

DMD #56630

## Title Page

### 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 SA 5042, Australia.

DMD #56630

## Running Title Page

**Running Title:** UDP glucuronosyltransferase 2B7 promoter haplotypes

### Address correspondence to

Professor Peter I Mackenzie, Department of Clinical Pharmacology, Flinders University  
School of Medicine, Flinders Medical Centre, Bedford Park SA 5042, Australia

Phone: 0061 8 82045394

Fax: 0061 8 82045114

Email: [peter.mackenzie@flinders.edu.au](mailto:peter.mackenzie@flinders.edu.au)

**Text pages:** 23

**Tables:** 1

**Figures:** 6 and 1 supplementary figure

**References:** 55

**Abstract:** 232 words

**Introduction:** 594 words

**Discussion:** 1471 words

### Abbreviations:

AP-1, activating protein 1; bp, base pair; ChIP-seq, Chromatin Immunoprecipitation Sequencing; ENCODE, encyclopedia of DNA elements; Fra-1, fos-related antigen 1; Fra-2, fos-related antigen 2; HUVECs, human umbilical vein endothelial cells; kb, kilo base pair; LD: linkage disequilibrium; INDEL, insertion and deletion; MAF, minor allele frequency; MPA, mycophenolic acid; M3G, morphine-3-glucuronide; M6G, morphine-6-glucuronide; MT, mutated construct; NCBI, National Center of Biotechnology Information; RAR, retinoic acid receptor; SNP, single nucleotide polymorphisms; UGT, UDP-glucuronosyltransferase.

DMD #56630

## Abstract

Identification of functional polymorphisms in the UDP-glucuronosyltransferase 2B7 (*UGT2B7*) gene predicting inter-patient 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 herein 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 -2,600/-4,000 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. Due to 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 and II. The addition of the minor allele of a rare polymorphism and allele exchanges between Haplotypes I and II generated sub-haplotypes 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 AP-1 site (TGAGTCAA) as occurred in Haplotype II to a mutated AP-1 site (TGAGTCG) 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 glucuronide 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, non-steroidal anti-inflammatory drugs, epirubicin, morphine, and mycophenolic acid (MPA) (Radomska-Pandya, Little and Czernik 2001).

Morphine is primarily metabolized by UGT2B7 in the liver to form morphine-3-glucuronide (M3G, the major inactive metabolite) and morphine-6-glucuronide (M6G, a more potent analgesic compound compared to 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 glucuronide-to-morphine plasma concentration ratios in cancer patients on chronic morphine therapy (Holthe et al. 2002, Holthe et al. 2003, Klepstad, Kaasa and Borchgrevink 2000, McQuay et al. 1990, Tiseo et al. 1995, Fladvad et al. 2013). This inter-patient difference in morphine glucuronide-to-morphine ratios is attributable to genetic and non-genetic factors. Non-genetic factors may include variability in daily morphine dosage, medical conditions, co-medications, cross-membrane transport, hepatic blood flow, volume of distribution, protein binding, and renal elimination (Holthe et al. 2002, Holthe et al. 2003, Fladvad et al. 2013, Court 2010).

Studies have shown up to 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 (Innocenti et al. 2008, Fisher et al. 2000). Thus the identification of polymorphisms that

