; **38:** 86-92.

18. Bellemare J, Rouleau M, Harvey M, Popa I, Pelletier G, Tetu B*, et al*. Immunohistochemical expression of conjugating UGT1A-derived isoforms in normal and tumoral drug-metabolizing tissues in humans. *J Pathol* 2011; **223:** 425-435.

19. Court MH, Zhang X, Ding X, Yee KK, Hesse LM, Finel M. Quantitative distribution of mRNAs encoding the 19 human UDP-glucuronosyltransferase enzymes in 26 adult and 3 fetal tissues. *Xenobiotica* 2012; **42:** 266-277.

20. Margaillan G, Rouleau M, Fallon JK, Caron P, Villeneuve L, Turcotte V*, et al*. Quantitative profiling of human renal UDP-glucuronosyltransferases and glucuronidation activity: a comparison of normal and tumoral kidney tissues. *Drug Metab Dispos* 2015; **43:** 611-619.

21. Margaillan G, Rouleau M, Klein K, Fallon JK, Caron P, Villeneuve L*, et al*. Multiplexed Targeted Quantitative Proteomics Predicts Hepatic Glucuronidation Potential. *Drug Metab Dispos* 2015; **43:** 1331-1335.

22. Ohno S, Nakajin S. Determination of mRNA expression of human UDP-glucuronosyltransferases and application for localization in various human tissues by real-time reverse transcriptase-polymerase chain reaction. *Drug Metab Dispos* 2009; **37:** 32-40.

23. Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* 2013; **14:** 778.

24. Menard V, Levesque E, Chen S, Eap O, Joy MS, Ekstrom L*, et al*. Expression of UGT2B7 is driven by two mutually exclusive promoters and alternative splicing in human tissues: changes from prenatal life to adulthood and in kidney cancer. *Pharmacogenet Genomics* 2013; **23:** 684-696.

25. 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:** 29-36.

26. Rouleau M, Tourancheau A, Girard-Bock C, Villeneuve L, Vaucher J, Duperre AM*, et al*. Divergent Expression and Metabolic Functions of Human Glucuronosyltransferases through Alternative Splicing. *Cell Rep* 2016; **17:** 114-124.

27. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 1998; **23:** 198-199.

28. Bushey RT, Dluzen DF, Lazarus P. Importance of UDP-glucuronosyltransferases 2A2 and 2A3 in tobacco carcinogen metabolism. *Drug Metab Dispos* 2013; **41:** 170-179.

29. Bushey RT, Lazarus P. Identification and functional characterization of a novel UDP-glucuronosyltransferase 2A1 splice variant: potential importance in tobacco-related cancer susceptibility. *J Pharmacol Exp Ther* 2012; **343:** 712-724.

30. Audet-Delage Y, Rouleau M, Rouleau M, Roberge J, Miard S, Picard F*, et al*. Crosstalk between alternatively spliced UGT1A isoforms and colon cancer cell metabolism. *Submitted* 2016.

31. Shkreta L, Bell B, Revil T, Venables JP, Prinos P, Elela SA*, et al*. Cancer-Associated Perturbations in Alternative Pre-messenger RNA Splicing. *Cancer Treat Res* 2013; **158:** 41-94.

32. Levesque E, Menard V, Laverdiere I, Bellemare J, Barbier O, Girard H*, et al*. Extensive splicing of transcripts encoding the bile acid-conjugating enzyme UGT2B4 modulates glucuronidation. *Pharmacogenet Genomics* 2010; **20:** 195-210.

33. Menard V, Eap O, Roberge J, Harvey M, Levesque E, Guillemette C. Transcriptional diversity at the UGT2B7 locus is dictated by extensive pre-mRNA splicing mechanisms that give rise to multiple mRNA splice variants. *Pharmacogenet Genomics* 2011; **21:** 631-641.

Figure Legends:

Figure 1: *UGT* expression across pools of human tissues. Significant expression of many *UGTs* is observed from multiple donors (n=3 donors per pool for each organ besides normal intestine/colon, pools of 5 donors).

Figure 2: Abundance of *UGT* expression in the context of the human transcriptome. *Top*. Global profile of human gene expression levels is partitioned in 3 categories representing low (below 1 FPKM), moderate (1 to 100 FPKM) and high (above 100 FPKM) expression, measured in human liver, kidney and intestine/colon tissues by RNA-Seq. The percent of human genes expressed at the defined levels are indicated (%). *Bottom*. Expression levels of representative *UGT1* and *UGT2* loci in normal livers, kidneys and intestines/colons are shown in the context of expression levels of other pharmacogenes. ABC: ATP-binding cassette transporters; CYP: cytochrome P450; UGT: UDP-glycosyltransferases; SLCO: Solute carrier organic transporters. Expression levels of all *UGTs* are given in supplemental **Table S1** and those of *CYPs* and transporters in supplemental **Table S2**.

Figure 3: Tissular expression of canonical *UGTs* and alternative variants. Expression levels (FPKM) of each *UGT* established by RNA-Seq is given for normal (N) and tumor (T) tissues. Stacked bars represent expression of canonical (blue) and alternative (red) *UGT* transcripts.

Figure 4: Expression of selected alternative *UGT* transcripts in 18 individual liver and kidney tissues. RNA-Seq data (GSE70503)[2] was aligned to the hg19 reference from iGenome complemented with the human *UGT* transcriptome comprising alternative *UGT* variants previously unannotated to quantify their expression. **A. *UGT1* gene variants; B. *UGT2B* gene variants;** Selected variants expressed in normal livers and kidneys above 2 FPKM are shown. CV: coefficient of variation; L: liver; K: kidney

Figure 5: Correlation between canonical and alternative variant expression in 18 individual liver and kidney tissues. A. Total *UGT1A_v1* expression correlated well with total alternate *UGT1A_v2/v3* expression in liver and kidney. Correlation coefficients for each *UGT1A v1 vs v2/v3* are given in C. **B.** Hepatic and renal expression of *UGT2B7_v1* does not correlate with expression of alternative transcripts encoding *UGT2B7_Δexon1*. **C.** Correlations between canonical variants v1 and selected alternative variants. *UGT1As* comprise all *UGT1A_v1* or *v2/v3* expressed in specified tissue. Significantly correlated expressions are in bold. Spearman's rank-order correlations were determined with GraphPad for shown variants. RNA-Seq data: GSE70503[2].

