Polymorphisms and Haplotypes of the UDP-Glucuronosyltransferase 2B7 Gene Promoter

Dong Gui Hu, Robyn Meech, Lu Lu, Ross A. McKinnon, and Peter I. Mackenzie

Department of Clinical Pharmacology and Flinders Centre for Innovation in Cancer, Flinders University School of Medicine, Flinders Medical Centre, Bedford Park, Australia

Received January 2, 2013; accepted February 21, 2014

ABSTRACT

Identification of functional polymorphisms in the UDP-glucuronosyltransferase 2B7 (UGT2B7) gene predicting interpatient variability in the glucuronidation of drugs that are primarily metabolized by UGT2B7 has been the subject of many studies. These studies have shown linkage disequilibrium (LD) covering the region from −2 kb to 16 kb of the UGT2B7 gene. We identified three novel single-nucleotide polymorphisms (SNPs) and extended this LD in the 5′-upstream direction to cover an additional nine prevalent polymorphisms in the distal −2600- to −4000-base pair (bp) promoter. We further showed complete LD between these distal promoter SNPs and the SNP (802C>T) in exon 2 in a panel of 26 livers. Because of this LD, we showed that all of the 23 prevalent polymorphisms in the 4-kb UGT2B7 promoter are linked together, defining two major haplotypes (i.e., I and II). The addition of the minor allele of a rare polymorphism and allele exchanges between haplotypes I and II generated subhaplotypes of I and II. We demonstrated a higher promoter activity of haplotype II over haplotype I, and this higher activity was abolished by an A to G change at a single SNP (−900A>G). This mutation changed a consensus activating protein-1 (AP-1) site (TGAAGTC) as occurred in haplotype II to a mutated AP-1 site (TGAATCG) as occurred in haplotype I. Finally, we showed that the previously reported Alu element resides exclusively in haplotype I and is a highly conserved CG-rich Alu Y element.

Introduction

The 19 human UDP-glucuronosyltransferases (UGTs) catalyze the transfer of glucuronic acid from the cofactor UDP-glucuronic acid to small hydrophobic compounds. The glucuronic products are generally inactive and water soluble, thus facilitating their excretion from the body. One UGT, UGT2B7, is involved in the glucuronidation of a variety of endobiotics and xenobiotics, including steroid hormones, fatty acids, bile acids, retinoids, nonsteroidal anti-inflammatory drugs, epiurubicin, morphine, and mycophenolic acid (MPA) (Radominska-Pandya et al., 2001).

Morphine is metabolized primarily by UGT2B7 in the liver to form morphine-3-glucuronide (M3G, the major inactive metabolite) and morphine-6-glucuronide (a more potent analgesic compound compared with morphine) (Coffman et al., 1997). The hepatic expression and enzyme activity of UGT2B7 are thus critical for systemic clearance of UGT2B7 substrates such as morphine. Studies have shown a huge variability in morphine glucuronidate-morphine plasma concentration ratios in cancer patients on chronic morphine therapy (McQuay et al., 1990; Tiseo et al., 1995; Klepstad et al., 2000; Holthe et al., 2002, 2003; Fladvat et al., 2013). This interpatient difference in morphine glucuronide-to-morphine ratios is attributable to genetic and nongenetic factors. Nongenetic factors may include variability in daily morphine dosage, medical conditions, comedinations, cross-membrane transport, hepatic blood flow, volume of distribution, protein binding, and renal elimination (Holthe et al., 2002, 2003; Court, 2010; Fladvat et al., 2013).

Studies have shown up to a 7-fold difference in hepatic UGT2B7 mRNA levels (Congiu et al., 2002) and up to 14-fold difference in M3G formation in human liver microsomes (Fisher et al., 2000; Innocenti et al., 2008). Thus, identification of polymorphisms that might influence the expression and enzymatic activity of UGT2B7 has been the subject of many studies. Early studies focused on the SNP 802C>T (rs7439366) located in exon 2, which produces two variant UGT2B7 proteins with either histidine or tyrosine at amino acid 286. This SNP, first reported by our group (Jin et al., 1993), is the only prevalent SNP within the UGT2B7 coding sequence in different ethnic populations with a global minor allele frequency (MAF) of around 0.35 according to the 1000Genome Project (May 2011 release). Most studies show that this SNP has little impact on substrate selectivity and enzyme activity (Coffman et al., 1997, 1998; Blasker et al., 2000; Holthe et al., 2002; Court et al., 2003; Sawyer et al., 2003; Wiener et al., 2004; Lin et al., 2005; Thibaudeau et al., 2006; Innocenti et al., 2008; Court, 2010). Further studies on polymorphisms in the promoter and elsewhere in the UGT2B7 gene have led to the discovery of a tight linkage disequilibrium (LD) covering the region from the proximal 2-kb promoter to the 16-kb entire gene, including all exons and introns (Sawyer et al., 2008). This SNP, identified by our group (Jin et al., 1993), is the only prevalent SNP within the UGT2B7 coding sequence in different ethnic populations with a global minor allele frequency (MAF) of around 0.35 according to the 1000Genome Project (May 2011 release). Most studies show that this SNP has little impact on substrate selectivity and enzyme activity (Coffman et al., 1997, 1998; Blasker et al., 2000; Holthe et al., 2002; Court et al., 2003; Sawyer et al., 2003; Wiener et al., 2004; Lin et al., 2005; Thibaudeau et al., 2006; Innocenti et al., 2008; Court, 2010).