DMD #56630

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). The majority of studies show that this SNP has little impact on substrate selectivity and enzyme activity (Bhasker et al. 2000, Court 2010, Thibaudau et al. 2006, Holthe et al. 2002, Coffman et al. 1997, Court et al. 2003, Sawyer et al. 2003, Lin et al. 2005, Wiener et al. 2004, Coffman et al. 1998, Innocenti et al. 2008). 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. 2003, Tian et al. 2012, Duguay et al. 2004, Nakamura et al. 2008). Two SNPs were identified in the promoter that negatively (-138A>G, rs7668258) (Duguay et al. 2004) or positively (-900A>G, rs7438135) (Duguay et al. 2004, Nakamura et al. 2008) regulate *UGT2B7* promoter activity; however, the underlying molecular mechanisms remained undefined. Another SNP (IVSI +985A>G, rs62298861) in intron 1 with a MAF of 0.18 has been shown to be significantly associated with hepatic UGT2B7 mRNA levels, and M3G and M6G formation (Innocenti et al. 2008). As deposited in the NCBI dbSNP database, nine other prevalent polymorphisms with a similar MAF of around 0.35 are present in the distal *UGT2B7* promoter between nucleotides -2,600 and -4,000. Whether these polymorphisms are in LD with the polymorphisms at the proximal promoter and whether they impact on *UGT2B7* promoter activity is unknown.

## 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 Haplotype 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. 2013, Menard et al. 2011). 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 1 (termed 1a, 1b, 1c, and 1d), exon 1b is the closest variant exon 1 to promoter 1 and resides 6871 bp upstream from promoter 1. The *UGT2B7* gene fragment cloned and analysed 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 just 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 (2B7-4026F or 2B7-4926F) and a

DMD #56630

common reverse primer (2B7-1R). The resultant PCR amplicons were subsequently cloned into the MluI and XhoI sites of the pGL3-basic vector (Promega).

According to the manufacturer's instructions (Roch), 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 4 clones containing a 4 kb *UGT2B7* promoter. The cloned promoter fragments in these 20 clones were sequenced in full bi-directionally 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-453), the prostate cancer cell lines (LNCaP, VCaP, PC3, and DU145), and the kidney HEK293 cell line were purchased from American Type Culture Collection. The HepG2 and VCaP cell lines were maintained in Dulbecco's Modified Eagle 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 prior to 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  $\mu$ l of Lipofectamine<sup>®</sup>2000 (Invitrogen). 24 h post-transfection, the transfection medium was replaced by fresh medium and cells were cultured for a further 24 h. 48 h post-transfection, cells were harvested to quantify Firefly and Renilla luciferase activities using the Dual-Luciferase Reporter Assay system (Promega, Madison, WI) as described previously (Hu and Mackenzie 2010). Firefly (*Photinus pyralis*) luciferase activity was normalized against the

DMD #56630

Renilla (*Renilla reniformis*) activity. The relative activity of the *UGT2B7* promoter was presented as -fold change compared to 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 TGAGTCA whereas Haplotype I has a mutated AP-1 site (TGAGTCCG, 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 given 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 9 human cell lines using primers 2B7-2579F and 2B7-1862R (Table 1). Briefly, PCR was performed in a 50 ul reaction consisting 200 ng of each genomic DNA sample, 1 unit of Phusion hot start high-fidelity DNA polymerase (Thermo Scientific), dNTPs (200 nM each), and the forward and reverse primers (200 nM each). 10 ul aliquots of each amplicon were resolved on 1% 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 (Kerdpin et al. 2009). We herein performed western blotting assays to further verify its specificity. As shown in (Supplemental Figure 1), this antibody displayed high binding specificity towards 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). Following 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 donkey anti-rabbit antibody as previously reported (Kerdpin et al. 2009). 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). Band intensity of the immunoblots and background subtraction was carried out using imaging software, Multi Gauge Ver3.0 (FUJIFILM, Japan).

**Statistical Analysis.** Analysis of Variance (ANOVA) 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 less than 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 (assessed 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, 17 polymorphisms within the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> 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 (rs61641343, 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 -2067 and -2390 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 minor allele frequency (MAF) of > 1%. Out 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 3 novel single nucleotide substitutions (Fig. 1A, 1B, 1C). The known polymorphisms included 22 of the above-mentioned 23 prevalent SNPs/INDELs, rs11730928C>T at -4324 (MAF: 0.14), rs62296941A>G at -3127 (MAF: 0.12), rs73823857T>C at -1681 (MAF: 0.3), and rs28458345G>T at -2821 (MAF: 0.3). In addition, three novel single nucleotide substitutions, -788C>T, -2036G>A, and -2655C>T, were seen in a single individual clone (Fig. 1A, 1B,