Figure 6: Expression of the major *UGT1A* gene alternative transcripts *UGT1A_v2/v3*. Global expression of *UGT1A_v2/v3* is given relative to the *UGT1A* transcriptome in each tissue.

Figure 7: Expression of *UGT2B4* and *UGT2B10* canonical and alternative transcripts in the normal liver. A. *UGT2B4*; **B.** *UGT2B10*. *Left* Quantitative expression

of canonical (v1, blue bars) and alternative (red bars) variants (FPKM values are given for each variant; threshold > 2 FPKM). Previously characterized variants are labeled with the prefix “v” whereas novel variants are labeled with the prefix “n”. *Right* Schematic representation of canonical *v1* transcript (*v1* transcript exons are shown as numbered boxes; alternative exons and positions are shown above main transcript). Encoded enzyme (i1) with domain organization is schematized as a reference (light green: substrate binding domain; dark green/light green box: co-substrate binding domain/position of the UGT signature; zigzag line: transmembrane domain and charged lysine tail). Lines below indicate exons included in each expressed transcript whereas dashed lines indicate missing sequences in alternative transcripts. Novel appended sequences are represented by boxes (color according to exon source of novel sequences). Previously described isoforms are labeled according to the literature[32] whereas isoforms predicted from novel transcripts (*n*) are not labeled.

Figure 8: Expression of *UGT2B7* canonical and alternative variants. Figure description is as in Figure 7. Predicted isoforms are labeled according to Ménard et al[33] whereas isoforms predicted from novel transcripts (*n*) are not labeled. The *UGT2B7_Δexon 1* variants consist of *v4, v10, v12, v14, v16, v17, v20, v22, n1, n3, n12, n13, n14, and n15* that all encode the isoform i5. The *n4, n5, andn6* variants comprise the novel *UGT2B7* exon 2b. The “alt exon 1” group comprises variants *n16, n17, n18, n19, n20, n21 and n22* transcribed from exon 1a but lacking all canonical exons.

Figure 9: Expression of *UGT2B15* and *UGT2B17* canonical and alternative variants. **A.** *UGT2B15*; **B.** *UGT2B17*. Figure description as in Figure 7. Not

represented: Expression of *UGT2B15* and *UGT2B17* in the kidney because they were not detected and expression of *UGT2B17* in the liver because alternative variants were below 2 FPKM.

Figure 10: **Expression of *UGT2A3* canonical and alternative variants.** Figure description is as in Figure 7. Expression of *UGT2A3* in liver is not represented because only canonical v1 was > 2 FPKM.

Table 1. Expression of alternate variants relative to total transcripts per *UGT* gene (%)[1]

Organ		*UGT1A*									*UGT2A*			*UGT2B*							*UGT3*		*UGT8*
		1A1	1A3	1A4	1A5	1A6	1A7	1A8	1A9	1A10	2A1	2A2	2A3	2B4	2B7	2B10	2B11	2B15	2B17	2B28	3A1	3A2	
Liver	Normal	10	5	9	-	35	-	-	25	-	-	20	14	16	8	62	-	6	37	-	0.5	-	-
Kidney	Normal	-	-	-	-	34	-	-	8	-	-	-	11	-	33	-	-	-	-	-	1.8	-	26
	Tumoral	-	-	-	-	53	-	-	27	-	-	-	29	-	96	-	-	-	-	-	1.7	-	21
Intestine/ Colon	Normal	12	16	46	22	56	-	61	-	23	-	-	13	-	9	-	-	100	28	-	1.0	-	54
	Tumoral	-	-	-	-	88	-	-	-	-	-	-	24	-	28	-	-	90	26	-	-	-	47

[1]Corresponding expression values in FPKM are given in Table S1.
"-" indicates *UGT* expression < 2 FPKM

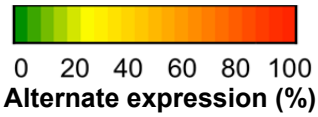


Fig. 1

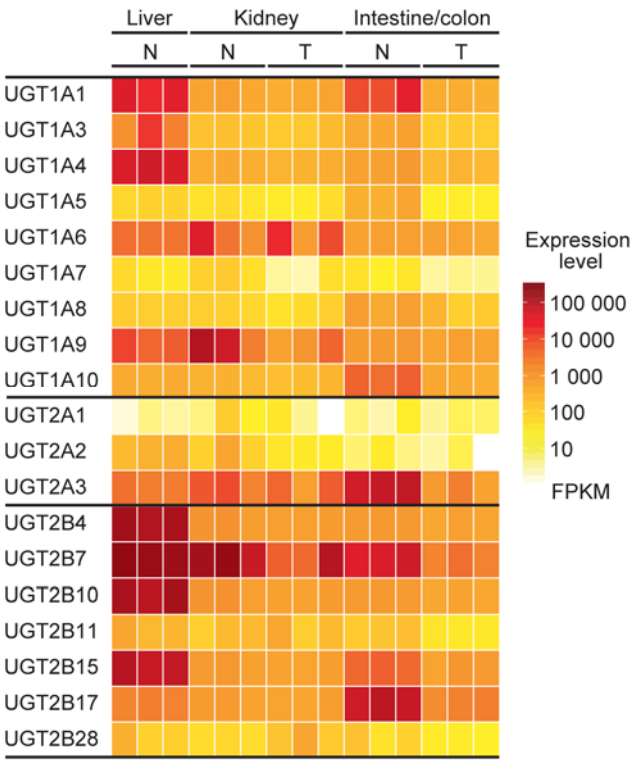

Liver
Kidney
Intestine/colon
N
N
T
N
T
UGT1A1
UGT1A3
UGT1A4
UGT1A5
UGT1A6
UGT1A7
UGT1A8
UGT1A9
UGT1A10
UGT2A1
UGT2A2
UGT2A3
UGT2B4
UGT2B7
UGT2B10
UGT2B11
UGT2B15
UGT2B17
UGT2B28
Expression level
100 000
10 000
1 000
100
10
FPKM