Further studies on polymorphisms in the promoter and elsewhere in the UGT2B7 gene have led to the discovery of a tight linkage disequilibrium (LD) covering the region from the proximal 2-kb promoter to the 16-kb entire gene, including all exons and introns (Sawyer et al., 2008). This SNP, identified by our group (Jin et al., 1993), is the only prevalent SNP within the UGT2B7 coding sequence in different ethnic populations with a global minor allele frequency (MAF) of around 0.35 according to the 1000Genome Project (May 2011 release). Most studies show that this SNP has little impact on substrate selectivity and enzyme activity (Coffman et al., 1997, 1998; Blasker et al., 2000; Holthe et al., 2002; Court et al., 2003; Sawyer et al., 2003; Wiener et al., 2004; Lin et al., 2005; Thibaudeau et al., 2006; Innocenti et al., 2008; Court, 2010). Further studies on polymorphisms in the promoter and elsewhere in the UGT2B7 gene have led to the discovery of a tight linkage disequilibrium (LD) covering the region from the proximal 2-kb promoter to the 16-kb entire gene, including all exons and introns (Sawyer et al., 2008). Two SNPs were identified in the promoter that negatively (−138A>G, rs7668258) or positively (−900A>G, rs7439366) regulate the expression and enzymatic activity of UGT2B7 (Duguay et al., 2004). This mutation changed a consensus activating protein-1 (AP-1) site (TGAAGTC) as occurred in haplotype II to a mutated AP-1 site (TGAATCG) as occurred in haplotype I. Finally, we showed that the previously reported Alu element resides exclusively in haplotype I and is a highly conserved CG-rich Alu Y element.
UDP Glucuronosyltransferase 2B7 Promoter Haplotypes

855

Table 1

Primer sequences used for cloning, sequencing, genotyping, and mutagenesis (5'-3')

<table>
<thead>
<tr>
<th>Cloning</th>
<th>Primer sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B7-1046F</td>
<td>ATAGACCGGTGACCTGAAAGGAATACCTCA</td>
</tr>
<tr>
<td>2B7-2076F</td>
<td>ATGTACGCGTTGCAGGTCTTCTAAGCATA</td>
</tr>
<tr>
<td>2B7-3026F</td>
<td>ATAGACCGGTGACCTGAAAGGAATACCTCA</td>
</tr>
<tr>
<td>2B7-4026F</td>
<td>ATCAACGCGATGTGATGCCTAGGAAAGGCT</td>
</tr>
<tr>
<td>2B7-4925F</td>
<td>TACTACGGTGGAGCCCTGAACCTTCAAT</td>
</tr>
<tr>
<td>2B7-ATC1862R</td>
<td>ATATCTGGACCTGGTGCAATGCAAT</td>
</tr>
<tr>
<td>Sequencing</td>
<td>GGTATATGTCGCCGCTGCTC</td>
</tr>
<tr>
<td>2B7-1917R</td>
<td>TCATTAGACCTCTTATGAGC</td>
</tr>
<tr>
<td>2B7-2907R</td>
<td>GCGGCTACAGGGGCTGAAG</td>
</tr>
<tr>
<td>2B7-3867R</td>
<td>CTGGATACAAAAGGTAAAGA</td>
</tr>
<tr>
<td>Genotyping</td>
<td>GGTATATGTCGCCGCTGCTC</td>
</tr>
<tr>
<td>2B7-2579F</td>
<td>GCATATGCTTCTCTGGTG</td>
</tr>
<tr>
<td>2B7-1862R</td>
<td>CCACCACTCGAAATGAGTTCAG</td>
</tr>
<tr>
<td>Mutagenesis</td>
<td>GGTGAGAAAGTGATGCGGAGAGCAACAGTACAA</td>
</tr>
<tr>
<td>2B7-900 A to G MTF</td>
<td>GGTGAGAAAGTGATGCGGAGAGCAACAGTACAA</td>
</tr>
<tr>
<td>2B7-900 A to G MTR</td>
<td>GGTGAGAAAGTGATGCGGAGAGCAACAGTACAA</td>
</tr>
</tbody>
</table>

Materials and Methods

Source of Liver Tissues. Human liver tissues from 26 Caucasian donors were obtained from the liver bank of the Department of Clinical Pharmacology of Flinders Medical Centre, Flinders University of South Australia, Australia. Genomic DNA and human liver microsomes were prepared from the liver tissues as described previously (Bhasker et al., 2000). Approval for the use of human liver tissues for pharmacogenetics and pharmacokinetics studies was granted by the Flinders Medical Centre Research Ethics Committee.

Cloning and Haplotypic Determination of the UGT2B7 Promoter. Recent studies have shown that UGT2B7 transcription is directed by two mutually exclusive promoters, promoter 1 (canonical) and promoter 1a (alternative) (Menard et al., 2011, 2013). Promoter 1 governs the transcription of the classic exon 1-containing UGT2B7 transcripts encoding the functional UGT2B7 enzyme, whereas promoter 1a is responsible for transcription of the alternatively spliced exon 1-containing UGT2B7 transcripts resulting in inactive UGT2B7 proteins. Promoter 1a is approximately 44 kb upstream from promoter 1. Among the four alternatively spliced exons (1a, 1b, 1c, and 1d), exon 1b is the closest variant exon 1 to promoter 1 and resides 6871 base pairs (bp) upstream from promoter 1. The UGT2B7 gene fragment cloned and analyzed in this study (~1 to ~4926) was downstream of all the variant exons 1 and contained promoter 1. For simplicity, we refer to the UGT2B7 promoter 1 as the UGT2B7 promoter throughout this article.

