UGT Genomic Diversity: Beyond Gene Duplication

Chantal Guillemette[1,2], Eric Lévesque[1], Mario Harvey[1], Judith Bellemare[1] and Vincent Menard[1]

[1]Pharmacogenomics Laboratory, CHUQ Research Center and Faculty of Pharmacy, Laval University, G1V 4G2, Quebec, Canada and [2]Canada Research Chair in Pharmacogenomics,

Short Title: Molecular Plasticity at the *UGT1A* and *UGT2B* loci.

Key words: Glucuronosyltransferase, *UGT*, Alternative splicing, Copy-number variation, Polymorphisms, Epigenetics; Pharmacogenomics.

Author to whom correspondence should be sent:
Chantal Guillemette, Ph.D.
Canada Research Chair in Pharmacogenomics
Pharmacogenomics Laboratory
CHUQ Research Center
2705 Boul. Laurier, T3-48
Québec, Canada, G1V 4G2
Tel. (418)654-2296; Fax.(418)654-2761
E-mail: Chantal.Guillemette@crchul.ulaval.ca

OUTLINE

1. Abstract

2. Introduction

3. The human UGT genes

3.1 The human UGT1A family members

3.2 The human UGT2 family members

4. Novel UGT-associated mRNA isoforms

4.1 UGT-derived EST clusters

5. Experimental evidence of new UGT1A-associated mRNAs

6. Alternative splicing at the UGT2 gene loci.

6.1 The UGT2B28 gene

6.2 The UGT2B7 gene

6.3 The UGT2B4 gene: A closer look at the genomic level.

6.4 Conclusions on the topic alternative splicing

7. Copy-number variations of UGT2B genes

8. Epigenetic regulation of UGT gene expression

9. Concluding Remarks

10. Acknowledgments

1. Abstract

The human UDP-glucuronosyltransferase (UGT) superfamily comprises enzymes responsible for a major biotransformation phase II pathway: the glucuronidation process. The UGT enzymes are located in the endoplasmic reticulum of almost all tissues where they catalyze the inactivation of several endogenous and exogenous molecules including bilirubin, sex-steroids, numerous prescribed drugs and environmental toxins. This metabolic pathway is particularly variable. The influence of inheritable polymorphisms in human UGT-encoding genes has been extensively documented, and was shown to be responsible for a fraction of the observed phenotypic variability. Other key genomic processes are likely underlying this diversity; these may include copy-number variations (CNV), epigenetic factors and newly discovered splicing mechanisms. This review will discuss novel molecular aspects which may be determinant to UGT phenotypes.

Abbreviation: *UGT*, UDP-glucuronosyltransferase; *SNP*, Single nucleotide polymorphism; CNV; Copy-number variations, *AS*; Alternative splicing.

2. Introduction

UDP-glucuronosyltransferase (UGT) enzymes are key metabolic proteins that prevent accumulation of possibly toxic lipophilic compounds and initiate their elimination through more hydrophilic vehicules, such as biliary and renal systems (Dutton, 1980). This process is performed by the addition of a hydrophilic sugar moiety (glucuronide) from uridine diphosphate glucuronic acid (UDPGA) by UGTs (Dutton, 1980). The number of molecules handled by UGT enzymes is structurally divergent and almost infinite (Mackenzie, Bock et al., 2005; Mackenzie et al., 1997; Radominska-Pandya, Czernik, Little, Battaglia, & Mackenzie, 1999; Tukey & Strassburg, 2000b, 2001; Wells et al., 2004). This is related to the ability of these enzymes to catalyze the transfer of the glucuronic acid moiety at multiple functional sites, such as carbonyl, carboxyl, sulfuryl, hydroxyl, and amine groups (Mackenzie et al., 1997; Radominska-Pandya, Czernik, Little, Battaglia, & Mackenzie, 1999; Tukey & Strassburg, 2000b). The substrates for UGTs may be endogenous compounds, environmental products, or synthetic drugs. One may speculate that the vast substrate diversity of UGT enzymes would necessitate several enzyme-producing genes, and would be highly energy consuming for cells. However, higher eukaryotic organisms have developed several strategies allowing protein diversity without increasing the complexity of their genetic background. Based on current development in the field of glucuronidation and recent discoveries, we now better appreciate that the human *UGT* genes are certainly a remarkable example of this concept. Besides, this pathway is particularly variable and currently both genetic and environmental factors have been shown to determine individual glucuronidation phenotypes. In fact, the number of cases partially explained by genetic factors described

in the literature has increased exponentially and they were mainly focused on single-nucleotide polymorphism (SNP) (especially regulatory and missense mutations), insertions and deletions (indels) of nucleotides, for review see (Burchell & Coughtrie, 1997; Burchell et al., 2000; Guillemette, 2003; Strassburg, Kalthoff, & Ehmer, 2008), and more recently on copy number variation (CNV) (McCarroll et al., 2006). In this review, we will discuss novel molecular aspects which may be determinant to UGTs protein diversity and glucuronidation phenotypes.

3. The human UGT genes

The human UGT are a very good example of evolving protein diversity, mainly arising from either exon or gene duplication, but also by alternative promoter usage and splicing events (Mackenzie, Bock et al., 2005; Mackenzie et al., 1997; Owens, Basu, & Banerjee, 2005; Ritter et al., 1992). Indeed, the nature of the eukaryotic gene structure has evolved to facilitate protein multiplicity. This is possible by combining independent smaller parts or modules in varying contexts. Eukaryotic genes, in contrast to lower complexity organisms, are mostly characterized by small coding sequences (exons) intercepted with large non-coding genomic regions (introns) (Boeger et al., 2005). The junction of successive exons by the excision of intronic sequences in pre-mRNA molecules (mRNA splicing process) conducts to mature mRNAs, which may then be translated to polypeptidic products (Gilbert, 1978). Therefore, exon assembly, which determines protein domains, is the basis of the proteome diversity. Mechanisms inducing exon mixing, such as segmental duplication and recombination (exon shuffling) of genomic

regions, together with alternative pre-mRNA splicing events, are the major driving forces of such an evolving protein complexity (Moore & Proudfoot, 2009).

The UGT enzymes have been classified in two major families in human, namely *UGT1A* and *UGT2* (subdivided in *UGT2A* and *UGT2B*) (UDP Glucuronosyltransferase Homepage http://som.flinders.edu.au/FUSA/ClinPharm/UGT/index.html). The *UGT1A* family members are all encoded by a single gene locus on human chromosome 2q37.1, while several likely duplicated genes, located on 4q13.2, define UGT2B and UGT2A3 enzymes (Mackenzie, Bock et al., 2005). On the other hand, UGT2A1 and 2A2 proteins, similarly to *UGT1A*, arise from one locus on chromosome 4q13.3 (Mackenzie, Bock et al., 2005) (Court, Hazarika, Krishnaswamy, Finel, & Williams, 2008).

3.1 The human UGT1A family members

The human *UGT1A* gene well exemplifies the usage of alternative promoter (consequently the first exon) to increase protein diversity. The complexity of the *UGT1A* locus has been firstly highlighted by the discovery of bilirubin transferase cDNAs and phenol transferase cDNAs having unique 5' sequences but identical 3' ends (Gong et al., 2001; Owens, Basu, & Banerjee, 2005; Ritter et al., 1992). Further experiments clearly indicated that these diverging cDNAs are encoded by one gene locus, and lead to transferases having various, sometimes overlapping and divergent substrate specificities. In fact, the human *UGT1A* locus is defined by 13 first exons, which are alternatively spliced to four (more recently five) common exons (**Figure 1**). From the 13 possible mRNA isoforms, nine conduct to functionally active polypeptides (UGT1A1, 1A3, 1A4, 1A5, 1A6, 1A7, 1A8, 1A9 and 1A10) and 4 pseudogenes (UGT1A2p, UGT1A11p and

UGT1A12p, UGT1A13p) (Gong et al., 2001; Ritter et al., 1992). The first exons most probably encode for the substrate-binding domain (Meech & Mackenzie, 1997a). Based upon their function and sequence similarity, UGT1A members are divided in two groups, the bilirubin-like-associated enzymes, which are defined by UGT1A1, UGT1A3, UGT1A4 and UGT1A5, and the phenol-like-associated group, encompassing UGT1A6, UGT1A7, UGT1A8, UGT1A9, and UGT1A10 (Owens, Basu, & Banerjee, 2005; Ritter et al., 1992). This family of enzymes has the ability to conjugate several xenobiotics, such as drugs and carcinogens (for a more detailed review see (Court, 2005; Dutton, 1980; Guillemette, 2003; Mackenzie, Bock et al., 2005; Mackenzie et al., 1997; Owens, Basu, & Banerjee, 2005; Ritter, 2000; Tukey & Strassburg, 2000a)), but are also involved in metabolism of endogenous compound, namely bilirubin (Ritter et al., 1992; Ritter, Crawford, & Owens, 1991), bile acids (Gall et al., 1999b; Ritter, Chen, Sheen, Lubet, & Owens, 1992; Trottier et al., 2006), serotonin (Court, 2005), androgenic and estrogenic steroid hormones (Belanger et al., 1998; Belanger, Pelletier, Labrie, Barbier, & Chouinard, 2003; Guillemette, Belanger, & Lepine, 2004; Lepine et al., 2004; Levesque et al., 1997; Starlard-Davenport, Lyn-Cook, & Radominska-Pandya, 2008). Nucleic acid sequence similarity clearly argues for ancestral exon duplication among UGT1A members. In addition, an inter-species sequence comparison of UGT1A first exons demonstrated that UGT1A1 and UGT1A6 are ancestrally common exons, and that other isoforms arose more recently in the evolution (Owens, Basu, & Banerjee, 2005).

A number of genetic polymorphisms, especially non-synonymous mutations, have been reported in UGT1A first exons (for a comprehensive list of UGT polymorphisms: http://www.ugtalleles.ulaval.ca). Considering the role of exon 1-derived UGT1A peptides

in substrate binding and the continuously changing environment, these coding polymorphisms indicate that *UGT1A* is an evolving locus armed to cope for new encountered exogenous components. In fact, some mutations are exclusively or mostly observed in particular human populations, supporting an environmental pressure in sequence selection. This is well exemplified by UGT1A1 mutations, UGT1A1*6, UGT1A1*7, and UGT1A1*27 solely detected in Asians (Aono et al., 1995; Aono et al., 1993). Moreover, examples of SNP association with glucuronidation rates and substrate selectivity support this evolutionary mechanism. A comparative study between *UGT1* loci from diverse vertebrates indicates that exon birth-and-death is likely involved in UGT evolution, as demonstrated by pseudogenes (Riedy et al., 2000; Turgeon, Carrier, Chouinard, & Belanger, 2003). The shared and also more conserved region of the protein (for UGT1A1 this corresponds to amino acids 289 to 533) includes the transmembrane domain, the binding domain for the co-substrate UDPGA (Meech & Mackenzie, 1997a). This conserved proteins domains are essential to preserve the fundamental glucuronic acid transferase function (Ouzzine, Barre, Netter, Magdalou, & Fournel-Gigleux, 2003; Patana, Kurkela, Goldman, & Finel, 2007; Radominska-Pandya, Czernik, Little, Battaglia, & Mackenzie, 1999).

The expression of each UGT1A mRNA isoform is determined by their own promoter region (Gong et al., 2001; Ritter et al., 1992). The use of multiple promoters by UGT1A gene provides a genomic framework enabling a cell to respond to diverse physiological and pathophysiological stimuli. Consequently, UGT1A enzymes also diverge in their expression profile in tissues and between individuals (Burchell & Coughtrie, 1997;

Guillemette, 2003; King, Rios, Green, & Tephly, 2000; Mackenzie, Gregory et al., 2005; Strassburg, Kalthoff, & Ehmer, 2008; Strassburg, Nguyen, Manns, & Tukey, 1998; Tukey & Strassburg, 2000b). For instance, while UGT1As are mostly produced in gastrointestinal tissues (GI) (eg. liver, stomach, small intestine, and colon), some are specifically encountered in extrahepatic tissues (UGT1A7, UGT1A8 and UGT1A10). As a major contributor in liver's detoxification process, UGT1A1 is widely expressed throughout human body, whereas UGT1A9 is largely expressed in liver and kidney (Ritter, 2007; Strassburg, Manns, & Tukey, 1998; Strassburg, Nguyen, Manns, & Tukey, 1998; Tukey & Strassburg, 2000a). UGT1A6, UGT1A3 and UGT1A4 are also produced in liver, and along with UGT1A10, in GI tract. Some UGTs are also present in brain, ovary, uterus, prostate and breast (Belanger, Pelletier, Labrie, Barbier, & Chouinard, 2003; Guillemette, Belanger, & Lepine, 2004; King, Rios, Green, & Tephly, 2000; Starlard-Davenport, Lyn-Cook, & Radominska-Pandya, 2008; Strassburg, Kalthoff, & Ehmer, 2008). Thus, it is clear that the alternative usage of duplicated first exons is the basis of UGT1A function and expression diversities. Besides, all this tissue-specific expression is well orchestrated by the cooperative interaction between transcription factors and the transcription initiation complex (for review see (Mackenzie et al., 2003; Owens, Basu, & Banerjee, 2005; Strassburg, Kalthoff, & Ehmer, 2008; Tukey & Strassburg, 2001) (Gardner-Stephen & Mackenzie, 2008) Studies concerning UGT1A promoter-interacting factors have revealed the binding of some important activators, such as HNF-1, cdx2, Nrf2, PXR, XRE (AhR) and CAR (Aueviriyavit, Furihata, Morimoto, Kobayashi, & Chiba, 2007; Bonzo, Belanger, & Tukey, 2007; Buckley & Klaassen, 2009; Gregory, Lewinsky, Gardner-Stephen, & Mackenzie, 2004; Senekeo-Effenberger

et al., 2007; Sugatani, Sueyoshi, Negishi, & Miwa, 2005; Verreault et al., 2006; Yuan, Li, & Yang, 2007). Transcriptional regulation of the *UGT1A* gene by transcription factors is more specifically addressed in some following chapters. Variability in these regulatory mechanisms are also accounting for interindividual differences in the glucuronidation pathway (Aueviriyavit, Furihata, Morimoto, Kobayashi, & Chiba, 2007; Mackenzie, Gregory et al., 2005; Ramirez et al., 2008; Toide et al., 2002). In addition, the recent discovery of a new alternatively used exon (referred as exon 5b) in the common region of the locus (exon 2-5) sheds light on innovative mechanisms likely implicated in the control UGT1A-mediated glucuronidation (see next section) (Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007).