Fig. 2

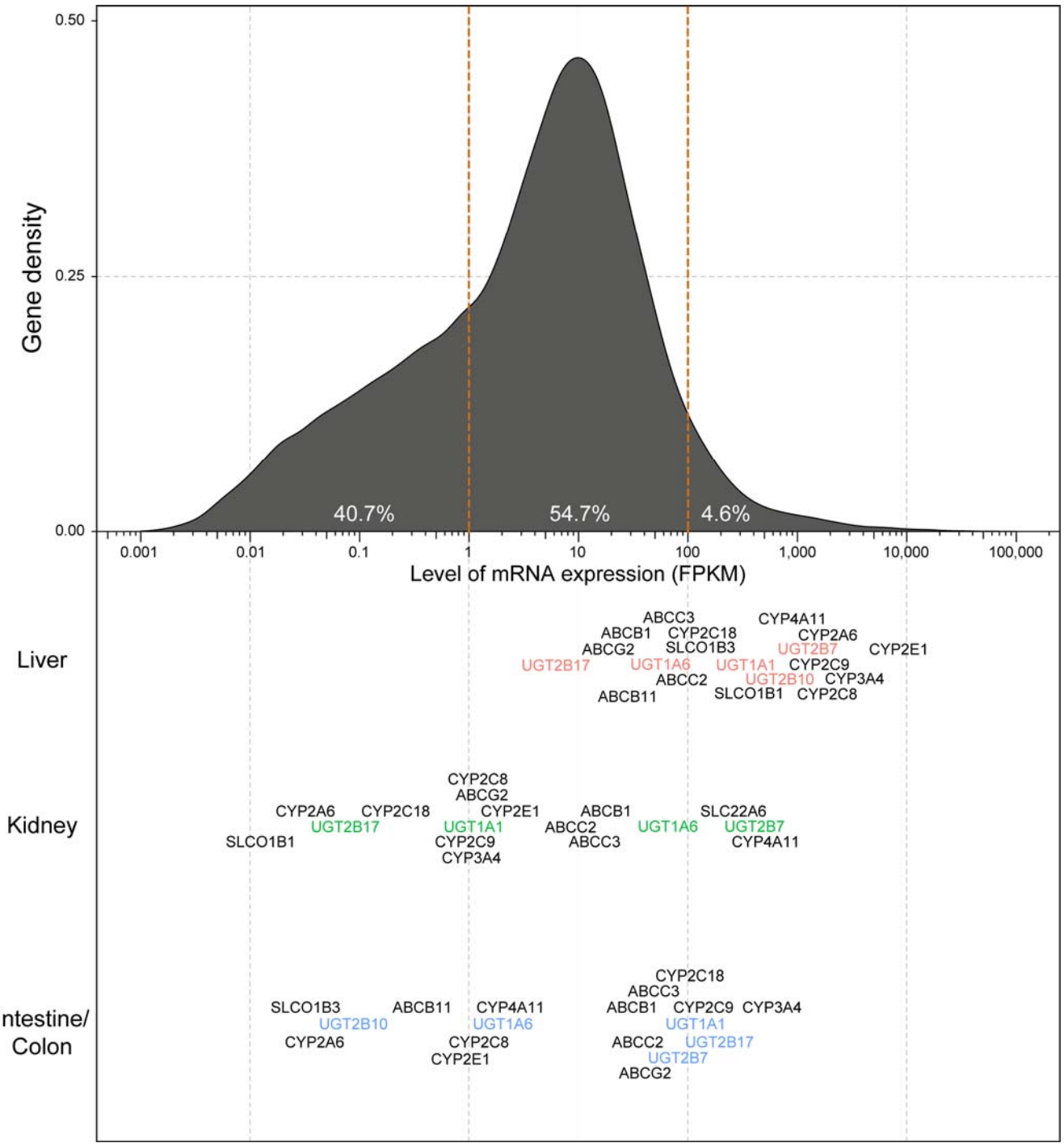