The human UGT2B7 promoter region from nucleotides ~1 to ~4026 or from nucleotides ~1 to ~4926 (hereafter termed 4-kb or 5-kb UGT2B7 promoter) relative to the translation start site was amplified from a commercial human genomic DNA sample (Roche Diagnosis, Indianapolis, IN) by Phusion hot-start high-fidelity DNA polymerase (Thermo Fisher Scientific, Pittsburgh, PA) using a specific forward primer (2B7-1046F or 2B7-4926F) and a common reverse primer (2B7-1R). The resultant polymerase chain reaction (PCR) amplicons were subsequently cloned into the MluI and XhoI sites of the pGL3-basic vector (Promega, Madison, WI).

According to the manufacturer’s instructions (Roche), the genomic DNA sample used for UGT2B7 promoter cloning was prepared from a pool of human blood (buffy coat) from 80–100 individuals of mainly German nationality. As a consequence, each cloned promoter represented a specific haplotype in the blood sampling population. As such, genotyping of individual cloned promoters would allow us to define the UGT2B7 promoter haplotypes. We obtained 16 clones carrying a 5-kb UGT2B7 promoter and four clones containing a 4-kb UGT2B7 promoter. The cloned promoter fragments in these 20 clones were sequenced in full bidirectionally using 10 primers as given in detail in Table 1.

Cell Culture and Luciferase Reporter Assays. The liver cancer cell line (HepG2), the breast cancer cell lines (MCF-7, ZR75.1, and MDA-MB-435), the prostate cancer cell lines (LNCaP, VCaP, PC3, and DU145), and the kidney HEK293 cell line were purchased from American Type Culture Collection (Manassas, VA). The HepG2 and VCaP cell lines were maintained in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum, whereas the remaining seven cell lines were cultured in RPMI 1640 medium containing 5% fetal bovine serum. For transient transfection, cells were plated into 96-well plates and reached approximately 60% confluence before transfection. Cells were transfected per well with 0.8 ng of pRL-null vector (Promega), 200 ng of each promoter luciferase construct, and 0.8 μl of Lipofectamine2000 (Invitrogen, Carlsbad, CA). Twenty-four hours post-transfection, the transfection medium was replaced by fresh medium and cells were cultured for a further 24 hours. Forty-eight hours post-transfection, the cells were harvested to quantify Firefly and Renilla luciferase activities using the Dual-Luciferase Reporter Assay system (Promega) as described previously (Hu and Mackenzie 2010). Firefly (Photinus pyralis) luciferase activity was normalized against the Renilla (Renilla reniformis) activity. The relative activity of the UGT2B7 promoter was presented as the -fold change compared with that of the control promoterless pGL3 vector.

Mutagenesis. Mutagenesis was conducted using the QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). Because of the SNP, ~900A>G, at the UGT2B7 promoter, haplotype II has a consensus AP-1 site of TGAAGCA, whereas haplotype I has a mutated activating protein-1 (AP-1) site (TGAAGCG, mismatched nucleotide underlined). An A-to-G mutation at ~900 was generated in the 5-kb UGT2B7 promoter construct carrying the haplotype II promoter (termed MT) as previously reported (Hu and Mackenzie 2009). The forward and reverse primers used for mutagenesis are listed in Table 1.

Genotyping of the Alu Insertion/Deletion Polymorphism at the UGT2B7 Promoter. The Alu insertion/deletion polymorphism (rs71955301) at the UGT2B7 promoter was previously reported; this Alu element (324 bp) is inserted between nucleotides ~2067 and ~2390 relative to the translation start site at the UGT2B7 promoter carrying haplotype I (Innocenti et al., 2008; Nakamura et al., 2008). To genotype this polymorphism, the promoter region between nucleotides ~1862 and ~2579 spanning the Alu element was amplified from 35 genomic DNA samples, extracted from 26 liver tissues and nine human cell lines using primers 2B7-2579F and 2B7-1862R (Table 1). Briefly, PCR was performed in a 50-μl reaction comprising 200 ng of each genomic DNA sample, 1 U of Phusion hot-start high-fidelity DNA polymerase (Thermo Scientific), deoxyribonucleotide triphosphates (200 μM each), and the forward and reverse primers (200 μM each). Then 10-μl aliquots of each amplicon were resolved on 4% agarose gels and imaged on a UV transilluminator. Homozygosity of the Alu insertion was defined by the presence of a single amplicon of 718 bp, whereas homozygosity of the Alu deletion was defined by the presence of a single amplicon of 394 bp. Heterozygosity was defined by the presence of both amplicons.