DMD #56630

1C). 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 5 clones, thus this INDEL represents the 23<sup>rd</sup> 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 8 unique sub-haplotypes that belonged to two major haplotypes: Haplotype I and haplotype II (Fig. 2). Linkage of all of the minor alleles or all of the major alleles of the 23 prevalent SNPs/INDELs observed in the present study gave rise to Haplotype I (3 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 (Duguay et al. 2004). Haplotypes I and II are named in other reports as \*2 and \*1, respectively (Nakamura et al. 2008, Saeki et al. 2004). Of note, Haplotype I matches the *UGT2B7* promoter reference sequence (NC\_000004.11). As shown in Fig. 2, our findings agreed with the previously reported linkage of the most proximal six (Duguay et al. 2004, Holthe et al. 2003) or fourteen (Nakamura et al. 2008) prevalent polymorphisms at the *UGT2B7* promoter. As mentioned above, the *UGT2B7* promoter region between nucleotides -4,000 and -5,000 contains 17 polymorphisms, one of which (rs11730928C>T) at -4342 has a MAF of 0.14 and the MAFs for the others are less than 0.03. We observed the minor allele of SNP rs11730928C>T (“T”) at -4342 in one clone, and the major alleles of the remaining 16 polymorphisms in all of the 16 cloned 5 kb *UGT2B7* promoters (data not shown). Therefore, our results were not informative for haplotype analysis for this promoter region. Collectively,

DMD #56630

our data extends the previously reported LD at the *UGT2B7* promoter in the 5' upstream direction from the INDEL of an Alu element (rs71955301) (Nakamura et al. 2008) at -2067 to the INDEL of TAACAGAC (rs61641343) at -3967. In addition to Haplotypes I and II, we observed allele exchanges between Haplotype I and Haplotype II generating sub-haplotypes including allele swapping at three polymorphic sites in one clone (Haplotype I-1) or at eight polymorphic sites in another clone (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, the replacement of the major allele with the minor allele at a rare polymorphic site gave rise to an additional four sub-haplotypes (I-2, II-2, II-3, and II-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 model cell line, namely HepG2. As shown in Fig. 3, all six haplotypes displayed strong promoter activity as compared to 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 found to be similar among the three Haplotypes II (II, II-1, and II-2) ( $p > 0.05$  in all cases). These results indicate that the nucleotide substitutions that differentiate the sub-haplotypes had no impact on promoter activity. However, the activities of Haplotypes II were about 2-fold higher compared to 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 (-900A>G) (Duguay et al. 2004, Nakamura et al. 2008). However, the *UGT2B7* promoter analysed in these studies covered only the proximal six (Duguay et al. 2004) or fourteen (Nakamura et al. 2008) prevalent polymorphisms. To assess whether the SNP (-900A>G) also contributed to the 2-fold difference in promoter activity between the 5 kb Haplotype I and Haplotype II promoters, we

DMD #56630

compared promoter-luciferase constructs carrying either the wild type Haplotype II -900A allele or the variant Haplotype II harbouring 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 (-900A>G) is functional in the 5 kb *UGT2B7* promoter.

***An Alu Y 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, Li and Liang 2013, Price, Eskin and Pevzner 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, Pai and Li 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 classical Alu element, including (1) the highly conserved left and right arms, a central polyA tract, an internal RNA polymerase III promoter consensus sequence (Boxes A and B), two AluI 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

DMD #56630

polymorphism at the *UGT2B7* promoter were obtained for 9 human cell lines and 26 liver tissues. Of the 9 cell lines, 2 cell lines (MCF7 and MDA-MB-453) were found to be homozygous for the AluY insertion, 3 cell lines (HepG2, HEK293, and PC3) homozygous for the AluY deletion, and 4 cell lines (ZR75, LNCaP, VCaP, and Du145) heterozygous for the AluY insertion and deletion (Fig. 5A). These results indicate that this INDEL is prevalent in commonly used human cell lines. Out of the 26 liver tissues, 6 were found to be homozygous for the AluY insertion, 7 homozygous for the AluY deletion, and 13 heterozygous for the AluY insertion and deletion (Fig. 5B, 5C, and 5D). 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. 5B, 5C, and 5D, 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>T in this panel of liver tissues.