3.2 The human UGT2 family members

In contrast to the *UGT1A* locus, the human *UGT2* family members are encoded for by unique and multiples genes on 4q13 (Mackenzie, Bock et al., 2005). As illustrated in **Figure 2**, *UGT2B* and *UGT2A* genes likely originated from segmental chromosome duplication. Indeed, the phylogenetic relationship well demonstrates the prominent similarity between these genes. *UGT2A3* and all *UGT2B* genes are composed of at least six coding exons (Riedy et al., 2000; Turgeon et al., 2000) (Court, Hazarika, Krishnaswamy, Finel, & Williams, 2008). UGT2A1 and UGT2A2 mRNAs are transcribed by the same locus, by the use of alternative first exons, as demonstrated for the *UGT1A* locus (Jedlitschky, Cassidy, Sales, Pratt, & Burchell, 1999; Mackenzie, Bock et al., 2005). The intergenic distances vary greatly and the gene transcription may occur on either strand. As pictured in **Figure 2**, and similar to the *UGT1A* family, the first exon

encodes for the substrate-binding domain, and the other coding exons are responsible for the UDPGA-binding and transmembrane domains. One of the most studied UGT2B enzyme is UGT2B7. Of great importance in the liver, it is abundantly expressed in GI tract, but also in breast, uterus, kidney, brain, and other tissues (Turgeon, Carrier, Levesque, Hum, & Belanger, 2001). UGT2B7 has the ability to glucuronidate several endogenous compounds, such as steroid hormones (Gall et al., 1999a), retinoids (Czernik, Little, Barone, Raufman, & Radominska-Pandya, 2000), fatty acids (Jude, Little, Bull, Podgorski, & Radominska-Pandya, 2001), but also exogenous molecules such as morphine (Coffman, Rios, King, & Tephly, 1997), AZT (Barbier et al., 2000), EFV (Mutlib et al., 1999), NSAIDS (C. Jin, J. O. Miners, K. J. Lillywhite, & P. I. Mackenzie, 1993; C. J. Jin, J. O. Miners, K. J. Lillywhite, & P. I. Mackenzie, 1993), and others (Court, 2005). Although the other UGT2B family members have been investigated to a smaller extent, they have important role, especially with endogenous compounds. For example, UGT2B4 is involved in the metabolism of catechol estrogens and bile acids (Fournel-Gigleux, Jackson, Wooster, & Burchell, 1989; Levesque, Beaulieu, Hum, & Belanger, 1999; Pillot et al., 1993; Ritter, Chen, Sheen, Lubet, & Owens, 1992). UGT2B15, UGT2B17 and UGT2B28 arise as the major UGT2B sex-steroid metabolizing transferases (Beaulieu, Levesque, Hum, & Belanger, 1996; Levesque et al., 1997; Levesque et al., 2001). In contrast to the UGT1A first exons, it appears that fewer mutations affecting the coding regions occurred in these genes, which indicates a more conservative structure for *UGT2B* genes [48-50]. Polymorphic deletions were shown to occur for UGT2B genes namely *UGT2B17* and *UGT2B28* (Jakobsson et al., 2006; McCarroll et al., 2006; Wilson et al., 2004). In fact, copy-number variations of these two

genes have been described with divergent frequencies among ethnic populations and surely contribute as a source of phenotypic differences (Spielman et al., 2007). These will be further discussed (see section 7).

4. Novel UGT-associated mRNA isoforms

With the recent genome-wide analyses, the alternative splicing becomes an important mechanism explaining the transcriptome (which is defined as all "transcripts" in a cell type at a given time) diversity, with the vast majority of human genes having alternatively spliced mRNA isoforms (Johnson et al., 2003; Kampa et al., 2004; Maniatis & Tasic, 2002; Stamm, 2008; E. T. Wang et al., 2008). There are several ways to generate alternative mRNA isoforms from one gene locus, namely the alternative promoter usage (observed in *UGT1A* gene), alternative exon usage, intron retention and the alternative 3'UTR (**Figure 2**) (Kornblihtt, 2005; Moore & Proudfoot, 2009). Beside to dramatically increases the transcriptome diversity, alternative splicing introduces novel mRNA molecules, which may lead to divergent polypeptides, in term of biological function and/or expression profile, but also having intrinsic regulatory role. Recent evidence support that this is the case for human *UGT* genes (Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007).

The human genome project has provided to the scientific community priceless data about genomic organization and sequences, but has also contributed to extend knowledge on gene products, especially RNA species. In fact, sequencing of large cDNA libraries (expressed-sequence tag, EST), the use of Tag-based sequencing methods such as SAGE (Anisimov, 2008)and CAGE (de Hoon & Hayashizaki, 2008), the development of

specialized expression arrays (exon and tiling arrays) (Samanta, Tongprasit, & Stolc, 2007) (Lin et al., 2009), and the use of "next-generation" sequencing system to perform RNA-seq (Z. Wang, Gerstein, & Snyder, 2009), are all important tools to better understand the transcriptome.

As abovementioned, the UGT1A locus is a perfect example of "one locus given several divergent mRNAs". The recent discovery of a new alternative exon in the common region of the *UGT1A* locus further supports such an increased mRNA diversity (Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007). Considering the high occurrence of gene-associated mRNA diversity, it is likely to investigate whether other *UGT* loci have additional mRNA species. In the following, we listed evidence for additional UGT mRNA isoforms by analyzing expressed-sequence tags (EST) gene clusters. We also mentioned some experimental supports for such mRNA species.

4.1 UGT-derived EST clusters

UniGene (http://www.ncbi.nlm.nih.gov/unigene) is an NCBI database attempting to clusterize transcribed sequences according to their similarity, thus providing a set of transcripts associated to one gene locus. Those transcripts came from several sources, namely mRNA sequences, but also from the EST database. Therefore, by searching *UniGene* database for *UGT* gene loci, we retrieved all associated transcripts, which have been then aligned between each others. By this simple exercise we sought to determine whether the *UGT* loci are associated with alternative mRNAs, i.e. transcripts having alternative exons1/promoters, internal exons or alternative ends.

The **Table 1** summarizes our findings, and presents new mRNAs for human *UGT* loci and supportive DNA clones. It is noteworthy that the transcript ID is an unofficial identifier only used to facilitate the description of each transcript. The latter includes sequences such as whole curated cDNA clones, but also single-passed sequences from either 5'-RACE, 3'-RACE, or shotgun-sequencing of cDNA clones. It is not surprising to observe alternative splicing events in all UGT loci, as presumed by the elevated ASE rate in human genes. The approach is solely qualitative and one may not deduce a relative abundance based upon the number of supportive sequences. We noted that ASEs mostly occur either at the beginning or the end of the mRNA molecules. Whether these mRNA species endogenously exist and generate proteins or, on the other hand, are modulator mRNA species (e.g. nonsense-mediated mRNA decay) have to be investigated. the deduced polypeptidic product for most ASEs would lead to shorter UGT proteins, each requiring molecular and biological characterization. One clear observation is the presence of diverse mRNA species likely having different functions and expression profile. Therefore, in future UGT expression experiment, researchers would face to clearly define the transcripts that would be addressed by their selected strategy to avoid further conflicting results. However, given the high sequence similarity between UGTs, this represents a significant challenge

Considering the already reported usage of alternative first exon in *UGT1A* locus, we only sought for ASEs in the common region. In fact, the use of an alternative exon 5 for UGT1A, previously described as exon 5b (Girard et al., 2007; Levesque, Girard,

Journault, Lepine, & Guillemette, 2007), is supported by five expressed sequences. It is noteworthy that one of these sequences is the reported mRNA from our previous published studies.(Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007) Moreover, this new exon has been further investigated in more detail, as presented in a next section. No other ASE has been found. As abovementioned, in contrast to *UGT1A* locus, the *UGT2* family members are encoded by single gene (Mackenzie, Bock et al., 2005; Riedy et al., 2000; Turgeon et al., 2000). We attempted to uncover ASEs associated to each gene. One major drawback of this approach is the high sequence similarity between either *UGT2A* or *UGT2B*. Indeed, the sequence quality of certain EST clones may impair the correct attribution to one gene cluster over another. In here, we aligned transcribed sequences associated with human *UniGene* clusters. This fast analysis allowed discovering several ASEs occurring in all *UGT2B* and *UGT2A* genes, excepted for *UGT2B11*, *UGT2B15*, and *UGT2B17*. We also have to note that splicing events occurring in the *UGT2B28* gene came from experimental evidence from the Dr. Alain Bélanger's laboratory (Levesque et al., 2001) and from several other monkey ESTs. The type of ASEs are various, we noted the use of either alternative 5' and 3' splice site (*UGT2B4*, *UGT2B10*, *UGT2B28*), the internal exon skipping (*UGT2B4*, *UGT2B7*, *UGT2B10*, *UGT2B28*), the alternative internal exon usage (*UGT2B4*). Although not alternative splicing events, we also noted the use of alternative promoter (alternative first exon in *UGT2B4* and *UGT2B7*), and alternative last exon (*UGT2B4* and *UGT2B7*). Some experimental evidences in support of these events have recently been obtained and are discussed below.

The major challenge is now to corroborate these new mRNA species in a cellular context, to define their abundance and expression profile, to determine their translational product (if it is so), and ultimately to establish their functional and physiological relevance. Firstly, assuming that these mRNAs may conduct to proteins, few of them should be functional transferase enzymes. These new translated polypeptides may serve for an unknown molecular process or act as a negative modulator of glucuronidation as discussed below. Otherwise, mRNA molecules bearing a prematured stop codon might intrinsically be involved in the nonsense-mediated mRNA decay (for review see (Chang, Imam, & Wilkinson, 2007). Several experimental studies are currently in progress and will be needed to clearly establish the role of all transcribed mRNA species, and whether these affect the net glucuronidation activity in a specific time and tissue.

5. Experimental evidence of new UGT1A-associated mRNAs

Extensive search for new putative exons at *UGT1* locus in several human tissues and cell lines was performed and led to discovery of a new exon, located in the common region of the gene, between coding exons 4 and 5 and was renamed exon 5b (**Figure 3**) (Levesque, Girard, Journault, Lepine, & Guillemette, 2007). This genomic sequence perfectly matches with the transcript UGT1A.2 reported in Table 1 found in EST database. However, according to our findings, this exon 5b might either be used as terminal exon (leading to the mRNA isoform variant 2 or v2) or be spliced with the common exon5 (exon5a) (leading to mRNA isoform variant 3 or v3)(**Figure 3**) (Girard et al., 2007). Thus, the discovery of the novel exon 5b initiated a search for a whole new class of UGT1A isoforms, and revealed the production of 18 new mRNAs (v2 and v3)

leading to 9 new UGT1A isoform 2 proteins (Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007). The presence of exon 5b (v2 and v3) in the mRNA sequence generates a shorter polypeptide product (named isoform i2 or UGT1A_i2) lacking the amino acids 435-533 encoded by exon 5a which are replaced by a sequence of 10 amino acids: RKKQQSGRQM (Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007) (**Figure 3**). These new UGT splice variant isoforms 2 lack the transmembrane domain encoded by the exon 5a but possess an ER-targeting signal peptide and complete binding sites for both the substrate and co-substrate. The heterologous expression of UGT1A_i2 isoforms in HEK293 cells indicate that despite abundant expression and ER localization of i2 proteins, they are inactive in the transfer of the glucuronic acid moiety to classical substrates of the UGT1A family members. The observations that the splice i2 forms are devoid of transferase activity (Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007) are in line with previous studies that reported the existence of a loss of enzyme activity for C-terminal truncated UGT proteins located in the ER (Levesque, Girard, Journault, Lepine, & Guillemette, 2007). However, they display a repressive effect on rates of formation of glucuronide products when co-expressed with their equivalent UGT1A_i1 isoforms in HEK293 cells (Girard et al., 2007; Levesque, Girard, Journault, Lepine, & Guillemette, 2007). Indeed, co-expression of i1 and i2 within the cells result in significant reduction of glucuronidation activity of UGT1A1, UGT1A7, UGT1A8 and UGT1A9 (Girard et al., 2007). These data support a potential dominant negative role for these splice forms of the *UGT1A* gene.

Using available polyclonal anti-UGT antibodies, the presence of 45-kd UGT1A_i2 proteins was shown to be widely observed in human tissues, including liver, kidney, oesophagus, small intestine and colon (Girard et al., 2007). In most tissues, UGT1A_i2 are expressed together with their UGT1A_i1 homologs. From the published data, the expression of i2 was predominant in extrahepatic tissues compared to the liver. The subcellular localization of i2 to the endoplasmic and perinuclear structure was firstly described in UGT1A-overexpressing HEK293 cells together with classical isoforms i1 where protein interactions might occur (Levesque, Girard, Journault, Lepine, & Guillemette, 2007). The endogenous expression of splice i2 proteins in human-derived colon cell lines indicates its intracellular location to ER membrane and perinuclear structure, as demonstrated for the UGT1A isoforms 1. Immunohistochemal experiments in human colon tissues further support localization of both isoforms i1 and i2 in similar structures (Bellemare *et al*. under review).