Immunoblotting. We previously developed an anti-UGT2B7 antibody (Kerdin et al., 2009). We herein performed Western blotting assays to further verify its specificity. As shown in Supplemental Fig. 1, this antibody displayed high binding specificity toward UGT2B7 protein; weakly cross-reacted with UGT2B4 and 2B28 proteins; and had no cross-reactivity with UGT2B10, 2B15, and 2B17 proteins. This antibody did not cross-react with any of the nine UGT1A family proteins (data not shown). After verification of the specificity of this antibody, we performed Western blotting assays using this antibody to quantify the UGT2B7 protein levels in 12 human liver microsomes. Equal amounts of human liver microsomes (30 μg) were resolved on SDS-polyacrylamide gels (10%) and transferred to nitrocellulose membrane. Blots were probed with an anti-UGT2B7 antibody, followed by horseradish peroxidase conjugated secondary antibodies.
peroxidase-conjugated donkey anti-rabbit antibody, as previously reported (Kerdpin et al., 2009). SuperSignalWest Pico Chemiluminescent substrate (Thermo Scientific) was used for immunodetection, and images were obtained using an ImageQuant LAS 4000 luminescent image analyzer (GE Healthcare Life Sciences, Piscataway, NJ). Band intensity of the immunoblots and background subtraction was carried out using imaging software, Multi Gauge Ver3.0 (FUJIFILM, Tokyo, Japan).

Statistical Analysis. Analysis of variance in the activities of UGT2B7 promoters carrying different haplotypes and in the UGT2B7 protein levels of different liver microsomes was assessed using Student’s t test. A P value < 0.05 was considered statistically significant.

Results

Polymorphisms in the 5-kb UGT2B7 Promoter in the dbSNP Database. There are 131 polymorphisms in the 5-kb UGT2B7 promoter deposited in the NCBI dbSNP database (accessed December
2013), including 116 single-nucleotide polymorphisms (SNPs) and 15 insertion/deletions (INDELs). This equates to one polymorphism every 38 bp; however, these polymorphisms distribute unevenly across this region. There are 21, 22, 33, 38, and 17 polymorphisms within the 1st, 2nd, 3rd, 4th, and 5th kb of the UGT2B7 promoter region, respectively. The length of the INDELs varies from a single nucleotide (rs34590280, rs71204074, rs138316705, rs138844955, rs57058697, and rs150919624), two nucleotides (rs200830029, eight nucleotides (rs16614134, rs33937977, rs35014193, and rs113605769), 21 nucleotides (rs6148500 and rs377337102), to 324 nucleotides (rs71955301). Of particular interest is the INDEL of a 324-bp Alu element (rs71955301) that spans between nucleotides 22067 and 22390 relative to the translation start site.

According to the 1000Genome phase 1 genotype dataset from 1094 worldwide individuals (May 2011 release), 35 of the 131 polymorphisms have a global MAF > 1%. Of these 35 polymorphisms, only 23 are prevalent with a similar MAF of approximately 0.35.

**Polymorphisms in the 5-kb UGT2B7 Promoter Found in This Study.** A comparison of the 20 cloned promoter sequences with the UGT2B7 promoter reference sequence (NC_000004.11) identified 26 known polymorphisms and three novel single-nucleotide substitutions (Fig. 1, A–C). The known polymorphisms included 22 of the aforementioned 23 prevalent SNPs/INDELs, rs11730928C>T at 4342 (MAF 0.14), rs62296941A>G at 3127 (MAF 0.12), rs73823857T>C at 1681 (MAF 0.03), and rs28458345G>T at 2821 (MAF 0.03). In addition, three novel single-nucleotide substitutions, rs28458345G, rs62296941A, and rs11730928C, were seen in a single individual clone (Fig. 1, A–C). According to the NCBI dbSNP database and as shown in Fig. 1D, four INDELs (rs35014193, rs61641343, rs33937977, and rs113605769) are present within six nucleotides between −3965 and −3970. One of the INDELs (rs113605769) has a MAF of 0.35, and the MAFs for the three other INDELs are unknown. Based on this frequency, we expected to see the minor allele of INDEL (rs113605769) in some of the 20 clones; however, none of the clones carried this minor allele. We did observe the minor allele of another INDEL (rs61641343) in five clones; thus, this INDEL represents the 23rd prevalent polymorphism in the 5-kb UGT2B7 promoter observed in this study (MAF 0.25 in our study).

**Haplotype Structures of the 5-kb UGT2B7 Promoter.** A comparison of the 20 cloned promoter haplotypes identified eight unique subhaplotypes that belonged to two major haplotypes: haplotype I and haplotype II (Fig. 2). Linkage of all the minor alleles or all the major alleles of the 23 prevalent SNPs/INDELs observed in the present study gave rise to haplotype I (three clones) or haplotype II (8 clones). As shown in Fig. 2, haplotypes I and II were termed herein to be consistent with the haplotype nomenclatures reported by Duguay et al. (2004). Haplotypes I and II are named in other reports as *2 and

![Fig. 2. Polymorphisms are in tight LD at the 4-kb UGT2B7 promoter. The UGT2B7 promoter region (4–5 kb) was amplified from a commercial genomic DNA that was prepared from a pool of human blood (buffy coat) of 80–100 individuals of mainly German nationality and subsequently cloned into the pGL3-basic vector. Shown are 29 polymorphisms and eight haplotypes that were found in 20 cloned promoter constructs. Both minor and major alleles (bold) of these polymorphisms are indicated by arrows. Twenty-three prevalent alleles with a similar global minor allele frequency of around 0.35 are indicated by *As a result of LD in the promoter, eight subhaplotypes of haplotype I (subtypes I, I-1, and I-2) and II (subtypes II, II-1, II-2, II-3, and II-4) can be defined. Alleles that are swapped between haplotypes I and II and the nucleotide that was mutated in MT promoter constructs are highlighted in bold. Previously reported haplotypes of the UGT2B7 proximal promoter are shown for comparison. a, INDEL of a 324-bp Alu element between nucleotides −2067 and −2390. b, INDEL of TAACAGAC between nucleotides −3966 and −3970. (Holthe et al., 2003. Duguay et al., 2004. Nakamura et al., 2008).](image-url)
Substitution at polymorphisms in all the 16 cloned 5-kb promoters (Holthe et al., 2003; Duguay et al., 2004) at haplotype II-1 (Fig. 2). This indicates that the LD at the promoter is not complete as previously proposed (Sawyer et al., 2003; Saito et al., 2006). Furthermore, replacement of the major allele with the minor allele at a rare polymorphic site gave rise to an additional four subhaplotypes (I-2, II-2, II-3, and I-4).