***Haplotype II Associated with Higher UGT2B7 Protein Levels Than Haplotype I in Human Liver Microsomes.*** The liver microsomes analysed included 4 samples homozygous for haplotype I, 4 samples homozygous for Haplotype II, and 4 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

DMD #56630

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 was consistent with our observed higher promoter activity of Haplotype II over Haplotype I as shown in Fig. 3.

DMD #56630

## Discussion

Sawyer et al first reported that -161C>T at the proximal promoter was in complete LD with 802C>T at exon 2 in the *UGT2B7* gene (Sawyer et al. 2003). This LD has been confirmed and extended in both directions spanning the gene from the 2 kb proximal promoter to the terminal exon (Duguay et al. 2004, Hirota et al. 2003, Holthe et al. 2003, Innocenti et al. 2008, Nakamura et al. 2008, Saeki et al. 2004, Saito et al. 2006, Sawyer et al. 2003, Tian et al. 2012, Hwang et al. 2010). This LD has been reported in different ethnic groups, including Caucasian (Duguay et al. 2004, Innocenti et al. 2008, Holthe et al. 2003), Chinese (Tian et al. 2012), Korean (Hwang et al. 2010), and Japanese populations (Nakamura et al. 2008, Saito et al. 2006, Saeki et al. 2004, Hirota et al. 2003). The present study further extended this LD in the 5' upstream direction spanning the -2000 to -4000 bp distal promoter region. All of the 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. 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.



DMD #56630

We showed in this study that the addition of the minor allele of a rare polymorphism to Haplotype I or II generated sub-haplotypes of I or II. We expect that a huge number of such sub-haplotypes exist in the population given that there are nearly 100 polymorphisms with a MAF of less than 0.03 within the 5 kb *UGT2B7* promoter. We showed in the present study that four rare polymorphisms (-1681T>C, 2036G>A, -2821G>T, and -3127A>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 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 (Duguay et al. 2004). 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 to 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, Hansen and Mackenzie 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 activator protein (AP-1) transcription factors comprise Jun (c-Jun, JunB, and JunD) and Fos (c-Fos, FosB, Fra-1, and Fra-2) sub-families. AP-1 proteins bind as dimers to their cognate regulatory elements [5'-TGA(G/C)TCA-3'] at promoters and enhancers to modulate gene transcription (Hess, Angel and Schorpp-Kistner 2004). The SNP (-900A>G) at the *UGT2B7* promoter is contained within a putative AP-1 binding site [TGAGTC(A/G)]. Haplotype II has a consensus AP-1 site (TGAGTCAA), whereas Haplotype I has a mutated AP-1 site (TGAGTCG). We showed that the activity of the Haplotype II 5 kb promoter was

DMD #56630

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-mycophenolic acid-glucuronide was significantly higher in AA- and GA-human liver microsomes, compared with GG-human liver microsomes (Djebli et al. 2007), and that GG and GA genotypes were associated with lower morphine glucuronide/morphine AUC (area under the concentration-time curve) ratios compared to the ratios of AA genotypes in Sickle cell disease patients treated with morphine (Darbari et al. 2008, Joly et al. 2012).