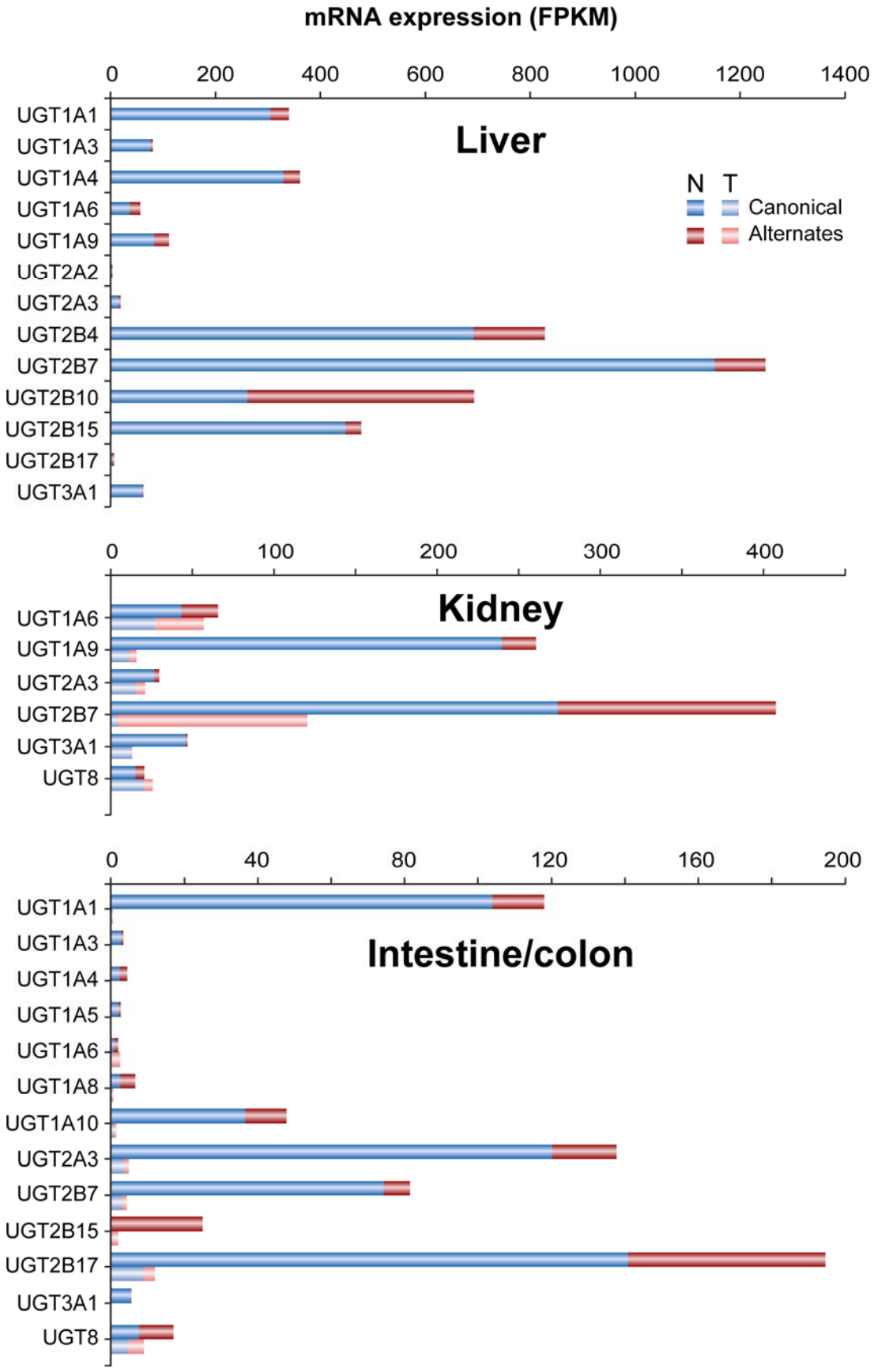
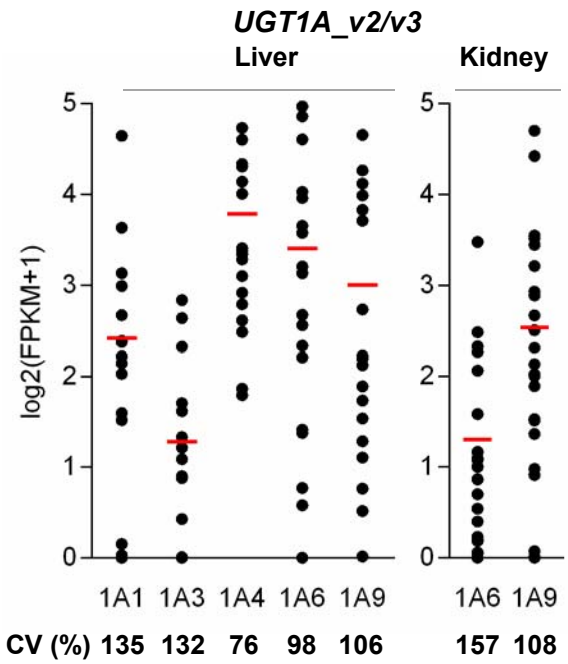