Down-regulation of UGT1A_i2 proteins by transfecting specific small interference RNA into in human colon cells enhances the glucuronidation activity for various UGT1A substrates (Bellemare *et al*. under review). This supports the dominant negative hypothesis for UGT1A_i2 proteins. These results not only demonstrate the biological significance of alternative splicing of the UGT1A system, but also suggest a new UGT regulatory pathway. These observations clearly unveil another level of complexicity in the regulation of intracellular glucuronidation activity and may have important physiological and pharmacological implications. We propose that protein interaction may underlie the negative role of shorter i2 proteins, and there is substantial literature

supporting that UGT proteins might interact together as dimers/oligomers or with other ER-resident proteins (Finel & Kurkela, 2008; Fremont, Wang, & King, 2005; Fujiwara et al., 2009; Fujiwara, Nakajima, Yamanaka, Katoh, & Yokoi, 2007; Fujiwara et al., 2007; Ghosh et al., 2001; Ikushiro, Emi, & Iyanagi, 1997; Ishii et al., 2001; Kurkela et al., 2003; Kurkela, Hirvonen, Kostiainen, & Finel, 2004; Kurkela et al., 2007; Meech & Mackenzie, 1997b; Nakajima, Yamanaka, Fujiwara, Katoh, & Yokoi, 2007; Operana & Tukey, 2007; Radominska-Pandya, Ouzzine, Fournel-Gigleux, & Magdalou, 2005; Takeda et al., 2009). Interaction between both splice forms is also supported by their proximity in ER membranes, and by the demonstration of physical interaction between both isoforms 1 and 2 of UGT1A1 through co-immunoprecipitation experiments (Levesque, Girard, Journault, Lepine, & Guillemette, 2007). Together, these observations begin to highlight a dynamic and functional antagonism between isoforms 1 and their splice i2 variants. A parallel example was reported by Ghosh *et al*. in 2001, where an inactive truncated mutant was acting as a negative modulator of UGT1A1 activity while interacting with wild-type UGT1A1 (Ghosh et al., 2001). Further investigations are definitely required to provide molecular basis for these effects. Yet, such a negative regulation by differential oligomerization of the alternatively spliced isoforms is widely reported for a variety of proteins (Hashimoto, Chiorazzi, & Gregersen, 1995; Lincz et al., 2008; McElvaine & Mayo, 2006; Nakamura, Hashimoto, Furuyama, & Kakudo, 1995; Nakamura et al., 1995; Sztainberg et al., 2009; W. Wang, Kwok, & Chan, 2007).

The evolutionary conservation of a particular gene structure as well as of a specific splice isoforms is generally a good indicator of the functionality of those variants. Indeed, the structure of the *UGT1* gene is highly conserved among orthologues vertebrate's

genes; exon 5b is also present in primates and rodents leading to shorter splice variants corresponding to human i2 orthologous (Girard et al., 2007). Furthermore, the search in public databases established the conservation of isoform 2 splice variants (UGT1A5_v2: Genbank accession number XM001150762) in chimpanzee (*Pan troglodytes*). Thus, the alternative splicing of the last exon at *UGT1A* locus might be part of the tightly regulated circuit that controls the overall cellular glucuronidation activity.

These new alternative splicing events occurring in the common region of the *UGT1A* gene increase the protein diversity derived from this single gene, while representing new post-transcriptional and post-translational regulatory mechanisms of the glucuronidation pathway. Besides, it is well appreciated that alterations in splicing patterns of genes contribute to the regulation of gene function by generating endogenous inhibitor or activator molecules. This raises the possibility that changes in the equilibrium between endogenous i1 an i2 forms of UGT1A possibly represent a significant contribution in global conjugation activity of the cell. At present, the physiological and pathological roles of this particular class of *UGT1A* gene products remain an unanswered and open question. Preliminary results tend to show a strong heterogeneity in expression and localization of i2 within tissues, but several studies are needed to understand the exact cellular functions of these newly discovered splice variants. Understanding the regulation of this complex transcriptional system will be of great interest in the near future and may certainly underlie part of the interindividual glucuronidation variability currently unexplained by *UGT1A* single nucleotide polymorphisms or other non-genetic causes.

6. Alternative splicing at the *UGT2* gene loci.

6.1 The UGT2B28 gene

Although EST clustering analyses suggest alternative splicing events in human *UGT2* genes, there is still little experimental evidence of such events. Effectively, the scarce evidence of splicing in this subfamily comes initially from the description of the *UGT2B28* gene in 2001 (Levesque et al., 2001). Indeed, two distinct alternatively spliced *UGT2B28* transcripts were isolated and characterized at the molecular level along with the active UGT2B28_v1 transcript, originally described as UGT2B28 type I (**Figure 4**) (Levesque et al., 2001). The UGT2B28 type III transcript lacks a portion of the exon 1, potentially via the utilization of a cryptic splicing site, whereas the UGT2B28 type II lacks the exons 4 and 5 in the mature mRNA (Levesque et al., 2001). These transcripts were the first evidence for the existence of alternative splicing in this human UGT subfamily, and are related to the transcripts UGT2B28.2 and UGT2B28.3 in Table 1. Functional data performed with the corresponding cDNA revealed no evidence of transferase activity for the UGT2B28-derived protein variants when stably expressed independently in HEK293 cells. However, no co-expression experiments were performed in this previous study and the physiological relevance of these transcripts remained to be assessed.

6.2 The UGT2B7 gene

More recently, Innocenti and coworkers demonstrated the presence of two novel exons, called exon 1A and 1B in the *UGT2B7* gene upstream of the first coding exon (**Figure 4**)

(Innocenti et al., 2008). The authors showed the existence of two additional UGT2B7-associated transcripts in human livers, of which one contain the exon 1A and 1B spliced with the coding exon 2 and 3 of UGT2B7 (UGT2B7_v2) and the other includes only the exon 1A joints to the coding exon 2 and 3 (UGT2B7_v3). They further reported that both isoforms ended in the intron 3. These isoforms are related to the predicted transcripts UGT2B7.3 and UGT2B7.4, respectively, described in Table 1, but, in contrast to Innocenti *et al*. (Innocenti et al., 2008), longer mRNA transcripts have been reported. Whether all these transcripts (shorter and longer) are produced by human cells remains to be uncovered. It is however clear that new UGT2B7 mRNAs are transcribed from an alternative promoter (upstream of the exon 1A). If these mRNAs were translated, it presumably start at the ATG in the coding exon 2, would lead to an UGT lacking the substrate-binding domain. Accordingly, whether the mRNA *per se,* the associated proteins, or both has functional relevance is still not known. Nevertheless, these observations will clearly influence the experimental design of subsequent studies about *UGT2B7* gene expression and activity knowing that several mRNAs are produced by this very important metabolizing gene in the human liver and extrahepatic tissues.

6.3 The UGT2B4 gene: A closer look at the genomic level

To explore in more detail the formation of alternative splice products at the *UGT2* loci, and the validity of using a computational approach to predict mRNA transcripts from one locus, the *UGT2B4* gene was chosen as model (**Figure 4**). Based upon data in Table 1, we designed amplification strategies to both determine the presence of UGT2B4-associated ASEs and full-length isoforms. All predicted alternative transcripts and ASEs

listed in Table 1 were observed in human liver tissues. In addition, we uncovered other mRNA isoforms revealing that alternative splicing at the *UGT2B4* locus produces significant gene expression diversity. Preliminary quantitative-PCR experiments show that splice variants are expressed at variable levels in the liver, whereas a more restricted pattern of expression seems to prevail in extrahepatic tissues (Lévesque, Ménard, Bellemare, Girard, & Guillemette). We further cloned and expressed individual splice UGT2B4 forms in HEK293 cells. Immunofluorescence analysis revealed that these transcripts conduct to proteins that are directed to the ER and perinuclear membranes despite lacking critical domains. Microsomal preparation from these stably transformed cells did not displayed detectable glucuronidation activity for eugenol, a broad substrate for UGT2Bs. However, in the presence of these new proteins, the rates of glucuronidation by the UGT2B4_i1 enzyme were significantly reduced by 30-50%. To further confirm the potential regulatory functions of these UGT2B4 splice proteins, additional experiments are required. However, as previously exemplified by the UGT1A_v2/v3 alternative transcript, the production of various mRNA species by the *UGT2B4* locus is not only related to erroneous transcription and/or splicing mechanisms, but is likely to have a functional role in the control of glucuronidation activity.

6.4 Conclusions on the topic of alternative splicing

Several questions remain to be answered before to substantiate functional role(s) for these new UGT transcripts *in vivo*. Nevertheless, we should now be aware of the

presence of such transcripts, particularly in the design of future gene expression analysis. Among these questions are: what is their relative abundance in various tissues and conditions? Do their changes in expression significantly influence the glucuronidation? by which mechanism(s) and under which circumstances? Are splice variants associated with diseases or specific physiological conditions? Is *in vivo* synthesis of splice forms controlled, if so, by which mechanism(s)? Does splice variants contribute to variation in drug metabolism at the magnitude of clinically significance? Does this affect the way patients metabolize drugs or endogenous molecules? Several examples in the literature support that splice variants of genes may affect drugs metabolism and even drug resistance (reviewed by (Passetti, Ferreira, & Costa, 2009)). We might not promptly answer these questions, and several studies will be needed to understand the function and importance of UGT splice variants in cellular metabolism.

7. Copy-number variations of *UGT2B* genes

The human genome displays extensive copy-number variations (CNV) of large DNA fragments (McCarroll et al., 2006). In particular circumstance, this genomic instability might either induce gene duplication or deletion, and thus influencing the expression levels of the targeted gene product (Spielman et al., 2007). When associated to a *UGT* gene, it may impact upon interindividual drug metabolism variability. Indeed, one of the most studied CNV in human includes the sequence of the *UGT2B17* sex-steroid and drug metabolizing gene. This gene was isolated from the human prostate and the human prostate cancer cell lines in 1996 by Dr Belanger's group in the search of hormone-related cancer genes (Beaulieu, Levesque, Hum, & Belanger, 1996). UGT2B17 was then

shown to be involved in the termination of the androgen signal in target cells, conjugating and inactivating the major androgen dihydrotestosterone (DHT), testosterone and androsterone (Beaulieu, Levesque, Hum, & Belanger, 1996). Then, the discovery of a common deletion of the entirety of gene leading to CNV has been reported and is characterized by either the presence of one gene copy or the absence of the entire gene sequence (Jakobsson et al., 2006; Wilson et al., 2004). Subsequently, *UGT2B17* has been showed to be one of the most commonly deleted genes in the human genome, this associated with the most significant variations in expression levels between ethnic groups (Jakobsson et al., 2006; McCarroll et al., 2006; Spielman et al., 2007). In addition, *UGT2B17* was selected as an initial candidate to study the evolutionary divergence of copy-number variations (CNV) in human populations (Xue et al., 2008). The clinical interests for the deletion of *UGT2B17* comes from its association with circulating testosterone levels (Jakobsson et al., 2006), its impact on doping in sports (Schulze et al., 2008), its association with prostate cancer risk revealed in some studies and its intrinsic role as an antigenic peptide in transplant procedures (Gallagher et al., 2008; Karypidis, Olsson, Andersson, Rane, & Ekstrom, 2008; Murata, Warren, & Riddell, 2003; Olsson et al., 2008; J. Park et al., 2006; J. Park et al., 2004; J. Y. Park et al., 2007; Xue et al., 2008). Moreover, CNV of *UGT2B17* has recently been associated with the pathogenesis of osteoporosis and higher concentrations of circulating testosterone and estradiol and, therefore, represents a susceptibility gene for this highly prevalent metabolic bone disease in Asians (Yang et al., 2008). The UGT2B17-including CNV might in fact be seen as a *human gene knockout* enabling a more precise characterization of the physiological role of this gene in complex diseases including hormone-dependent cancers. Definitely, the

major advances in the understanding of the role of *UGT2B17* gene comes from CNV studies, which are critical for appreciating the significance of this particular gene product within this high homology subfamily of UGT2B enzymes.

To further emphasize the importance of CNV in this human subfamily, another CNV on chromosome 4q13 encompassing the *UGT2B28* gene has been found in human populations. *UGT2B28* is located 500 kb away from *UGT2B17* and is also involved in androgen and estrogen metabolism (Levesque et al., 2001). *UGT2B28* is the more recent characterized gene in this subfamily. As opposed to the well characterized impact of *UGT2B17* deletion in human populations, the role of *UGT2B28* CNVs has never been addressed. Nevertheless, a recent association study using tag-SNPs (associated with the *UGT2B28* genomic deletion) suggested that foetal development and gestation length are influenced by the presence of *UGT2B28* CNVs; a feature which may be related to its detoxification function or its role in androgen and/or estrogen metabolism (C. L. Relton, 2008). In support, changes in expression levels of *UGT2B28* was also associated with progression of oesophageal dysplasia in a Chinese chemoprevention trial (Joshi et al., 2006), underscoring the needs to study *UGT2B28* CNVs as well.