Nakamura et al 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 (Nakamura et al. 2008). As annotated in the NCBI GRCh37/hg19 Assembly, recent Chromatin Immunoprecipitation Sequencing (ChIP-seq) data from the 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 (HUVECs). 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

DMD #56630

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 (Cheng et al. 2012, Ebihara et al. 2002, Le Goff et al. 2003). It has been reported that Alu elements harbour putative response elements for various transcription factors, including thyroid receptor, estrogen receptor, retinoic acid receptor (RAR), hepatocyte nuclear factor 4 alpha, nuclear factor- $\kappa$ B, and the tumour suppressor p53 protein (Piedrafita et al. 1996, Vansant and Reynolds 1995, Antonaki et al. 2011, Bolotin et al. 2011, Polak and Domany 2006, Babich et al. 1999, Cui, Sirotin and Zhurkin 2011). For example, the evolutionarily recent Alu sub-families contain three highly conserved hormone-responsive element (HRE) half-sites [(A/G)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* (cholesteryl 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 as methylation centres 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.

DMD #56630

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.

In summary, this study showed that all of 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.

DMD #56630

### **Authorship Contributions**

*Participated in research design:* Hu and Mackenzie

*Conducted experiments:* Hu and Lu

*Performed data analysis:* Hu and Mackenzie

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

McKinnon

DMD #56630

## References

- Ahmed, M., W. Li & P. Liang (2013) Identification of three new Alu Yb subfamilies by source tracking of recently integrated Alu Yb elements. *Mob DNA*, 4, 25.
- Antonaki, A., C. Demetriades, A. Polyzos, A. Banos, G. Vatsellas, M. D. Lavigne, E. Apostolou, E. Mantouvalou, D. Papadopoulou, G. Mosialos & D. Thanos (2011) Genomic analysis reveals a novel nuclear factor-kappaB (NF-kappaB)-binding site in Alu-repetitive elements. *J Biol Chem*, 286, 38768-38782.
- Babich, V., N. Aksenov, V. Alexeenko, S. L. Oei, G. Buchlow & N. Tomilin (1999) Association of some potential hormone response elements in human genes with the Alu family repeats. *Gene*, 239, 341-349.
- Batzer, M. A. & P. L. Deininger (2002) Alu repeats and human genomic diversity. *Nat Rev Genet*, 3, 370-379.
- Bhasker, C. R., W. McKinnon, A. Stone, A. C. Lo, T. Kubota, T. Ishizaki & J. O. Miners (2000) Genetic polymorphism of UDP-glucuronosyltransferase 2B7 (UGT2B7) at amino acid 268: ethnic diversity of alleles and potential clinical significance. *Pharmacogenetics*, 10, 679-685.
- Bolotin, E., K. Chellappa, W. Hwang-Verslues, J. M. Schnabl, C. Yang & F. M. Sladek (2011) Nuclear receptor HNF4alpha binding sequences are widespread in Alu repeats. *BMC Genomics*, 12, 560.
- Cheng, L. C., T. W. Pai & L. A. Li (2012) Regulation of human CYP11B1 and CYP11B2 promoters by transposable elements and conserved cis elements. *Steroids*, 77, 100-109.
- Coffman, B. L., C. D. King, G. R. Rios & T. R. Tephly (1998) The glucuronidation of opioids, other xenobiotics, and androgens by human UGT2B7Y(268) and UGT2B7H(268). *Drug Metab Dispos*, 26, 73-77.