Activities of UGT2B7 Promoter Haplotypes. To test the potential impact of different haplotypes on promoter activity, we transfected UGT2B7 promoter luciferase constructs carrying six different haplotypes (I, I-1, I-2, II, II-2, and II-3) into the most frequently used hepatic cell line cell, namely, HepG2. As shown in Fig. 3, all six haplotypes displayed strong promoter activity compared with the activity of the control promoter-less pGL3-basic vector. The three haplotypes I (I, I-1, and I-2) had a similar activity (P > 0.05 in all cases). Likewise, promoter activity was similar among the three haplotypes (II, I-1, and II-2) (P > 0.05 in all cases). These results indicate that the nucleotide substitutions that differentiate the subhaplotypes had no impact on promoter activity. However, the activities of haplotypes II were about 2-fold higher compared with those of haplotypes I (P < 0.05 in all cases). A similar 2-fold higher activity of haplotype II over haplotype I in HepG2 cells has been previously reported and attributed to the SNP (SNP –900A>G) (Duguay et al., 2004; Nakamura et al., 2008). However, the UGT2B7 promoter analyzed in these studies covered only the proximal six (Duguay et al., 2004) or 14 (Nakamura et al., 2008) prevalent polymorphisms. To assess whether the SNP (SNP –900A>G) also contributed to the 2-fold difference in promoter activity between the 5-kb haplotype I and haplotype II promoters, we compared promoter-luciferase constructs carrying either the wild-type haplotype II –900A allele or the variant haplotype II harboring an A-to-G mutation at –900. Transfection of these two constructs into HepG2 cells revealed that the A-to-G mutation reduced the activity of the haplotype II promoter to a level equivalent to that of the haplotype I promoter (Fig. 3). This indicates that the SNP (SNP –900A>G) is functional in the 5-kb UGT2B7 promoter.

An AluY Element in the UGT2B7 Promoter. Alu elements are classified into three subfamilies based on the time of their insertion into the human genome and mutations acquired in their sequences since their insertion, namely, AluJ, AluS, and AluY (Ahmed et al., 2013; Price et al., 2004). The AluY subfamily is the evolutionarily youngest Alu family (Ahmed et al., 2013). A canonical Alu element (–300 bp) has a characteristic two-armed (left and right) structure with a polyA tract of varying lengths between the two arms and at the 3′-terminal end (Cheng et al., 2012; Kaer and Speek, 2013). As shown in Fig. 4, the UGT2B7 Alu element is inserted in an antisense orientation in the promoter between nucleotides –2067 and –2390. This –324-bp Alu element differs from the AluY consensus sequence by only two single nucleotide substitutions in the right arm and an insertion of nine “A” in the central polyA tract. This Alu element has all the features of a classic Alu element, including the highly conserved left and right arms, a central polyA tract, an internal RNA polymerase III promoter consensus sequence (boxes A and B), two Alu restriction sites, a polyA tail, and a target-site duplication signal sequence of 5′AGAAACC3′ flanking the element. A consensus AluY element is considered a typical CpG island containing 24 CpG dinucleotides, one of which is mutated in the UGT2B7 Alu element. Overall, the UGT2B7 Alu element can be considered a highly conserved AluY element and potential target of CpG methylation.

Complete Linkage Disequilibrium between the AluY INDEL and the 802C>T SNP in the UGT2B7 Gene in a Panel of 26 Liver Tissues. The genotypes of the AluY INDEL polymorphism at the UGT2B7 promoter were obtained for nine human cell lines and 26 liver tissues. Of the nine cell lines, two (MCF7 and MDA-MB-453) were homozygous for the AluY insertion, three cell lines (HepG2, HEK293, and PC3) were homozygous for the AluY deletion, and four cell lines (ZR75, LNCaP, VCaP, and Du145) were heterozygous for the AluY insertion and deletion (Fig. 5A). These results indicate that this INDEL is prevalent in commonly used human cell lines. Of the 26 liver tissues, six were homozygous for the AluY insertion, seven were homozygous for the AluY deletion, and 13 were heterozygous for the AluY insertion and deletion (Fig. 5, B–D). The frequencies of the AluY insertion and deletion were estimated to be 48.1% and 51.9%, respectively.