To facilitate genomic studies in this high homology region on chromosome 4q13, the molecular identification of the breakpoints of these two common deletions was localized precisely (Menard, Eap, Harvey, Guillemette, & Lévesque, In press). Segmental duplications of 4.9 kb for *UGT2B17* and 6.8 kb for *UGT2B28* are localized 117 kb and 108 kb apart respectively on both ends of the deletions and both comprise purine-rich

recombination sites. CNVs of *UGT2B17* and *UGT2B28* occur at a frequency in Caucasians of 27% and 13.5%, respectively. We also demonstrate that only 43% of individuals have 2 copies of both genes, whereas 57% harbour at least one deletion. Their co-occurrence on 5% of the chromosomes creates a large genomic gap of 225 kb in these individuals. Their co-existence with a functional polymorphism of the androgen-metabolizing gene *UGT2B15*, p.D85Y (rs1902023:G>T) in Caucasians is further shown. Haplotype analysis revealed that the long arm of chromosome 4 can harbour seven different haplotypes, the 2B17+/2B28+/2B15Y^{85} (Hap I), 2B17+/2B28+/2B15D^{85} (Hap II), 2B17-/2B28+/2B15D^{85} (Hap III), 2B17+/2B28-/2B15Y^{85} (Hap IV), 2B17-/2B28+/2B15Y^{85} (Hap V), 2B17-/2B28-/2B15D/Y^{85} (Hap VI and VII) with frequencies of 34.5%, 29.5%, 15.5% and 8.5%, 7% and 5% respectively (Figure 5) (Menard, Eap, Harvey, Guillemette, & Lévesque, In press). The exact haplotypes frequencies of these three metabolizing genes in other ethnic group remains undetermined but is expected to be different because of the greater prevalence of UGT2B CNVs in other populations compared to Caucasians (McCarroll et al., 2006). The implications of these findings are critical because the polymorphic deletions of *UGT2B17* and *UGT2B28* are associated with one of the UGT2B15 codon 85 allele, which can have a profound impact on steroids metabolism in presence of CNVs in both these genes. Of note, in the last years, several studies were performed with respect to the *UGT2B17* deletion genotypes mainly in prostate cancer patients with conflicting results (Gallagher et al., 2008; Karypidis, Olsson, Andersson, Rane, & Ekstrom, 2008; J. Park et al., 2006; J. Park et al., 2004; J. Y. Park et al., 2007). CNVs of *UGT2B17* and *UGT2B28*, their distinct associations with a polymorphism of the *UGT2B15* gene and their variable prevalence in diverse ethnic

groups may indeed lead to divergent conclusions. Furthermore, several studies demonstrate the impact of CNV of *UGT2B17* on hormonal status in humans; however, the co-existence of *UGT2B17* and *UGT2B28* CNVs in 15% of the studied cohort is likely to influence the overall sex steroid metabolism in these subjects, the latter genes being involved in both androgens and estrogens elimination. Therefore, the genomic status of all these steroid metabolizing genes should be evaluated systematically in order to refine the molecular understanding of this complex locus and their physiological consequences on the endocrine system. Based on our data, it seems clear that haplotype analyses are critical because co-existence of these two common major deletion polymorphisms, in conjunction with their genetic associations with SNPs in the *UGT2B15* gene, may have even more profound physiological consequences on the overall hormonal metabolism then each of these functional polymorphisms taken individually.

8. Epigenetic regulation of *UGT* gene expression.

Epigenetic regulation such as DNA methylation is another key mechanism to either activate or silence gene transcription, and abnormal epigenetic regulation has been described as an important characteristic of tumor progression and malignancy (Grady & Carethers, 2008; Shelton, Misso, Shaw, Arthaningtyas, & Bhoola, 2008). Epigenetic mechanisms are also important to maintain tissue-specific gene expression. Furthermore, abnormal methylation of genes is by far more frequent and has often more drastic consequences than classical genetic mutation (Al-Romaih et al., 2007). Currently few evidence support the epigenetic modulation of human UGT expression and one of the few published example is the significant interindividual variability in the level of

methylation of the *UGT1A1* gene in cellular models derived from human colon tumors. UGT1A1 is well known to be responsible for the inactivation of SN-38 (Gagne et al., 2002), the active metabolite of the prodrug irinotecan use to treat metastatic colorectal cancer. Therefore, any epigenetic mechanisms altering the *UGT1A1* gene has the potential to modify the glucuronidation capacity at the target cell level and, hence, the biological effect of the drug. Aberrant methylation of specific CpG islands of the *UGT1A1* gene in UGT1A1-negative cells (HCT116, HCT-15, and COLO-320DM). Data support that *UGT1A1* promoter hypermethylation results in complete repression of gene transcriptional activity (Gagnon, Bernard, Villeneuve, Tetu, & Guillemette, 2006). Reversal of promoter methylation by treatment with demethylating and histone deacetylase inhibitor agents had the capacity to restore *UGT1A1* expression in hypermethylated UGT1A1-negative cells. Loss of *UGT1A1* methylation was further associated with an increase in UGT1A1 protein content and with an enhanced inactivation of SN-38 by 300% in HCT116 colon cancer cells (Gagnon, Bernard, Villeneuve, Tetu, & Guillemette, 2006). An analysis of the *UGT1A1* 5′-flanking sequence (−540 to −1) in human colon cancer cells has revealed that hypomethylation of this region is important for *UGT1A1* transcription (Gagnon, Bernard, Villeneuve, Tetu, & Guillemette, 2006). More specifically, the extent of *UGT1A1* promoter methylation between CpG -1 (-4nt relative to the ATG) and -4 (-99nt relative to the ATG) of the promoter was shown to significantly predict *UGT1A1* gene expression in colon cancer cell lines (Gagnon, Bernard, Villeneuve, Tetu, & Guillemette, 2006). Based on the above findings, it seems that epigenetic modifications are part of the complexity underlying *UGT* gene expression. At this point, it remains an open question whether the expression

of additional UGTs is regulated by epigenetic mechanism in a tissue-specific manner and if such regulation has phenotypic consequences.

9. Concluding Remarks

Over the last few years, the emphasis was directed towards the characterization of genetic variations in all previously characterized *UGT* genes in the hope of understanding the considerable genetic variability observed for this major molecular pathway. Despite finding some hints, the global genetic plasticity of the glucuronidation pathway remains unsolved. As discussed here, part of this diversity is expected to occur via other genomic processes, including the production of alternative transcripts, copy-number variations and epigenetic processes (**Figure 6**). Alternative splicing occurs in most genes and the UGT genes are definitely not an exception to this principle. At present, despite the fact that very few studies have been conducted, it is evident that extensive splicing patterns occur in the *UGT* genes and these processes are just beginning to emerge especially for *UGT2* family members. Predicted transcripts from public databases and experimental analyses undoubtedly sustain the presence of alternative transcripts in *UGT* genes, and at present, this should change the design of future experiment particularly regarding gene expression. Altogether, these cellular processes drastically increase the genetic variability arising from the *UGT1A* and *UGT2B* genes and protein diversity. By these observations, the analysis of high interindividual variability and diversity of the glucuronidation pathway is becoming even more complex and several years of investigation will certainly

be needed to understand the physiological and pharmacological roles of these new processes.

10. Acknowledgments

Some of the work presented herein was supported by the Canadian Institutes of Health Research (CIHR grants MOP-42392 and MOP-84223), The Natural Sciences and Engineering Research Council of Canada (NSERC) and The Canada Research Chair Program (C.G.). E.L. is a recipient of a CIHR clinician-scientist award. J.B and V.M. are recipients of a PhD studentship awards from the CIHR and Le Fonds de la recherche en santé du Québec (FRSQ), respectively. C.G. holds a Canada Research Chair in Pharmacogenomics.

Conflict of Interest statement: The authors have no conflict of interests to declare.

References

Al-Romaih, K., Somers, G. R., Bayani, J., Hughes, S., Prasad, M., Cutz, J. C., et al. (2007). Modulation by decitabine of gene expression and growth of osteosarcoma U2OS cells in vitro and in xenografts: Identification of apoptotic genes as targets for demethylation. *Cancer Cell Int, 7*, 14.

Anisimov, S. V. (2008). Serial Analysis of Gene Expression (SAGE): 13 years of application in research. *Curr Pharm Biotechnol., 9*(5), 338-350.

Aono, S., Adachi, Y., Uyama, E., Yamada, Y., Keino, H., Nanno, T., et al. (1995). Analysis of genes for bilirubin UDP-glucuronosyltransferase in Gilbert's syndrome. *Lancet, 345*(8955), 958-959.

Aono, S., Yamada, Y., Keino, H., Hanada, N., Nakagawa, T., Sasaoka, Y., et al. (1993). Identification of defect in the genes for bilirubin UDP-glucuronosyl-transferase in a patient with Crigler-Najjar syndrome type II. *Biochem Biophys Res Commun, 197*(3), 1239-1244.

Aueviriyavit, S., Furihata, T., Morimoto, K., Kobayashi, K., & Chiba, K. (2007). Hepatocyte nuclear factor 1 alpha and 4 alpha are factors involved in interindividual variability in the expression of UGT1A6 and UGT1A9 but not UGT1A1, UGT1A3 and UGT1A4 mRNA in human livers. *Drug Metab Pharmacokinet., 22*(5), 391-398.

Barbier, O., Turgeon, D., Girard, C., Green, M. D., Tephly, T. R., Hum, D. W., et al. (2000). 3'-azido-3'-deoxythimidine (AZT) is glucuronidated by human UDP-glucuronosyltransferase 2B7 (UGT2B7). *Drug Metab Dispos, 28*(5), 497-502.

Beaulieu, M., Levesque, E., Hum, D. W., & Belanger, A. (1996). Isolation and characterization of a novel cDNA encoding a human UDP-glucuronosyltransferase active on C19 steroids. *J Biol Chem, 271*(37), 22855-22862.

Belanger, A., Hum, D. W., Beaulieu, M., Levesque, E., Guillemette, C., Tchernof, A., et al. (1998). Characterization and regulation of UDP-glucuronosyltransferases in steroid target tissues. *J Steroid Biochem Mol Biol., 65*(1-6), 301-310.

Belanger, A., Pelletier, G., Labrie, F., Barbier, O., & Chouinard, S. (2003). Inactivation of androgens by UDP-glucuronosyltransferase enzymes in humans. *Trends Endocrinol Metab., 14*(10), 473-479.

Bellemare, J., Rouleau, M., Harvey, M., Tetu, B., Guillemette, C. Alternative spliced forms of the major phase II conjugating *UGT1A* gene regulate glucuronidation in human cells. *Under review*

Boeger, H., Bushnell, D. A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J. S., et al. (2005). Structural basis of eukaryotic gene transcription. *FEBS Lett, 579*(4), 899-903.

Bonzo, J. A., Belanger, A., & Tukey, R. H. (2007). The role of chrysin and the ah receptor in induction of the human UGT1A1 gene in vitro and in transgenic UGT1 mice. *Hepatology., 45*(2), 349-360.

Buckley, D. B., & Klaassen, C. D. (2009). Induction of mouse UDP-glucuronosyltransferase mRNA expression in liver and intestine by activators of aryl-hydrocarbon receptor, constitutive androstane receptor, pregnane X receptor,

peroxisome proliferator-activated receptor alpha, and nuclear factor erythroid 2-related factor 2. *Drug Metab Dispos., 37*(4), 847-856. Epub 2009 Jan 2014.

Burchell, B., & Coughtrie, M. W. (1997). Genetic and environmental factors associated with variation of human xenobiotic glucuronidation and sulfation. *Environ Health Perspect., 105*(Suppl 4), 739-747.

Burchell, B., Soars, M., Monaghan, G., Cassidy, A., Smith, D., & Ethell, B. (2000). Drug-mediated toxicity caused by genetic deficiency of UDP-glucuronosyltransferases. *Toxicol Lett., 112-113*, 333-340.

C. L. Relton, R. M. F., S. M. Ring, A. Groom, R. Gray, A. T. Hattersley, T. M. Frayling, G. Davey Smith (2008). The influence of metabolic polymorphisms and smoking during pregnancy on fetal growth and preterm delivery. .

Chang, Y. F., Imam, J. S., & Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem., 76*, 51-74.

Coffman, B. L., Rios, G. R., King, C. D., & Tephly, T. R. (1997). Human UGT2B7 catalyzes morphine glucuronidation. *Drug Metab Dispos, 25*(1), 1-4.

Court, M. H. (2005). Isoform-selective probe substrates for in vitro studies of human UDP-glucuronosyltransferases. *Methods Enzymol., 400*, 104-116.

Court, M. H., Hazarika, S., Krishnaswamy, S., Finel, M., & Williams, J. A. (2008). Novel polymorphic human UDP-glucuronosyltransferase 2A3: cloning, functional characterization of enzyme variants, comparative tissue expression, and gene induction. *Mol Pharmacol., 74*(3), 744-754. Epub 2008 Jun 2003.

Czernik, P. J., Little, J. M., Barone, G. W., Raufman, J. P., & Radominska-Pandya, A. (2000). Glucuronidation of estrogens and retinoic acid and expression of UDP-glucuronosyltransferase 2B7 in human intestinal mucosa. *Drug Metab Dispos, 28*(10), 1210-1216.

de Hoon, M., & Hayashizaki, Y. (2008). Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. *Biotechniques., 44*(5), 627-628, 630, 632.

Dutton, G. J. (1980). *Glucuronidation of drugs and other compounds*. Boca Raton, Florida: CRC Press.

Finel, M., & Kurkela, M. (2008). The UDP-glucuronosyltransferases as oligomeric enzymes. *Curr Drug Metab., 9*(1), 70-76.

Fournel-Gigleux, S., Jackson, M. R., Wooster, R., & Burchell, B. (1989). Expression of a human liver cDNA encoding a UDP-glucuronosyltransferase catalysing the glucuronidation of hyodeoxycholic acid in cell culture. *FEBS Lett, 243*(2), 119-122.

Fremont, J. J., Wang, R. W., & King, C. D. (2005). Coimmunoprecipitation of UDP-glucuronosyltransferase isoforms and cytochrome P450 3A4. *Mol Pharmacol., 67*(1), 260-262. Epub 2004 Oct 2014.

Fujiwara, R., Nakajima, M., Oda, S., Yamanaka, H., Ikushiro, S. I., Sakaki, T., et al. (2009). Interactions between human UDP-glucuronosyltransferase (UGT) 2B7 and UGT1A enzymes. *J Pharm Sci, 27*, 27.

Fujiwara, R., Nakajima, M., Yamanaka, H., Katoh, M., & Yokoi, T. (2007). Interactions between human UGT1A1, UGT1A4, and UGT1A6 affect their enzymatic activities. *Drug Metab Dispos., 35*(10), 1781-1787. Epub 2007 Jul 1789.

Fujiwara, R., Nakajima, M., Yamanaka, H., Nakamura, A., Katoh, M., Ikushiro, S., et al. (2007). Effects of coexpression of UGT1A9 on enzymatic activities of human UGT1A isoforms. *Drug Metab Dispos., 35*(5), 747-757. Epub 2007 Feb 2009.