0.50
0.25
0.00
Gene density
40.7%
54.7%
4.6%
0.001
0.01
0.1
1
10
100
1,000
10,000
100,000
Level of mRNA expression (FPKM)
Liver
ABCC3
CYP4A11
ABCB1
CYP2C18
CYP2A6
ABCG2
SLCO1B3
UGT2B7
CYP2E1
UGT2B17
UGT1A6
UGT1A1
CYP2C9
ABCC2
UGT2B10
CYP3A4
ABCB11
SLCO1B1
CYP2C8
Kidney
CYP2C8
ABCG2
CYP2A6
CYP2C18
CYP2E1
ABCB1
SLC22A6
UGT2B17
UGT1A1
ABCC2
UGT1A6
UGT2B7
SLCO1B1
CYP2C9
ABCC3
CYP4A11
CYP3A4
Intestine/
Colon
CYP2C18
ABCC3
SLCO1B3
ABCB11
CYP4A11
ABCB1
CYP2C9
CYP3A4
UGT2B10
UGT1A6
UGT1A1
CYP2A6
CYP2C8
ABCC2
UGT2B17
CYP2E1
UGT2B7
ABCG2

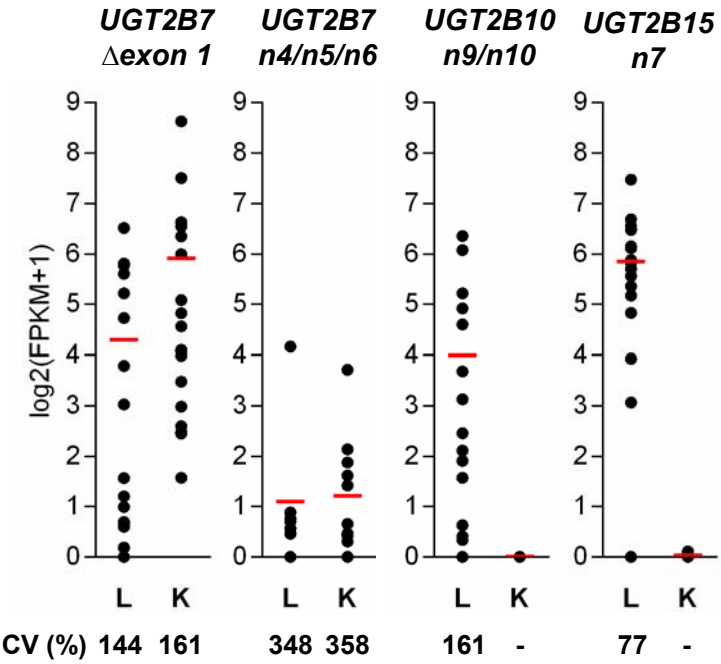
Fig. 3


mRNA expression (FPKM)
Liver
N T
Canonical
Alternates
UGT1A1
UGT1A3
UGT1A4
UGT1A6
UGT1A9
UGT2A2
UGT2A3
UGT2B4
UGT2B7
UGT2B10
UGT2B15
UGT2B17
UGT3A1
0 200 400 600 800 1000 1200 1400
Kidney
UGT1A6
UGT1A9
UGT2A3
UGT2B7
UGT3A1
UGT8
0 100 200 300 400
Intestine/colon
UGT1A1
UGT1A3
UGT1A4
UGT1A5
UGT1A6
UGT1A8
UGT1A10
UGT2A3
UGT2B7
UGT2B15
UGT2B17
UGT3A1
UGT8
0 40 80 120 160 200

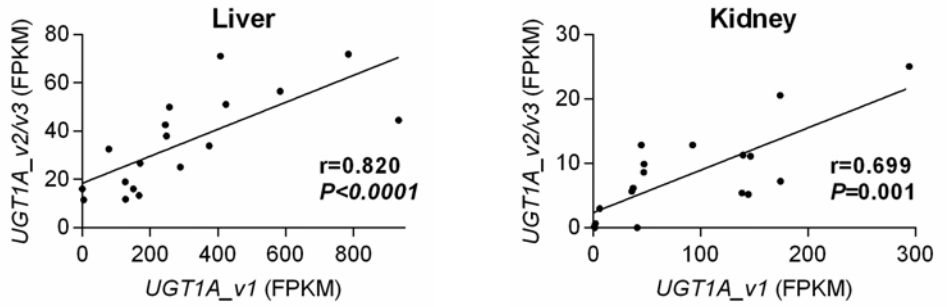
A.


UGT1A_v2/v3
Liver
Kidney
log2(FPKM+1)
0
1
2
3
4
5
1A1
1A3
1A4
1A6
1A9
1A6
1A9
CV (%) 135 132 76 98 106 157 108

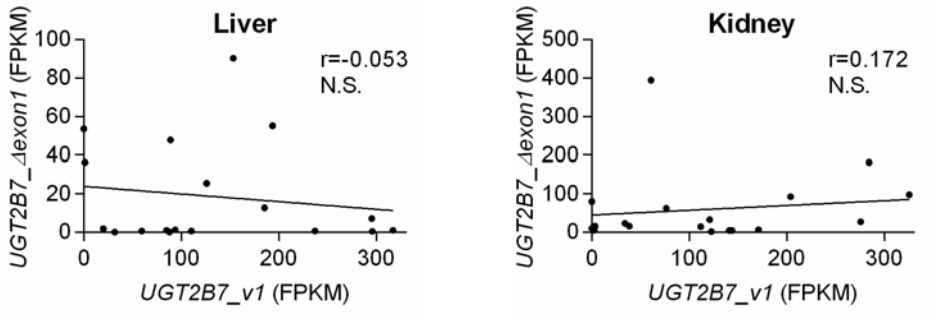
B.


UGT2B7 Δexon 1
UGT2B7 n4/n5/n6
UGT2B10 n9/n10
UGT2B15 n7
log2(FPKM+1)
0
1
2
3
4
5
6
7
8
9
L
K
CV (%) 144 161 348 358 161 - 77 -

A. UGT1A1_v1 vs. UGT1A1_v2/v3



B. UGT2B7_v1 vs. UGT2B7_Δexon1



C.

Tissue	Canonical UGT	Alternative UGT variant	r	*P-value*
Liver	1As	v2/v3	**0.820**	**<0.0001**
	1A1	1A1_v2/v3	**0.709**	**0.001**
	1A3	1A3_v2/v3	**0.741**	**<0.001**
	1A4	1A4_v2/v3	0.358	0.144
	1A6	1A6_v2/v3	**0.917**	**<0.0001**
	1A9	1A9_v2/v3	**0.781**	**<0.0001**
Kidney	1As	v2/v3	**0.699**	**0.001**
	1A6	1A6_v2/v3	0.351	0.153
	1A9	1A9_v2/v3	**0.781**	**<0.0001**
Liver	2B7	2B7_Δexon1	-0.053	0.836
Kidney	2B7	2B7_Δexon1	0.172	0.494

Fig. 6

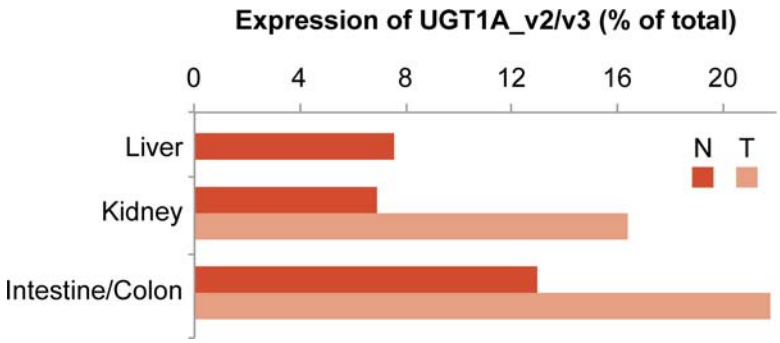

Expression of UGT1A_v2/v3 (% of total)
0
4
8
12
16
20
Liver
Kidney
Intestine/Colon
N
T