We previously reported the genotypes of SNP 802C>T for the same panel of liver samples (Bhasker et al., 2000). As shown in Fig. 5, B–D, the samples homozygous for the AluY insertion and deletion were found to be homozygous for 802T and 802C, respectively.
whereas the samples heterozygous for the AluY insertion and deletion had both alleles of 802T and 802C. We further amplified the 5-kb UGT2B7 promoter region from two liver genomic DNA samples that were homozygous for either 802T or 802C. Our sequencing of the resultant amplicons showed that the liver samples homozygous for 802T and 802C were homozygous for haplotype I and II, respectively. These results indicate that the AluY insertion was linked to 802T and the AluY deletion linked to 802C. In conclusion, the AluY INDEL is in complete LD with SNP 802C. Haplotype II Associated with Higher UGT2B7 Protein Levels than Haplotype I in Human Liver Microsomes. The liver microsomes analyzed included four samples homozygous for haplotype I, four samples homozygous for haplotype II, and four samples heterozygous for haplotype I and II. As shown in Fig. 6, UGT2B7 proteins were detected in all 12 liver microsomes and the positive control sample but not in the negative control sample. After background subtraction, quantitative analyses of the band intensity showed that the UGT2B7 protein levels in the liver microsomes homozygous for haplotype II was significantly higher than those in the liver microsomes homozygous for haplotype I ($P < 0.05$). This finding was consistent with our observed higher promoter activity of haplotype II over haplotype I, as shown in Fig. 3.

Discussion

Sawyer et al. (2003) first reported that $-161C>T$ at the proximal promoter was in complete LD with 802C>T at exon 2 in the UGT2B7 gene. This LD has been confirmed and extended in both directions spanning the gene from the 2-kb proximal promoter to the terminal exon (Hirota et al., 2003; Holthe et al., 2003; Sawyer et al., 2003; Duguay et al., 2004; Saeki et al., 2004; Saito et al., 2006; Innocenti et al., 2008; Nakamura et al., 2008; Hwang et al., 2010; Tian et al., 2012). This LD has been reported in different ethnic groups, including Caucasian (Holthe et al., 2003; Duguay et al., 2004; Innocenti et al., 2008), Chinese (Tian et al., 2012), Korean (Hwang et al., 2010), and Japanese populations (Hirota et al., 2003; Saeki et al., 2004; Saito et al., 2006; Nakamura et al., 2008). The present study further extended this LD in the 5'-upstream direction spanning the $-2000$ to $-4000$-bp distal promoter region. All 23 prevalent polymorphisms in the 4-kb UGT2B7 promoter found in the present study are linked to give rise to haplotype I and haplotype II. Haplotype I has the minor allele at all 23 prevalent polymorphic sites, whereas haplotype II contains the major allele at all 23 prevalent polymorphic sites, whereas haplotype II contains the major allele at all 23 prevalent polymorphic sites. This LD should facilitate association studies by using a single haplotype-tagging polymorphism representative of all 23 polymorphisms. For example, the SNP ($-161C>T$) was previously used as a Tag SNP in predicting genotypes of five linked SNPs ($-327G>A$, $-900A>G$, $-1112C>T$, $-1299C>T$, and $-1306G>A$) at the proximal UGT2B7 promoter (promoter 1) (Holthe et al., 2003). However, it is of note that this approach may lead to errors in genotype prediction when allele exchanges between haplotypes I and II occur, as observed in the present study. However, this LD makes it difficult or impossible to identify the functionally associated polymorphism (s), as the functional polymorphism could be the genotyped polymorphism or...
a polymorphism elsewhere in the gene but linked to the genotyped polymorphism.

We showed in this study that the addition of the minor allele of a rare polymorphism to haplotype I or II generated subhaplotypes of I or II. We expect that a huge number of such subhaplotypes exist in the population given that there are nearly 100 polymorphisms with a MAF, 0.03 within the 5-kb $UGT2B7$ promoter. We showed in the present study that four rare polymorphisms ($-1681T>C$, $-2036G>A$, $-22821G>T$, and $-23127A>G$) had no impact on promoter activity; however, it remains possible that other rare polymorphisms in the gene may influence promoter activity. Indeed, Duguay et al. (2004) reported that the presence of the minor allele A of the SNP ($-138G>A$) with a MAF of 0.03 decreased the activity of haplotype I in the $-1362/+34$ promoter context by approximately 7-fold in HepG2 cells. This study also showed that morphine-treated cancer patients carrying this minor allele A at $-138$ had reduced morphine glucuronide-to-morphine serum ratios compared with the ratios of the control group (Duguay et al., 2004). Interestingly, the SNP $-138G>A$ is located within the promoter region B that we previously found to be bound by an uncharacterized transcription factor complex in HepG2 cells (Ishii et al., 2000). Thus, SNP $-138G>A$ may modulate promoter activity through altering the binding of transcription factors. We propose that functional rare polymorphisms like this SNP at the promoter may contribute to the reported individual variability in $UGT2B7$ expression and enzymatic activity (Holthe et al., 2003; Innocenti et al., 2008).

The AP-1 transcription factors comprise Jun (c-Jun, JunB, and JunD) and Fos (c-Fos, FosB, Fra-1, and Fra-2) subfamilies. AP-1 proteins bind as dimers to their cognate regulatory elements $[5'\text{-TGA(G/C)TCA-3'}]$ at promoters and enhancers to modulate gene transcription (Hess et al., 2004). The SNP ($-900A>G$) at the $UGT2B7$ promoter is contained within a putative AP-1 binding site $[\text{TGAGTC(A/G)}]$. Haplotype II has a consensus AP-1 site $[\text{TGAGTC A}]$, whereas haplotype I has a mutated AP-1 site $[\text{TGAGTC G}]$. We showed that the activity of the haplotype II 5-kb promoter was significantly higher than that of haplotype I in HepG2 cells, and this higher activity was abolished by a single A-to-G change at $-900$. Previous studies reported a similar higher promoter activity of haplotype II over haplotype I in the context of much shorter promoter regions; however, the functionality of this SNP was not verified by mutagenesis in these studies (Duguay et al., 2004; Nakamura et al., 2008). To further support the importance of this SNP, recent studies have shown that the production of acyl- 