Gagne, J. F., Montminy, V., Belanger, P., Journault, K., Gaucher, G., & Guillemette, C. (2002). Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). *Mol Pharmacol, 62*(3), 608-617.

Gagnon, J. F., Bernard, O., Villeneuve, L., Tetu, B., & Guillemette, C. (2006). Irinotecan inactivation is modulated by epigenetic silencing of UGT1A1 in colon cancer. *Clin Cancer Res., 12*(6), 1850-1858.

Gall, W. E., Zawada, G., Mojarrabi, B., Tephly, T. R., Green, M. D., Coffman, B. L., et al. (1999a). Differential glucuronidation of bile acids, androgens and estrogens by human UGT1A3 and 2B7. *J Steroid Biochem Mol Biol, 70*(1-3), 101-108.

Gall, W. E., Zawada, G., Mojarrabi, B., Tephly, T. R., Green, M. D., Coffman, B. L., et al. (1999b). Differential glucuronidation of bile acids, androgens and estrogens by human UGT1A3 and 2B7. *J Steroid Biochem Mol Biol., 70*(1-3), 101-108.

Gallagher, C. J., Kadlubar, F. F., Muscat, J. E., Ambrosone, C. B., Lang, N. P., & Lazarus, P. (2008). The UGT2B17 gene deletion polymorphism and risk of prostate cancer: A case-control study in Caucasians. *Cancer Detect Prev, 32*(3), 186-187.

Gardner-Stephen, D. A., & Mackenzie, P. I. (2008). Liver-enriched transcription factors and their role in regulating UDP glucuronosyltransferase gene expression. *Curr Drug Metab., 9*(5), 439-452.

Ghosh, S. S., Sappal, B. S., Kalpana, G. V., Lee, S. W., Chowdhury, J. R., & Chowdhury, N. R. (2001). Homodimerization of human bilirubin-uridine-diphosphoglucuronate glucuronosyltransferase-1 (UGT1A1) and its functional implications. *J Biol Chem, 276*(45), 42108-42115.

Gilbert, W. (1978). Why genes in pieces? *Nature., 271*(5645), 501.

Girard, H., Levesque, E., Bellemare, J., Journault, K., Caillier, B., & Guillemette, C. (2007). 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, 17*(12), 1077-1089.

Gong, Q. H., Cho, J. W., Huang, T., Potter, C., Gholami, N., Basu, N. K., et al. (2001). Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus. *Pharmacogenetics, 11*(4), 357-368.

Grady, W. M., & Carethers, J. M. (2008). Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology., 135*(4), 1079-1099. Epub 2008 Sep 1074.

Gregory, P. A., Lewinsky, R. H., Gardner-Stephen, D. A., & Mackenzie, P. I. (2004). Regulation of UDP glucuronosyltransferases in the gastrointestinal tract. *Toxicol Appl Pharmacol, 199*(3), 354-363.

Guillemette, C. (2003). Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J, 3*(3), 136-158.

Guillemette, C., Belanger, A., & Lepine, J. (2004). Metabolic inactivation of estrogens in breast tissue by UDP-glucuronosyltransferase enzymes: an overview. *Breast Cancer Res., 6*(6), 246-254. Epub 2004 Sep 2027.

Hashimoto, S., Chiorazzi, N., & Gregersen, P. K. (1995). Alternative splicing of CD79a (Ig-alpha/mb-1) and CD79b (Ig-beta/B29) RNA transcripts in human B cells. *Mol Immunol, 32*(9), 651-659.

Ikushiro, S., Emi, Y., & Iyanagi, T. (1997). Protein-protein interactions between UDP-glucuronosyltransferase isozymes in rat hepatic microsomes. *Biochemistry., 36*(23), 7154-7161.

Innocenti, F., Liu, W., Fackenthal, D., Ramirez, J., Chen, P., Ye, X., et al. (2008). Single nucleotide polymorphism discovery and functional assessment of variation in the UDP-glucuronosyltransferase 2B7 gene. *Pharmacogenet Genomics, 18*(8), 683-697.

Ishii, Y., Miyoshi, A., Watanabe, R., Tsuruda, K., Tsuda, M., Yamaguchi-Nagamatsu, Y., et al. (2001). Simultaneous expression of guinea pig UDP-glucuronosyltransferase 2B21 and 2B22 in COS-7 cells enhances UDP-glucuronosyltransferase 2B21-catalyzed morphine-6-glucuronide formation. *Mol Pharmacol., 60*(5), 1040-1048.

Jakobsson, J., Ekstrom, L., Inotsume, N., Garle, M., Lorentzon, M., Ohlsson, C., et al. (2006). Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J Clin Endocrinol Metab., 91*(2), 687-693. Epub 2005 Dec 2006.

Jedlitschky, G., Cassidy, A. J., Sales, M., Pratt, N., & Burchell, B. (1999). Cloning and characterization of a novel human olfactory UDP-glucuronosyltransferase. *Biochem J., 340*(Pt 3), 837-843.

Jin, C., Miners, J. O., Lillywhite, K. J., & Mackenzie, P. I. (1993). Complementary deoxyribonucleic acid cloning and expression of a human liver uridine diphosphate-glucuronosyltransferase glucuronidating carboxylic acid-containing drugs. *J Pharmacol Exp Ther., 264*(1), 475-479.

Jin, C. J., Miners, J. O., Lillywhite, K. J., & Mackenzie, P. I. (1993). cDNA cloning and expression of two new members of the human liver UDP-glucuronosyltransferase 2B subfamily. *Biochem Biophys Res Commun, 194*(1), 496-503.

Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C. D., et al. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science, 302*(5653), 2141-2144.

Joshi, N., Johnson, L. L., Wei, W. Q., Abnet, C. C., Dong, Z. W., Taylor, P. R., et al. (2006). Gene expression differences in normal esophageal mucosa associated with regression and progression of mild and moderate squamous dysplasia in a high-risk Chinese population. *Cancer Res, 66*(13), 6851-6860.

Jude, A. R., Little, J. M., Bull, A. W., Podgorski, I., & Radominska-Pandya, A. (2001). 13-hydroxy- and 13-oxooctadecadienoic acids: novel substrates for human UDP-glucuronosyltransferases. *Drug Metab Dispos, 29*(5), 652-655.

Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., et al. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res, 14*(3), 331-342.

Karypidis, A. H., Olsson, M., Andersson, S. O., Rane, A., & Ekstrom, L. (2008). Deletion polymorphism of the UGT2B17 gene is associated with increased risk

for prostate cancer and correlated to gene expression in the prostate. *Pharmacogenomics J, 8*(2), 147-151.

King, C. D., Rios, G. R., Green, M. D., & Tephly, T. R. (2000). UDP-glucuronosyltransferases. *Curr Drug Metab., 1*(2), 143-161.

Kornblihtt, A. R. (2005). Promoter usage and alternative splicing. *Curr Opin Cell Biol., 17*(3), 262-268.

Kurkela, M., Garcia-Horsman, J. A., Luukkanen, L., Morsky, S., Taskinen, J., Baumann, M., et al. (2003). Expression and characterization of recombinant human UDP-glucuronosyltransferases (UGTs). UGT1A9 is more resistant to detergent inhibition than other UGTs and was purified as an active dimeric enzyme. *J Biol Chem., 278*(6), 3536-3544. Epub 2002 Nov 3514.

Kurkela, M., Hirvonen, J., Kostiainen, R., & Finel, M. (2004). The interactions between the N-terminal and C-terminal domains of the human UDP-glucuronosyltransferases are partly isoform-specific, and may involve both monomers. *Biochem Pharmacol., 68*(12), 2443-2450.

Kurkela, M., Patana, A. S., Mackenzie, P. I., Court, M. H., Tate, C. G., Hirvonen, J., et al. (2007). Interactions with other human UDP-glucuronosyltransferases attenuate the consequences of the Y485D mutation on the activity and substrate affinity of UGT1A6. *Pharmacogenet Genomics., 17*(2), 115-126.

Lepine, J., Bernard, O., Plante, M., Tetu, B., Pelletier, G., Labrie, F., et al. (2004). Specificity and regioselectivity of the conjugation of estradiol, estrone, and their catecholestrogen and methoxyestrogen metabolites by human uridine diphospho-glucuronosyltransferases expressed in endometrium. *J Clin Endocrinol Metab, 89*(10), 5222-5232.

Levesque, E., Beaulieu, M., Green, M. D., Tephly, T. R., Belanger, A., & Hum, D. W. (1997). Isolation and characterization of UGT2B15(Y85): a UDP-glucuronosyltransferase encoded by a polymorphic gene. *Pharmacogenetics, 7*(4), 317-325.

Levesque, E., Beaulieu, M., Hum, D. W., & Belanger, A. (1999). Characterization and substrate specificity of UGT2B4 (E458): a UDP-glucuronosyltransferase encoded by a polymorphic gene. *Pharmacogenetics, 9*(2), 207-216.

Levesque, E., Girard, H., Journault, K., Lepine, J., & Guillemette, C. (2007). Regulation of the UGT1A1 bilirubin-conjugating pathway: role of a new splicing event at the UGT1A locus. *Hepatology, 45*(1), 128-138.

Lévesque, E., Ménard, V., Bellemare, J., Girard, H., & Guillemette, C. Extensive splicing processes in the bile acid-conjugating gene UGT2B4 enrich diversity and modulate glucuronidation. *Under review*.

Levesque, E., Turgeon, D., Carrier, J. S., Montminy, V., Beaulieu, M., & Belanger, A. (2001). Isolation and characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. *Biochemistry, 40*(13), 3869-3881.

Lin, L., Liu, S., Brockway, H., Seok, J., Jiang, P., Wong, W. H., et al. (2009). Using high-density exon arrays to profile gene expression in closely related species. *Nucleic Acids Res, 27*, 27.

Lincz, L. F., Mudge, L. M., Scorgie, F. E., Sakoff, J. A., Hamilton, C. S., & Seldon, M. (2008). Quantification of hTERT splice variants in melanoma by SYBR green

real-time polymerase chain reaction indicates a negative regulatory role for the beta deletion variant. *Neoplasia, 10*(10), 1131-1137.

Mackenzie, P. I., Bock, K. W., Burchell, B., Guillemette, C., Ikushiro, S., Iyanagi, T., et al. (2005). Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics, 15*(10), 677-685.

Mackenzie, P. I., Gregory, P. A., Gardner-Stephen, D. A., Lewinsky, R. H., Jorgensen, B. R., Nishiyama, T., et al. (2003). Regulation of UDP glucuronosyltransferase genes. *Curr Drug Metab., 4*(3), 249-257.

Mackenzie, P. I., Gregory, P. A., Lewinsky, R. H., Yasmin, S. N., Height, T., McKinnon, R. A., et al. (2005). Polymorphic variations in the expression of the chemical detoxifying UDP glucuronosyltransferases. *Toxicol Appl Pharmacol., 207*(2 Suppl), 77-83.

Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., et al. (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics, 7*(4), 255-269.

Maniatis, T., & Tasic, B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature, 418*(6894), 236-243.

McCarroll, S. A., Hadnott, T. N., Perry, G. H., Sabeti, P. C., Zody, M. C., Barrett, J. C., et al. (2006). Common deletion polymorphisms in the human genome. *Nat Genet, 38*(1), 86-92.

McElvaine, A. T., & Mayo, K. E. (2006). A dominant-negative human growth hormone-releasing hormone (GHRH) receptor splice variant inhibits GHRH binding. *Endocrinology, 147*(4), 1884-1894.

Meech, R., & Mackenzie, P. I. (1997a). Structure and function of uridine diphosphate glucuronosyltransferases. *Clin Exp Pharmacol Physiol, 24*(12), 907-915.

Meech, R., & Mackenzie, P. I. (1997b). UDP-glucuronosyltransferase, the role of the amino terminus in dimerization. *J Biol Chem, 272*(43), 26913-26917.

Menard, V., Eap, O., Harvey, M., Guillemette, C., & Lévesque, E. (In press). Copy-number variations of the human sex steroid metabolizing genes UGT2B17 and UGT2B28 along with their associations with a UGT2B15 functional polymorphism. *Human Mutation*.

Moore, M. J., & Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell., 136*(4), 688-700.

Murata, M., Warren, E. H., & Riddell, S. R. (2003). A human minor histocompatibility antigen resulting from differential expression due to a gene deletion. *J Exp Med, 197*(10), 1279-1289.

Mutlib, A. E., Chen, H., Nemeth, G. A., Markwalder, J. A., Seitz, S. P., Gan, L. S., et al. (1999). Identification and characterization of efavirenz metabolites by liquid chromatography/mass spectrometry and high field NMR: species differences in the metabolism of efavirenz. *Drug Metab Dispos, 27*(11), 1319-1333.

Nakajima, M., Yamanaka, H., Fujiwara, R., Katoh, M., & Yokoi, T. (2007). Stereoselective glucuronidation of 5-(4'-hydroxyphenyl)-5-phenylhydantoin by human UDP-glucuronosyltransferase (UGT) 1A1, UGT1A9, and UGT2B15: effects of UGT-UGT interactions. *Drug Metab Dispos., 35*(9), 1679-1686. Epub 2007 Jun 1618.

Nakamura, M., Hashimoto, T., Furuyama, J., & Kakudo, K. (1995). Multiple alternative splice isoforms of parathyroid hormone-related peptide mRNA in human cell lines. *J Mol Endocrinol, 15*(3), 245-249.

Nakamura, M., Hashimoto, T., Nakajima, T., Ichii, S., Furuyama, J., Ishihara, Y., et al. (1995). A new type of human calcitonin receptor isoform generated by alternative splicing. *Biochem Biophys Res Commun, 209*(2), 744-751.

Olsson, M., Lindstrom, S., Haggkvist, B., Adami, H. O., Balter, K., Stattin, P., et al. (2008). The UGT2B17 gene deletion is not associated with prostate cancer risk. *Prostate, 4*, 4.