Fig. 7

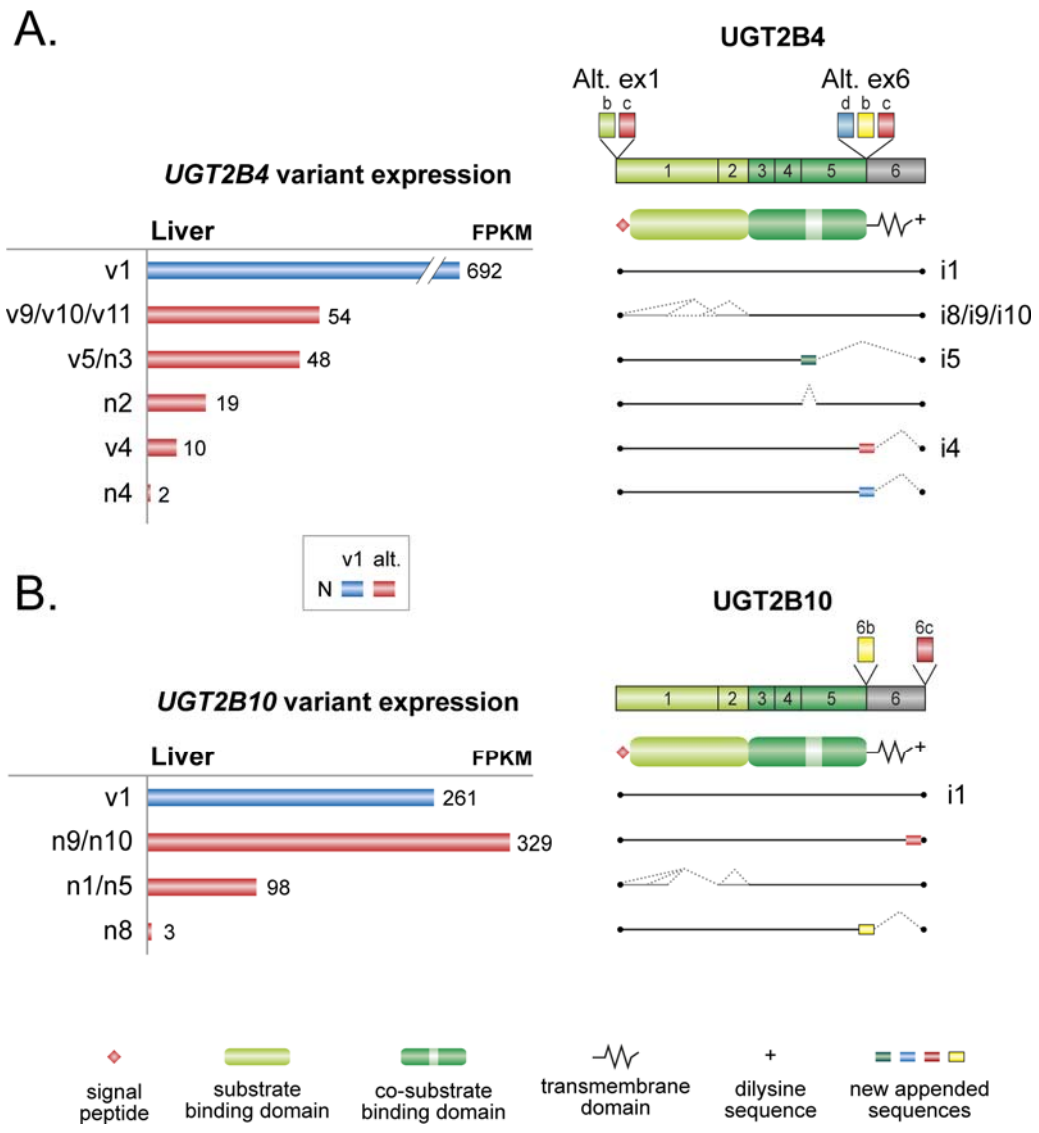

A.
UGT2B4
Alt. ex1
Alt. ex6
b c
d b c
1 2 3 4 5 6
UGT2B4 variant expression
Liver
FPKM
v1
692
v9/v10/v11
54
v5/n3
48
n2
19
v4
10
n4
2
i1
i8/i9/i10
i5
i4
v1 alt.
N
B.
UGT2B10
6b
6c
1 2 3 4 5 6
UGT2B10 variant expression
Liver
FPKM
v1
261
n9/n10
329
n1/n5
98
n8
3
i1
signal peptide
substrate binding domain
co-substrate binding domain
transmembrane domain
dilysine sequence
new appended sequences