![Fig. 5. The INDEL of the AluY element at the $UGT2B7$ promoter is in complete linkage disequilibrium with the SNP $802C>T$ in exon 2 in a panel of 26 liver tissues. The AluY-containing promoter region between nucleotides $-1862$ and $-2579$ was amplified from 35 genomic DNA samples, extracted from 26 liver tissues and nine human cell lines (MCF7, ZR75, MDA-MB-453, HepG2, LNcAP, HEK293, VCaP, PC3, and Du145) by PCR using Phusion hot start high-fidelity DNA polymerase. 10 ml of each of the amplicons was analyzed on agarose gels as shown in (A) for cell lines and (B, C, and D) for liver tissues. Homozygosity of the AluY insertion was defined by the presence of a single amplicon of 718 bp, whereas homozygosity of the AluY deletion was defined by the presence of a single amplicon of 394 bp. Heterozygosity of the AluY insertion and deletion was defined by the presence of both amplicons. The previously reported genotypes of the SNP $802C>T$ for the same panel of liver tissues are shown under the gel images. “Control 1 and Control 2” were PCR products from samples known to be homozygous for the AluY insertion and deletion, respectively. Control 3 comprised PCR products from sample heterozygous for the Alu Y polymorphism. M was DNA marker (100-bp ladder).](https://www.aspetjournals.org/article/62727/fig5)
HepG2 cells. UGT2B7 promoter consensus AP-1 site as Fra-2/JunD dimers in shown). Collectively, these data suggest that AP-1 proteins bind the occupancy of c-Fos and c-Jun at this promoter region (data not haplotype II carrying the consensus AP-1 site). Our results showed no assays to assess whether the c-Fos/c-Jun dimers could also bind to endothelial cells. We performed chromatin immunoprecipitation c-Fos/c-Jun dimers bind to the same region in human umbilical vein cipitation sequencing data from the Encyclopedia of DNA Elements, in the NCBI GRCh37/hg19 Assembly, recent chromatin immunopre- phoretic mobility shift assays; however, the identity of the tran- (Darbari et al., 2008; Joly et al., 2012).

Fig. 6. Haplotype II is associated with higher UGT2B7 protein levels than haplotype I in human liver microsomes. Human liver microsomes analyzed included four samples homozygous for haplotype I, four samples homozygous for haplotype II, and four samples heterozygous for haplotype I and II. Whole-cell lysates from HEK293 cells stably overexpressing UGT2B7 or expressing no UGT were used as HEK293 cells stably overexpressing UGT2B7 or expressing no UGT were used as the positive control (+) and negative control (−), respectively. Protein samples (−30 µg) were resolved by SDS-PAGE and transferred onto nitrocellulose membrane. Blots were probed with an anti-UGT2B7 antibody, followed by horseradish peroxidase-conjugated donkey anti-rabbit antibody and immunodetection using chemiluminescent substrate. (A) Images captured using ImageQuant LAS 4000 luminescent image analyzer. (B) Quantification of the relative UGT2B7 protein levels in samples of the three genotypes (II/II, I/II, and II/II); after background subtraction, data were presented as the means of each genotype from four samples (arbitrary units), error bars representing 1 S.D. P < 0.05 was considered statistically significant (Student’s t test).

In summary, this study showed that all the 23 prevalent polymorphisms in the 4-kb UGT2B7 promoter (promoter 1) are linked together to define two major haplotypes, I and II. We demonstrated that the higher promoter activity of haplotype II over haplotype I was mediated by the SNP (−900A>G), which is contained within a putative AP-1 binding site. Finally, we propose that the polymorphic AluY element linked to haplotype I and in close proximity to the transcription start site might control UGT2B7 promoter activity through methylation.

**Authorship Contributions**

**Participated in research design:** Hu, Mackenzie.

**Conducted experiments:** Hu, Lu.

**Performed data analysis:** Hu, Mackenzie.

**Wrote or contributed to the writing of the manuscript:** Hu, Mackenzie, Meech, McKinnon.

**References**


Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, and Tomilin N, et al. (1999) Association of AluY element with lower morphine glucuronide/morphine AUC (area under the concentration-time curve) ratios compared with the ratios of AA genotypes in patients with sickle cell disease treated with morphine (Darbari et al., 2008; Joly et al., 2012).

Nakamura et al. (2008) showed binding of transcription factors to the UGT2B7 promoter region covering SNP -900A>G by electrophoretic mobility shift assays; however, the identity of the transcription factor(s) involved was not defined in this study. As annotated in the NCBI GRCh37/hg19 Assembly, recent chromatin immunoprecipitation sequencing data from the Encyclopedia of DNA Elements, (ENCODE), Project has shown that Fra-2/JunD dimers bind to the region covering the SNP (−900A>G) in HepG2 cells and that c-Fos/c-Jun dimers bind to the same region in human umbilical vein endothelial cells. We performed chromatin immunoprecipitation assays to assess whether the c-Fos/c-Jun dimers could also bind to the same promoter region in HepG2 cells (which are homozygous for haplotype II carrying the consensus AP-1 site). Our results showed no occupancy of c-Fos and c-Jun at this promoter region (data not shown). Collectively, these data suggest that AP-1 proteins bind the UGT2B7 promoter consensus AP-1 site as Fra-2/JunD dimers in HepG2 cells.