Operana, T. N., & Tukey, R. H. (2007). Oligomerization of the UDP-glucuronosyltransferase 1A proteins: homo- and heterodimerization analysis by fluorescence resonance energy transfer and co-immunoprecipitation. *J Biol Chem, 282*(7), 4821-4829.

Ouzzine, M., Barre, L., Netter, P., Magdalou, J., & Fournel-Gigleux, S. (2003). The human UDP-glucuronosyltransferases: structural aspects and drug glucuronidation. *Drug Metab Rev., 35*(4), 287-303.

Owens, I. S., Basu, N. K., & Banerjee, R. (2005). UDP-glucuronosyltransferases: gene structures of UGT1 and UGT2 families. *Methods Enzymol, 400*, 1-22.

Park, J., Chen, L., Ratnashinge, L., Sellers, T. A., Tanner, J. P., Lee, J. H., et al. (2006). Deletion polymorphism of UDP-glucuronosyltransferase 2B17 and risk of prostate cancer in African American and Caucasian men. *Cancer Epidemiol Biomarkers Prev., 15*(8), 1473-1478.

Park, J., Chen, L., Shade, K., Lazarus, P., Seigne, J., Patterson, S., et al. (2004). Asp85tyr polymorphism in the udp-glucuronosyltransferase (UGT) 2B15 gene and the risk of prostate cancer. *J Urol., 171*(6 Pt 1), 2484-2488.

Park, J. Y., Tanner, J. P., Sellers, T. A., Huang, Y., Stevens, C. K., Dossett, N., et al. (2007). Association between polymorphisms in HSD3B1 and UGT2B17 and prostate cancer risk. *Urology., 70*(2), 374-379.

Passetti, F., Ferreira, C. G., & Costa, F. F. (2009). The impact of microRNAs and alternative splicing in pharmacogenomics. *Pharmacogenomics J., 9*(1), 1-13.

Patana, A. S., Kurkela, M., Goldman, A., & Finel, M. (2007). The human UDP-glucuronosyltransferase: identification of key residues within the nucleotide-sugar binding site. *Mol Pharmacol., 72*(3), 604-611. Epub 2007 Jun 2019.

Pillot, T., Ouzzine, M., Fournel-Gigleux, S., Lafaurie, C., Radominska, A., Burchell, B., et al. (1993). Glucuronidation of hyodeoxycholic acid in human liver. Evidence for a selective role of UDP-glucuronosyltransferase 2B4. *J Biol Chem, 268*(34), 25636-25642.

Radominska-Pandya, A., Czernik, P. J., Little, J. M., Battaglia, E., & Mackenzie, P. I. (1999). Structural and functional studies of UDP-glucuronosyltransferases. *Drug Metab Rev., 31*(4), 817-899.

Radominska-Pandya, A., Ouzzine, M., Fournel-Gigleux, S., & Magdalou, J. (2005). Structure of UDP-glucuronosyltransferases in membranes. *Methods Enzymol., 400*, 116-147.

Ramirez, J., Mirkov, S., Zhang, W., Chen, P., Das, S., Liu, W., et al. (2008). Hepatocyte nuclear factor-1 alpha is associated with UGT1A1, UGT1A9 and UGT2B7

mRNA expression in human liver. *Pharmacogenomics J., 8*(2), 152-161. Epub 2007 Apr 2017.

Riedy, M., Wang, J. Y., Miller, A. P., Buckler, A., Hall, J., & Guida, M. (2000). Genomic organization of the UGT2b gene cluster on human chromosome 4q13. *Pharmacogenetics, 10*(3), 251-260.

Ritter, J. K. (2000). Roles of glucuronidation and UDP-glucuronosyltransferases in xenobiotic bioactivation reactions. *Chem Biol Interact., 129*(1-2), 171-193.

Ritter, J. K. (2007). Intestinal UGTs as potential modifiers of pharmacokinetics and biological responses to drugs and xenobiotics. *Expert Opin Drug Metab Toxicol., 3*(1), 93-107.

Ritter, J. K., Chen, F., Sheen, Y. Y., Lubet, R. A., & Owens, I. S. (1992). Two human liver cDNAs encode UDP-glucuronosyltransferases with 2 log differences in activity toward parallel substrates including hyodeoxycholic acid and certain estrogen derivatives. *Biochemistry, 31*(13), 3409-3414.

Ritter, J. K., Chen, F., Sheen, Y. Y., Tran, H. M., Kimura, S., Yeatman, M. T., et al. (1992). A novel complex locus UGT1 encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini. *J Biol Chem, 267*(5), 3257-3261.

Ritter, J. K., Crawford, J. M., & Owens, I. S. (1991). Cloning of two human liver bilirubin UDP-glucuronosyltransferase cDNAs with expression in COS-1 cells. *J Biol Chem, 266*(2), 1043-1047.

Samanta, M. P., Tongprasit, W., & Stolc, V. (2007). In-depth query of large genomes using tiling arrays. *Methods Mol Biol., 377*, 163-174.

Schulze, J. J., Lundmark, J., Garle, M., Skilving, I., Ekstrom, L., & Rane, A. (2008). Doping test results dependent on genotype of uridine diphospho-glucuronosyl transferase 2B17, the major enzyme for testosterone glucuronidation. *J Clin Endocrinol Metab, 93*(7), 2500-2506.

Senekeo-Effenberger, K., Chen, S., Brace-Sinnokrak, E., Bonzo, J. A., Yueh, M. F., Argikar, U., et al. (2007). Expression of the human UGT1 locus in transgenic mice by 4-chloro-6-(2,3-xylidino)-2-pyrimidinylthioacetic acid (WY-14643) and implications on drug metabolism through peroxisome proliferator-activated receptor alpha activation. *Drug Metab Dispos., 35*(3), 419-427. Epub 2006 Dec 2006.

Shelton, B. P., Misso, N. L., Shaw, O. M., Arthaningtyas, E., & Bhoola, K. D. (2008). Epigenetic regulation of human epithelial cell cancers. *Curr Opin Mol Ther., 10*(6), 568-578.

Spielman, R. S., Bastone, L. A., Burdick, J. T., Morley, M., Ewens, W. J., & Cheung, V. G. (2007). Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet, 39*(2), 226-231.

Stamm, S. (2008). Regulation of alternative splicing by reversible protein phosphorylation. *J Biol Chem, 283*(3), 1223-1227.

Starlard-Davenport, A., Lyn-Cook, B., & Radominska-Pandya, A. (2008). Identification of UDP-glucuronosyltransferase 1A10 in non-malignant and malignant human breast tissues. *Steroids, 73*(6), 611-620.

Strassburg, C. P., Kalthoff, S., & Ehmer, U. (2008). Variability and function of family 1 uridine-5'-diphosphate glucuronosyltransferases (UGT1A). *Crit Rev Clin Lab Sci, 45*(6), 485-530.

Strassburg, C. P., Manns, M. P., & Tukey, R. H. (1998). Expression of the UDP-glucuronosyltransferase 1A locus in human colon. Identification and characterization of the novel extrahepatic UGT1A8. *J Biol Chem, 273*(15), 8719-8726.

Strassburg, C. P., Nguyen, N., Manns, M. P., & Tukey, R. H. (1998). Polymorphic expression of the UDP-glucuronosyltransferase UGT1A gene locus in human gastric epithelium. *Mol Pharmacol, 54*(4), 647-654.

Sugatani, J., Sueyoshi, T., Negishi, M., & Miwa, M. (2005). Regulation of the human UGT1A1 gene by nuclear receptors constitutive active/androstane receptor, pregnane X receptor, and glucocorticoid receptor. *Methods Enzymol., 400*, 92-104.

Sztainberg, Y., Kuperman, Y., Issler, O., Gil, S., Vaughan, J., Rivier, J., et al. (2009). A novel corticotropin-releasing factor receptor splice variant exhibits dominant negative activity: a putative link to stress-induced heart disease. *FASEB J*.

Takeda, S., Ishii, Y., Iwanaga, M., Nurrochmad, A., Ito, Y., Mackenzie, P. I., et al. (2009). Interaction of cytochrome P450 3A4 and UDP-glucuronosyltransferase 2B7: evidence for protein-protein association and possible involvement of CYP3A4 J-helix in the interaction. *Mol Pharmacol., 75*(4), 956-964. Epub 2009 Jan 2021.

Toide, K., Takahashi, Y., Yamazaki, H., Terauchi, Y., Fujii, T., Parkinson, A., et al. (2002). Hepatocyte nuclear factor-1alpha is a causal factor responsible for interindividual differences in the expression of UDP-glucuronosyltransferase 2B7 mRNA in human livers. *Drug Metab Dispos., 30*(6), 613-615.

Trottier, J., Verreault, M., Grepper, S., Monte, D., Belanger, J., Kaeding, J., et al. (2006). Human UDP-glucuronosyltransferase (UGT)1A3 enzyme conjugates chenodeoxycholic acid in the liver. *Hepatology, 44*(5), 1158-1170.

Tukey, R. H., & Strassburg, C. P. (2000a). Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu Rev Pharmacol Toxicol., 40*, 581-616.

Tukey, R. H., & Strassburg, C. P. (2000b). Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu Rev Pharmacol Toxicol, 40*, 581-616.

Tukey, R. H., & Strassburg, C. P. (2001). Genetic multiplicity of the human UDP-glucuronosyltransferases and regulation in the gastrointestinal tract. *Mol Pharmacol, 59*(3), 405-414.

Turgeon, D., Carrier, J. S., Chouinard, S., & Belanger, A. (2003). Glucuronidation activity of the UGT2B17 enzyme toward xenobiotics. *Drug Metab Dispos, 31*(5), 670-676.

Turgeon, D., Carrier, J. S., Levesque, E., Beatty, B. G., Belanger, A., & Hum, D. W. (2000). Isolation and characterization of the human UGT2B15 gene, localized within a cluster of UGT2B genes and pseudogenes on chromosome 4. *J Mol Biol, 295*(3), 489-504.

Turgeon, D., Carrier, J. S., Levesque, E., Hum, D. W., & Belanger, A. (2001). Relative enzymatic activity, protein stability, and tissue distribution of human steroid-metabolizing UGT2B subfamily members. *Endocrinology, 142*(2), 778-787.

Verreault, M., Senekeo-Effenberger, K., Trottier, J., Bonzo, J. A., Belanger, J., Kaeding, J., et al. (2006). The liver X-receptor alpha controls hepatic expression of the human bile acid-glucuronidating UGT1A3 enzyme in human cells and transgenic mice. *Hepatology, 44*(2), 368-378.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature., 456*(7221), 470-476.

Wang, W., Kwok, A. M., & Chan, J. Y. (2007). The p65 isoform of Nrf1 is a dominant negative inhibitor of ARE-mediated transcription. *J Biol Chem, 282*(34), 24670-24678.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet., 10*(1), 57-63.

Wells, P. G., Mackenzie, P. I., Chowdhury, J. R., Guillemette, C., Gregory, P. A., Ishii, Y., et al. (2004). Glucuronidation and the UDP-glucuronosyltransferases in health and disease. *Drug Metab Dispos., 32*(3), 281-290.

Wilson, W., 3rd, Pardo-Manuel de Villena, F., Lyn-Cook, B. D., Chatterjee, P. K., Bell, T. A., Detwiler, D. A., et al. (2004). Characterization of a common deletion polymorphism of the UGT2B17 gene linked to UGT2B15. *Genomics, 84*(4), 707-714.

Xue, Y., Sun, D., Daly, A., Yang, F., Zhou, X., Zhao, M., et al. (2008). Adaptive evolution of UGT2B17 copy-number variation. *Am J Hum Genet, 83*(3), 337-346.

Yang, T. L., Chen, X. D., Guo, Y., Lei, S. F., Wang, J. T., Zhou, Q., et al. (2008). Genome-wide Copy-Number-Variation Study Identified a Susceptibility Gene, UGT2B17, for Osteoporosis. *Am J Hum Genet*.

Yuan, J. H., Li, Y. Q., & Yang, X. Y. (2007). Inhibition of epigallocatechin gallate on orthotopic colon cancer by upregulating the Nrf2-UGT1A signal pathway in nude mice. *Pharmacology., 80*(4), 269-278. Epub 2007 Jul 2026.

Figure Legends

Figure 1. Schematic representation of the mRNA diversity generate by known alternative splicing patterns. Some of these splicing events have been observed in UGTs.

Figure 2. A) Phylogram of the human UGT1A and UGT2B proteins and their corresponding cluster of genes on chromosomes 2 and 4, respectively. Schematic representation of the UGT1A9 (B) and UGT2B7 (C) protein primary structures derived from each locus. Similarities between the two classes of proteins are shown.

Figure 3. Schematic representation of the *UGT1A* gene and the alternative sharing and splicing events taking place at the 5' and 3' ends of the locus. (Top) The previous genomic structure of *UGT1A* included 17 exons: 13 first exons and four common exons (2–5a). We found that the *UGT1A* locus encodes one additional exon (referred to as the new exon 5b) and that the common region comprises five shared exons instead of four, as currently reported in the literature. (Bottom) The newly identified *UGT1A* gene products (mRNAs) are referred to as variants, or v, whereas new proteins are named isoforms, or UGT1A_i, according to Human Gene Nomenclature Guidelines. The previously described 'classical' UGT1As are named UGT1A_i1 and the newly described UGT1A proteins are referred to as UGT1A_i2. Exon 5a and the new exon 5b are alternatively spliced, generating variant UGT1A mRNAs: v1 (containing 5a), v2 (5b) and v3 (5a+5b). The proteins translated from v3 are identical to those encoded by the UGT1A_v2 mRNAs, as they have identical open reading frames. These data suggest the existence of nine additional UGT1A_i2 proteins, increasing to 18 the number of proteins encoded by UGT1.

Figure 4. Known Alternatively spliced UGT2B mRNA transcripts: Depicted here are distinct mRNA transcripts which were discovered by our group and/or be found on the NCBI website. The mRNA sequences of these transcripts are present in the Genbank database, the identification number on the website is also indicated.