Fig. 8

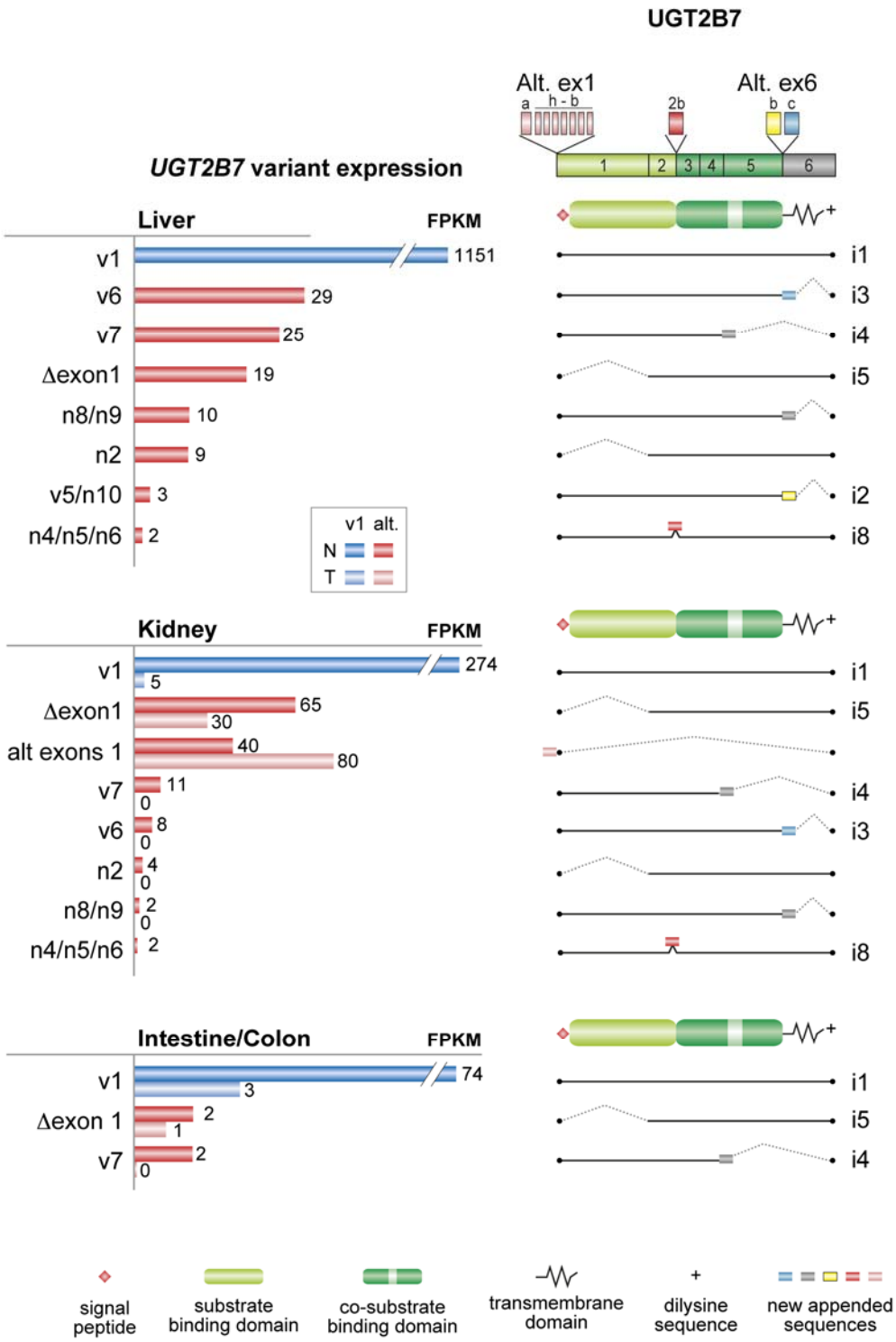

UGT2B7
Alt. ex1
Alt. ex6
a
h - b
2b
b c
1
2
3
4
5
6
UGT2B7 variant expression
Liver
FPKM
v1
v6
v7
Δexon1
n8/n9
n2
v5/n10
n4/n5/n6
1151
29
25
19
10
9
3
2
v1 alt.
N
T
i1
i3
i4
i5
i2
i8
Kidney
FPKM
v1
Δexon1
alt exons 1
v7
v6
n2
n8/n9
n4/n5/n6
274
5
65
30
40
80
11
0
8
0
4
0
2
0
2
i1
i5
i4
i3
i8
Intestine/Colon
FPKM
v1
Δexon 1
v7
74
3
2
1
2
0
i1
i5
i4
signal peptide
substrate binding domain
co-substrate binding domain
transmembrane domain
dilysine sequence
new appended sequences