In addition to the SNP -900A>G, the INDEL of the AluY element represents another prevalent potentially functional polymorphism at the UGT2B7 promoter. Alu elements can modulate gene expression when inserted into a target gene in either the same or opposite orientation with respect to the direction of target gene transcription (Ebihara et al., 2002; Le Goff et al., 2003; Cheng et al., 2012). It has been reported that Alu elements harbor putative response elements for various transcription factors, including thyroid receptor, estrogen receptor, RAR, hepatocyte nuclear factor 4α, nuclear factor-xb, and the tumor suppressor p53 protein (Vansant and Reynolds, 1995; Piedrafita et al., 1996; Babich et al., 1999; Polak and Domany 2006; Antonaki et al., 2011; Bolotin et al., 2011; Cui et al., 2011). For example, the evolutionarily recent Alu subfamilies contain three highly conserved hormone-responsive element (HRE) half-sites [(A/G)(G/T)(T/C)(G/A)(G)] (Vansant and Reynolds, 1995). These three sites are arranged as direct repeats separated by 2 bp (DR2), a structure that permits binding of the retinoic acid receptors (RARs) (Vansant and Reynolds, 1995). Recombinant RAR proteins bound these HRE sites in electrophoretic mobility shift assays and promoted the activity of reporter constructs containing these HRE sites (Vansant and Reynolds, 1995; Li et al., 2002). As shown in Fig. 4, three DR2 HRE sites are present in the UGT2B7 AluY element. Whether these sites could be bound by RARs remains to be investigated. The UGT2B7 AluY element is inserted between nucleotides −2067 and −2390 in the UGT2B7 promoter. An AluSx element located in a similar position (−2153/−2414) in the CETP (cholesterol ester transfer protein) promoter has been shown to negatively regulate promoter activity (Le Goff et al., 2003).

Alu elements account for 30% of the methylation sites in the human genome (Batzer and Deininger 2002; Deininger et al., 2003) and are considered methylation centers in the human genome (Turker and Bestor, 1997). The UGT2B7 AluY element has 23 CpG sites; however, its methylation status in cell lines and tissues remains unexplored. In addition to the AluY element, the UGT2B7 gene contains another two Alu elements: a 285-bp AluJr element (−4580/−4295) in the distal promoter and a 251-bp AluSx3 element (+7486/+7747) in intron 3. Only 0.5% of Alu elements in the human genome are polymorphic (Roy-Engel et al., 2001). The AluY element is the only polymorphic Alu element in the UGT2B7 gene and is exclusively linked to haplotype I.
McPhee DJ, Carroll D, Faur CC, Gavaghan DJ, Hand CW, and Moore RA (1990) Oral mor-
Ménard V, Roberge J, Harvey M, Lévesque E, and Guillemette C (2011) Transcriptional diversity at the UGT2B7 locus is dictated by extensive pre-mRNA splicing mechanisms that give rise to more than 200 mRNA splice variants. Pharmacogenet Genomics 21:631–641.

Address correspondence to: Professor Peter I. Mackenzie, Department of Clinical Pharmacology, Flinders University School of Medicine, Flinders Medical Centre, Bedford Park SA 5042, Australia. E-mail: peter.mackenzie@flinders.edu.au.
Polymorphisms and Haplotypes of the UDP-Glucuronosyltransferase 2B7 Gene Promoter.

Supplemental Methods

Determination of the specificity of anti-UGT2B7 antibody. Full length cDNAs of UGT2B4, 2B7, 2B10, 2B15, 2B17, or 2B28 were cloned into the pEF-IRESpuro6 expression vector, which contains a puromycin resistance gene. The resultant expression vectors were transfected into human embryonic kidney (HEK293T) cells, and cell lines stably expressing the respective UGTs were obtained following treatment using puromycin (2 µg/ml). UGT expression was confirmed by RT-PCR and a pan-UGT antibody. Whole cell lysates were prepared from these cell lines using RIPA buffer (150 mM NaCl, 1.0% IGEPAL® CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, pH 8.0). Protein concentration was determined using BioRad Protein Assay Reagents. To verify the specificity of our previously reported anti-UGT2B7 antibody (Kerdpin et al. 2009), 30 µg of each of the above-mentioned whole cell lysates were resolved on SDS-polyacrylamide gels (10%) and transferred to nitrocellulose membrane. Blots were probed with this anti-UGT2B7 antibody, followed by horseradish peroxidase-conjugated donkey anti-rabbit antibody. SuperSignal® West Pico Chemiluminescent substrate (Thermo Scientific) was used for immunodetection, and images were obtained using an ImageQuant LAS 4000 luminescent image analyser (GE Healthcare Life Sciences).

Supplemental Fig. 1. Specific binding of anti-UGT2B7 antibody to UGT2B7 proteins. Aliquots (30 µg) of whole cell lysates prepared from HEK293T cell lines stably expressing UGT2B4, 2B7, 2B10, 2B15, 2B17, or 2B28 was separated on SDS-polyacrylamide gels (10%) and transferred to nitrocellulose membrane. Blots were probed with our custom-made anti-UGT2B7 antibody as described in detail in Supplementary Methods.