Figure 5. Haplotypes of the *UGT2B17*, *UGT2B28* CNV along with the *UGT2B15* polymorphism on chromosome 4. A total of seven different haplotypes of *UGT2B17*, *UGT2B28* and *UGT2B15* have been inferred, using genotyping data. White boxes represent wild-type state (no polymorphism), while deletions are depicted by a crossed-out shaded box and variant nucleotides by a red box.

Figure 6. Current and emerging genomic mechanisms likely underlying interindividual glucuronidation phenotypes.

Table 1. Predicted transcripts from EST clusters at the UGT1 and UGT2 loci.

Locus (Unigene cluster) (coding strand)[1]	**Number of Associated transcripts[2]**	**Transcript ID[3]**	**Exon Boundaries[4]**	**Supportive Seq[5] (gi number)**	**Comments**
UGT1A (Hs.554822) (+ strand)	404	UGT1A.1	182291-182422 183106-183193 183477-183696 187579-187816		Only Common exon 2 to 5
		UGT1A.2	182291-182422 183106-183193 183477-183696 **184788-185275**	19140059 19010676 12671933 31657195	Use of an alternative last exon (exon 5b).
UGT2B4 (Hs.285887) (- strand)	78	UGT2B4.1	1020761-1020006 1018707-1018559 1014436-1014305 1011562-1011475 1010293-1010074 1005776-1005031		Total protein length= 529 aa.
		UGT2B4.2	**1050880-1050553** **1020276-**1020006 1018707-1018559 1014436-1014305 1011562-1011475 1010293-1010074 1005776-1005031	194390249 146076957	A novel first exon is spliced at an alternative 3' splice acceptor site in coding exon 1. Deletion of aa 1 to 150. New ATG start codon in the novel first exon encoding for a new peptide MFFALLHVSSTNGL in frame with the remaining UGT2B4 protein (aa 151-529) .
		UGT2B4.3	1020761-1020006 1018707-1018559 1014436-1014305	37777528	The insertion of an alternative exon between coding exon5 and 6. Prematured ending of the proteins

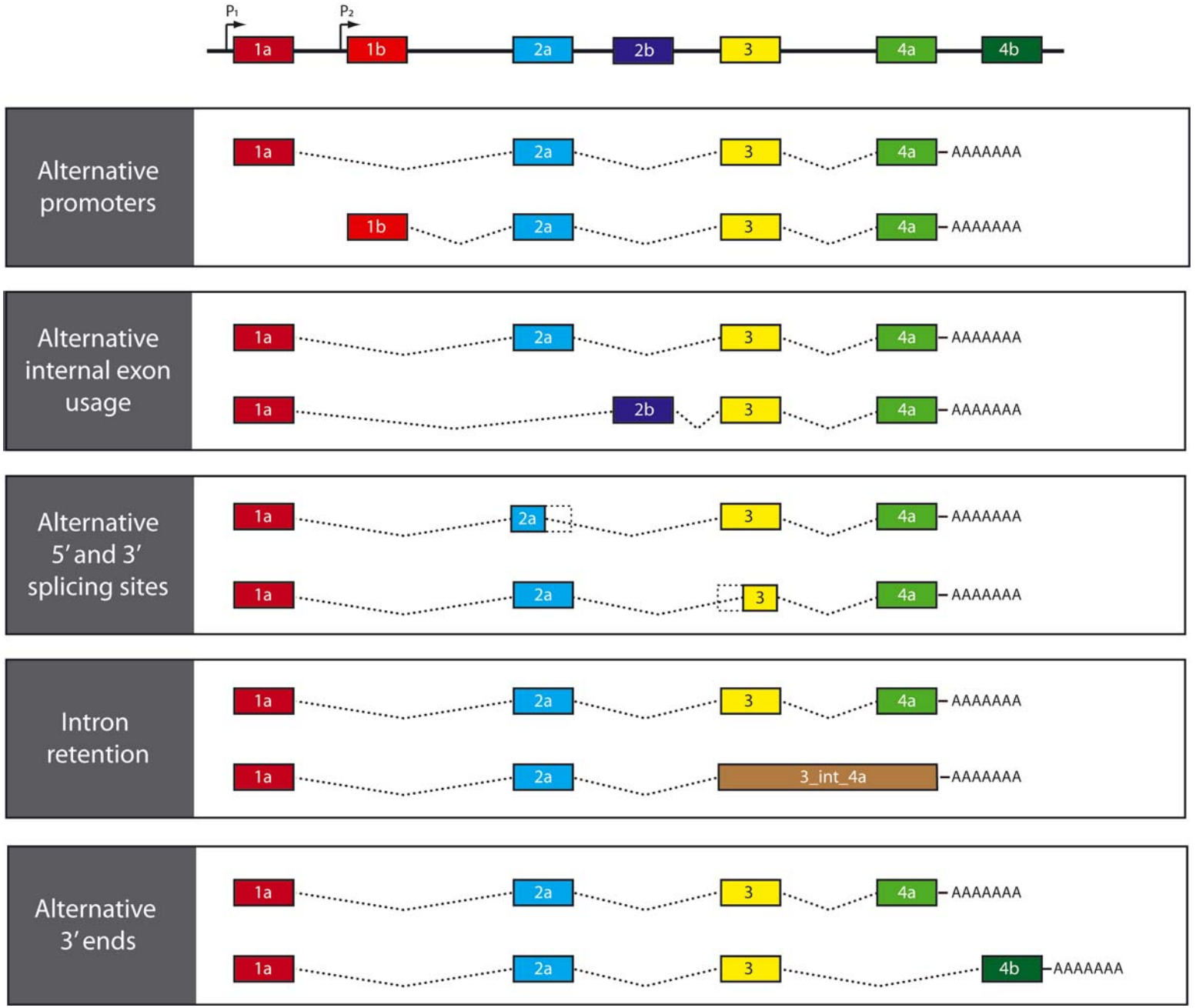
			1011562-1011475 1010293-1010074 **1006131-1005990** 1005776-1005031		at aa 436.
		UGT2B4.4	1020761-**1020259** 1018707-1018559 1014436-1014305 1011562-1011475 1005776-1005031	31023872 37777529	Alternative 5' donor splice site in the exon1, and skipping of the coding exon5. Prematured ending of the proteins at aa 157 followed by 82 UGT2B4-unrelated amino acids.
		UGT2B4_ASE.1	**1050880-1050553** **1020832-**1020006	80808841	Alternative 5'UT-exon, which is spliced with the first coding exon. Full-length UGT2B4 protein is expected.
		UGT2B4_ASE.2	**1050880-1050553** 1018707-1018559	52172943	Alternative 5'UT-exon which is spliced with the coding exon2. The translation is presumed to start in the novel first exon.
UGT2B7 (Hs.654424) (+ strand)	251	UGT2B7.1	621373-622107 623406-623554 627673-627804 632041-632128 632969-633188 637323-637853		Total protein length= 529 aa.
		UGT2B7.2	621373-622107 623406-623554 627673-627804 632041-632128 637323-637853	10735032	Deletion of the coding exon5. Prematured ending of the proteins at aa 363.
		UGT2B7.3	**575965-576468** **614338-614469**	13333538 28109514	Two alternative 5'UT-exons which are spliced with the coding exon2.

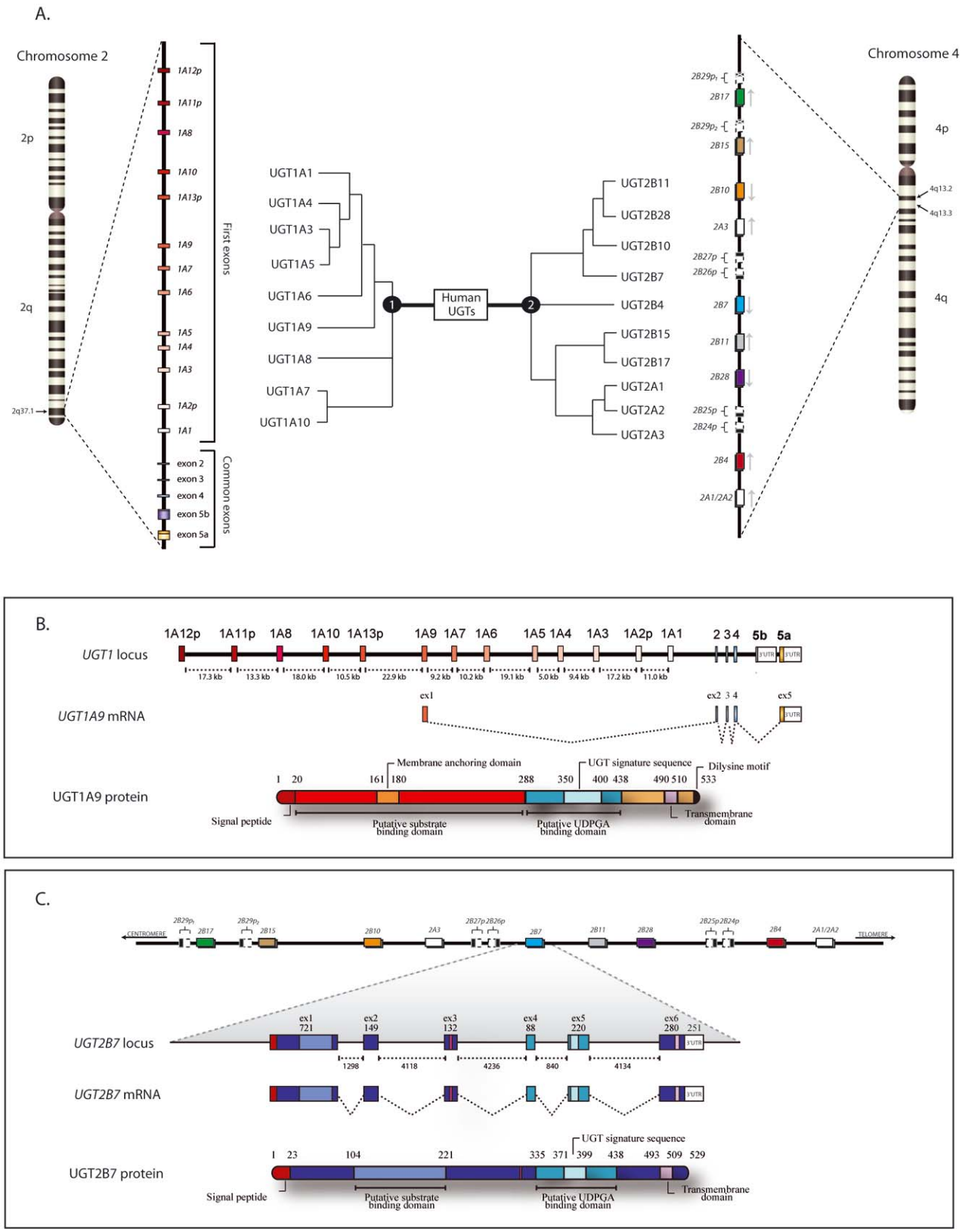
			623406-623554 627673-627804 632041-632128 632969-633188 637323-637853	28111933 15748889 13582705	The translation is presumed to start in the coding exon 2, and would lead to UGT2B7 lacking the substrate binding domain.
		UGT2B7.4	**575965-576468** 623406-623554 627673-627804 632041-632128 632969-633188 637323-637853	12673874	One alternative 5'UT-exon spliced with the coding exon2. The translation is presumed to start in the coding exon 2 as abovementioned.
UGT2B10 (Hs.201634) (+ strand)	75	UGT2B10.1	340861-341603 342895-343042 347136-347268 351276-351363 352195-352414 355466-355745		Total protein length=529 aa.
		UGT2B10.2	340861-**341351** 342895-343042 347136-347268 351276-351363 352195-352414 355466-355745	22119058 37777525 194391069	Alternative 5' donor splice site in the coding exon1. Do not change the coding frame. Deletion of aa 156 to 239.
		UGT2B10_ASE.1	340861-**341038** 342895-343042	52111972	Alternative 5' donor splice site in the coding exon1. Change in the coding frame and Prematured ending of the proteins.
		UGT2B10_ASE.2	351276-351363	13715524	Deletion of the coding exon5.

			355466-355745		
UGT2B28 (Hs.653154) (+ strand)	9	UGT2B28.1	805365-806087 807380-807528 811618-811749 814531-814618 815458-815677 819396-819675		Total protein length= 529 aa.
		UGT2B28.2	**805365-806087** **(□ 805685-806038)** 807380-807528 811618-811749 814531-814618 815458-815677 819396-819675		Deletion of aa 107-224 in the coding exon1.
		UGT2B28.3	805365-806087 807380-807528 811618-811749 819396-819675		Deletion of the coding exon4 and 5.
UGT2A1(2A2) (Hs.225950) (- strand)	21	UGT2A1.1	1172566-1171796 1124260-1124112 1121247-1121116 1120115-1120028 1119562-1119343 1114517-1113283		Total protein length= 527 aa.
		UGT2A1.2	1172566-1171796 **1160688-1160557** 1124260-1124112 1120115-1120028 1119562-1119343	148753379	The use of an alternative exon between the coding exon1 and 2. Deletion of the coding exon3.