Fig. 9

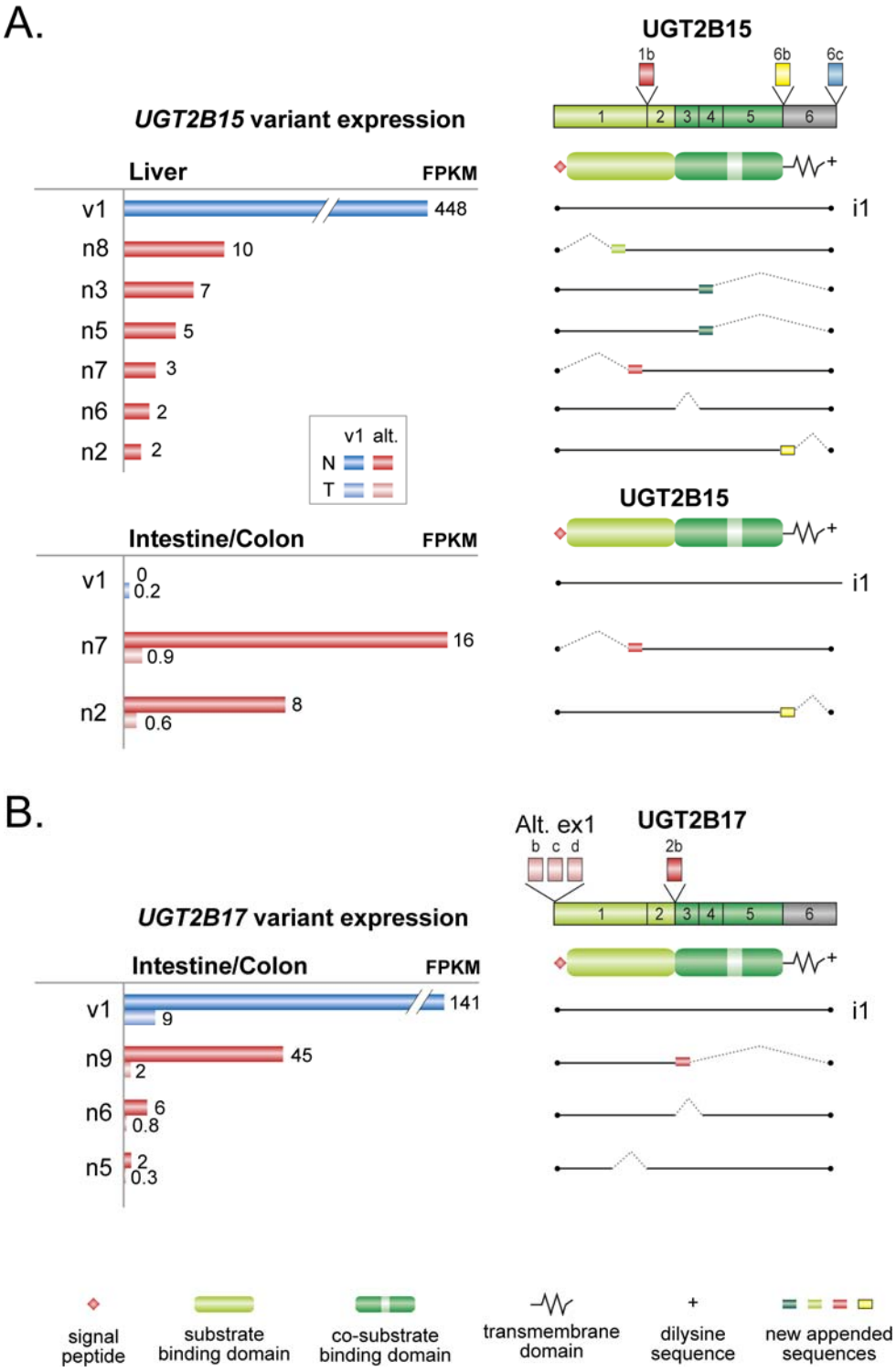
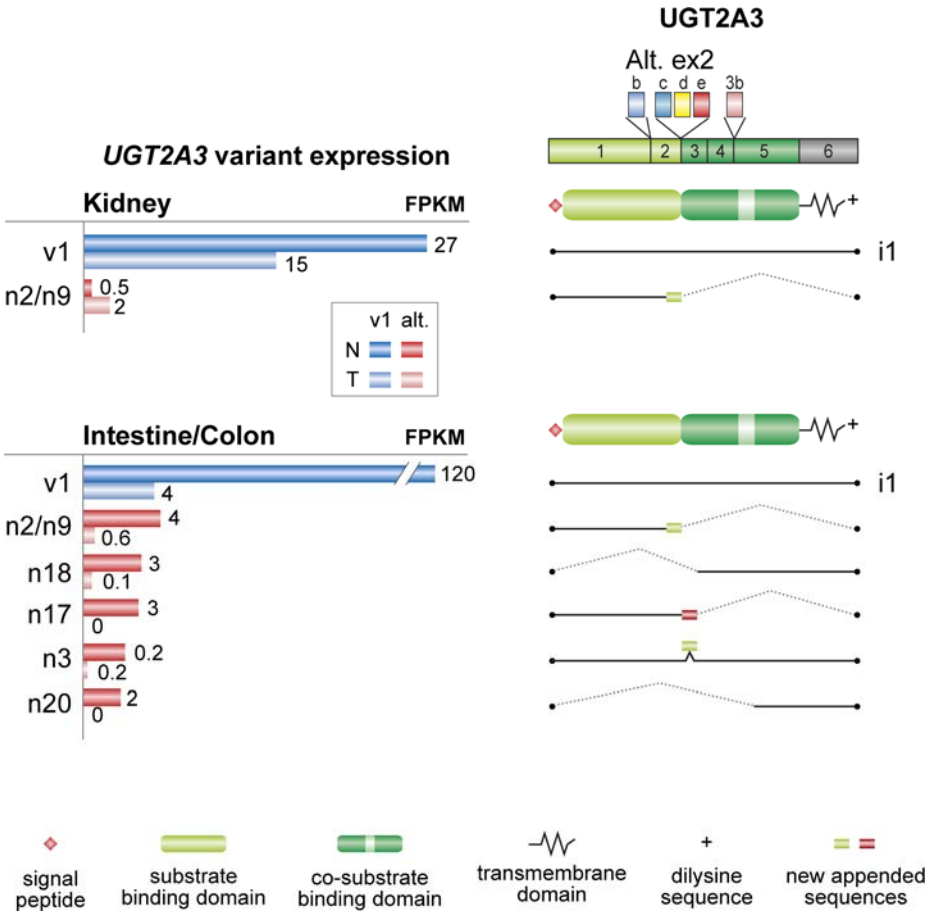

A.
UGT2B15
1b
6b
6c
1
2
3
4
5
6
UGT2B15 variant expression
Liver
FPKM
v1
n8
n3
n5
n7
n6
n2
448
10
7
5
3
2
2
v1 alt.
N
T
i1
Intestine/Colon
FPKM
v1
n7
n2
0
0.2
16
0.9
8
0.6
UGT2B15
i1
B.
Alt. ex1
b c d
UGT2B17
2b
UGT2B17 variant expression
Intestine/Colon
FPKM
v1
n9
n6
n5
141
9
45
2
6
0.8
2
0.3
i1
signal peptide
substrate binding domain
co-substrate binding domain
transmembrane domain
dilysine sequence
new appended sequences

Fig. 10


UGT2A3
Alt. ex2
b c d e 3b
1 2 3 4 5 6
UGT2A3 variant expression
Kidney
FPKM
v1
n2/n9
27
15
0.5
2
v1 alt.
N
T
Intestine/Colon
FPKM
v1
n2/n9
n18
n17
n3
n20
120
4
4
0.6
3
0.1
3
0
0.2
0.2
2
0
i1
i1
signal peptide
substrate binding domain
co-substrate binding domain
transmembrane domain
dilysine sequence
new appended sequences