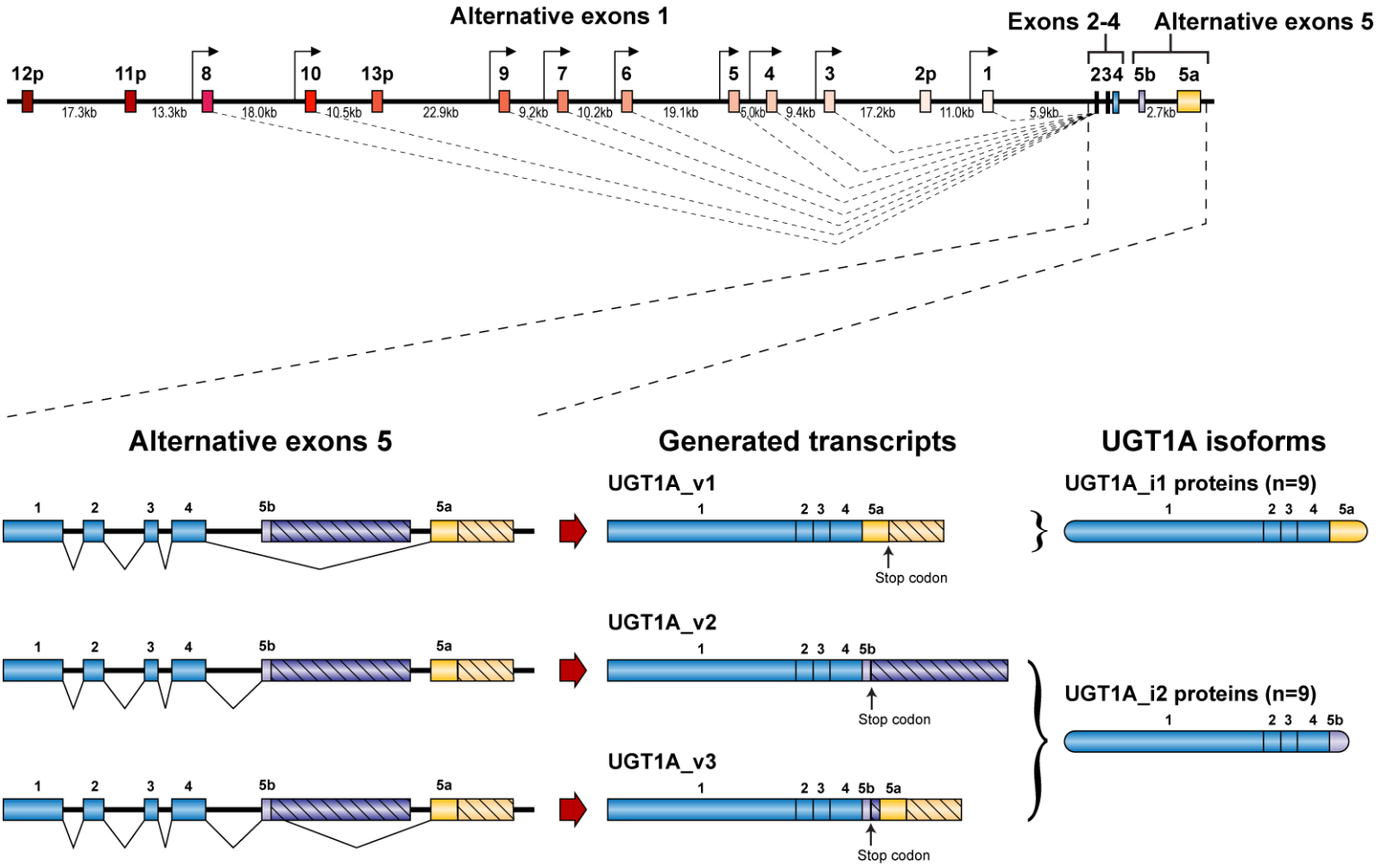
			1114517-1113283		
		UGT2A2.1	1164482-1163765 1124260-1124112 1121247-1121116 1120115-1120028 1119562-1119343 1114517-1113283		Total protein length= 528 aa.
UGT2A3 (Hs.122583) (- strand)	64	UGT2A3.1	476698-475912 470320-470172 457625-457494 456108-456021 455631-455412 454958-453326		Total protein length= 527 aa.
		UGT2A3_ASE.1	470320-470172 **468244-468148** 457625-457494 456108-456021 455631-455412	194373594	Alternative exon between exon 2 and 3.

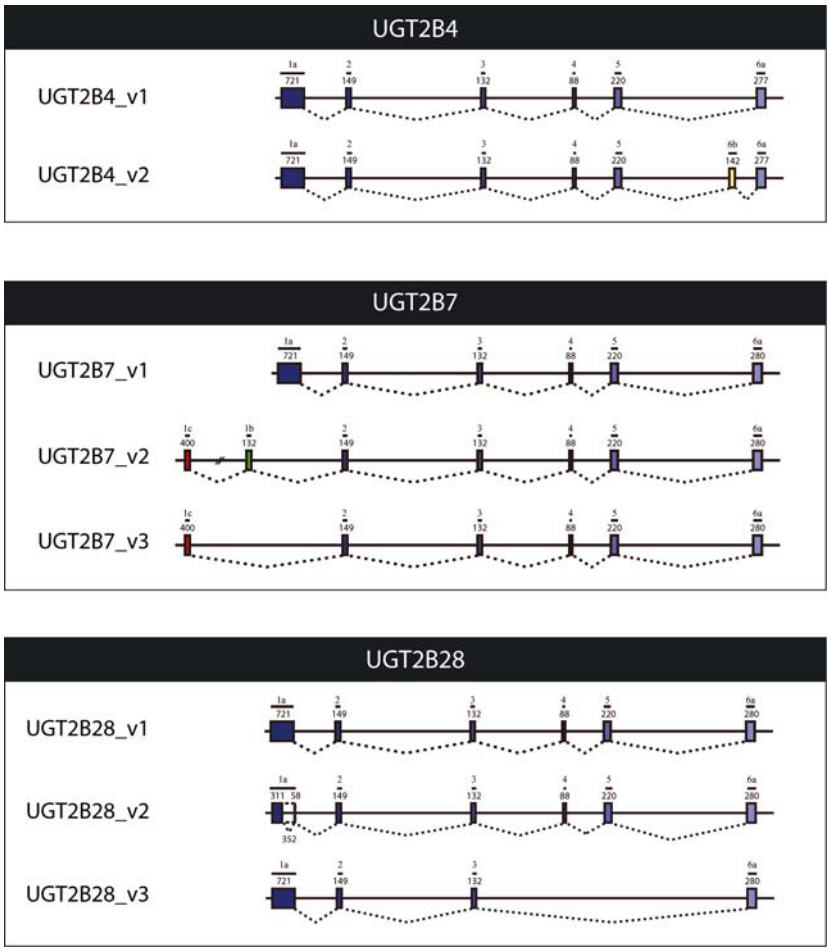
1. The coding strand indicates the orientation of the transcription, (+) from centromere to telomere, (-) from telomere to centromere.
2. The number of transcripts is the total number of mRNA and EST sequences associated with the corresponding locus in the Unigene cluster.
3. It is noteworthy that the transcript ID is an unofficial identifier only used to facilitate the description of each transcript. The transcript ID UGTXXX.X, indicate a full-length mRNA. The transcript ID UGTXXX_ASE.X indicate only the observation of an alternative splicing event (ASE).
4. Nucleotide position is given according to the human genome contig NG_002601 for UGT1A locus, and the contig NT_077444 for all UGT2 genes.
5. The supportive sequences for the major mRNA transcript for each locus, UGTXXX.1, have not been listed.

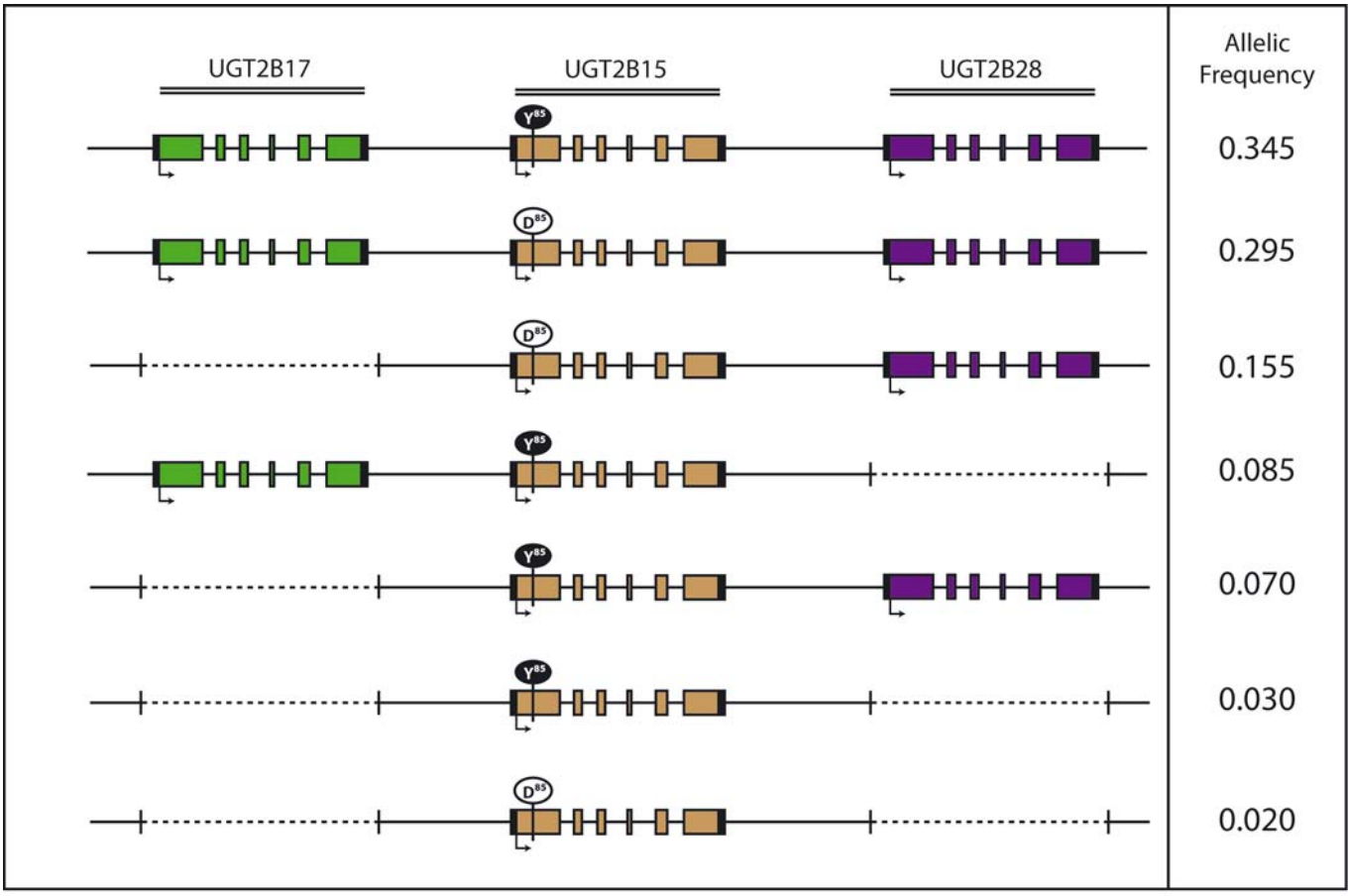

P1
P2
1a
1b
2a
2b
3
4a
4b
Alternative promoters
Alternative internal exon usage
Alternative 5′ and 3′ splicing sites
Intron retention
3_int_4a
Alternative 3′ ends
AAAAAAA


A.
Chromosome 2
2p
2q
2q37.1
1A12p
1A11p
1A8
1A10
1A13p
1A9
1A7
1A6
1A5
1A4
1A3
1A2p
1A1
First exons
exon 2
exon 3
exon 4
exon 5b
exon 5a
Common exons
UGT1A1
UGT1A4
UGT1A3
UGT1A5
UGT1A6
UGT1A9
UGT1A8
UGT1A7
UGT1A10
1
Human UGTs
2
UGT2B11
UGT2B28
UGT2B10
UGT2B7
UGT2B4
UGT2B15
UGT2B17
UGT2A1
UGT2A2
UGT2A3
Chromosome 4
4p
4q
4q13.2
4q13.3
2B29p1
2B17
2B29p2
2B15
2B10
2A3
2B27p
2B26p
2B7
2B11
2B28
2B25p
2B24p
2B4
2A1/2A2
B.
UGT1 locus
1A12p 1A11p 1A8 1A10 1A13p 1A9 1A7 1A6 1A5 1A4 1A3 1A2p 1A1 2 3 4 5b 5a
5'UTR
3'UTR
17.3 kb 13.3 kb 18.0 kb 10.5 kb 22.9 kb 9.2 kb 10.2 kb 19.1 kb 5.0 kb 9.4 kb 17.2 kb 11.0 kb
UGT1A9 mRNA
ex1
ex2 3 4
ex5
UGT1A9 protein
Membrane anchoring domain
UGT signature sequence
Dilysine motif
1 20 161 180 288 350 400 438 490 510 533
Signal peptide
Putative substrate binding domain
Putative UDPGA binding domain
Transmembrane domain
C.
CENTROMERE
TELOMERE
UGT2B7 locus
ex1 721
ex2 149
ex3 132
ex4 88
ex5 220
ex6 280 251
1298 4118 4236 840 4134
UGT2B7 mRNA
UGT2B7 protein
UGT signature sequence
1 23 104 221 335 371 399 438 493 509 529
Signal peptide
Putative substrate binding domain
Putative UDPGA binding domain
Transmembrane domain

UGT1A Locus


Alternative exons 1
Exons 2-4
Alternative exons 5
12p
11p
8
10
13p
9
7
6
5
4
3
2p
1
234
5b
5a
17.3kb
13.3kb
18.0kb
10.5kb
22.9kb
9.2kb
10.2kb
19.1kb
5.0kb
9.4kb
17.2kb
11.0kb
5.9kb
2.7kb
Alternative exons 5
Generated transcripts
UGT1A isoforms
UGT1A_v1
UGT1A_i1 proteins (n=9)
Stop codon
UGT1A_v2
UGT1A_i2 proteins (n=9)
Stop codon
UGT1A_v3
Stop codon


UGT2B4
UGT2B4_v1
UGT2B4_v2
UGT2B7
UGT2B7_v1
UGT2B7_v2
UGT2B7_v3
UGT2B28
UGT2B28_v1
UGT2B28_v2
UGT2B28_v3



2B15_D^{85} = 2B15*1, 2B15_Y^{85} = 2B15*2