DMD #56630

- Coffman, B. L., G. R. Rios, C. D. King & T. R. Tephly (1997) Human UGT2B7 catalyzes morphine glucuronidation. *Drug Metab Dispos*, 25, 1-4.
- Congiu, M., M. L. Mashford, J. L. Slavin & P. V. Desmond (2002) UDP glucuronosyltransferase mRNA levels in human liver disease. *Drug Metab Dispos*, 30, 129-134.
- Court, M. H. (2010) Interindividual variability in hepatic drug glucuronidation: studies into the role of age, sex, enzyme inducers, and genetic polymorphism using the human liver bank as a model system. *Drug Metab Rev*, 42, 209-224.
- Court, M. H., S. Krishnaswamy, Q. Hao, S. X. Duan, C. J. Patten, L. L. Von Moltke & D. J. Greenblatt (2003) Evaluation of 3'-azido-3'-deoxythymidine, morphine, and codeine as probe substrates for UDP-glucuronosyltransferase 2B7 (UGT2B7) in human liver microsomes: specificity and influence of the UGT2B7\*2 polymorphism. *Drug Metab Dispos*, 31, 1125-1133.
- Cui, F., M. V. Sirotin & V. B. Zhurkin (2011) Impact of Alu repeats on the evolution of human p53 binding sites. *Biol Direct*, 6, 2.
- Darbari, D. S., R. H. van Schaik, E. V. Capparelli, S. Rana, R. McCarter & J. van den Anker (2008) UGT2B7 promoter variant -840G>A contributes to the variability in hepatic clearance of morphine in patients with sickle cell disease. *Am J Hematol*, 83, 200-202.
- Deininger, P. L., J. V. Moran, M. A. Batzer & H. H. Kazazian, Jr. (2003) Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev*, 13, 651-658.
- Djebli, N., N. Picard, J. P. Rerolle, Y. Le Meur & P. Marquet (2007) Influence of the UGT2B7 promoter region and exon 2 polymorphisms and comedications on Acyl-MPAG production in vitro and in adult renal transplant patients. *Pharmacogenet Genomics*, 17, 321-330.

DMD #56630

- Duguay, Y., C. Baar, F. Skorpen & C. Guillemette (2004) A novel functional polymorphism in the uridine diphosphate-glucuronosyltransferase 2B7 promoter with significant impact on promoter activity. *Clin Pharmacol Ther*, 75, 223-233.
- Ebihara, M., H. Ohba, S. I. Ohno & T. Yoshikawa (2002) Genomic organization and promoter analysis of the human nicotinic acetylcholine receptor alpha6 subunit (CHNRA6) gene: Alu and other elements direct transcriptional repression. *Gene*, 298, 101-108.
- Fisher, M. B., M. Vandenbranden, K. Findlay, B. Burchell, K. E. Thummel, S. D. Hall & S. A. Wrighton (2000) Tissue distribution and interindividual variation in human UDP-glucuronosyltransferase activity: relationship between UGT1A1 promoter genotype and variability in a liver bank. *Pharmacogenetics*, 10, 727-739.
- Fladvad, T., P. Klepstad, M. Langaas, O. Dale, S. Kaasa, A. Caraceni & F. Skorpen (2013) Variability in UDP-glucuronosyltransferase genes and morphine metabolism: observations from a cross-sectional multicenter study in advanced cancer patients with pain. *Pharmacogenet Genomics*, 23, 117-126.
- Hess, J., P. Angel & M. Schorpp-Kistner (2004) AP-1 subunits: quarrel and harmony among siblings. *J Cell Sci*, 117, 5965-5973.
- Hirota, T., I. Ieiri, H. Takane, H. Sano, K. Kawamoto, H. Aono, A. Yamasaki, H. Takeuchi, M. Masada, E. Shimizu, S. Higuchi & K. Otsubo (2003) Sequence variability and candidate gene analysis in two cancer patients with complex clinical outcomes during morphine therapy. *Drug Metab Dispos*, 31, 677-680.
- Holthe, M., P. Klepstad, K. Zahlse, P. C. Borchgrevink, L. Hagen, O. Dale, S. Kaasa, H. E. Krokan & F. Skorpen (2002) Morphine glucuronide-to-morphine plasma ratios are unaffected by the UGT2B7 H268Y and UGT1A1\*28 polymorphisms in cancer patients on chronic morphine therapy. *Eur J Clin Pharmacol*, 58, 353-356.



DMD #56630

Holthe, M., T. N. Rakvag, P. Klepstad, J. R. Idle, S. Kaasa, H. E. Krokan & F. Skorpen (2003)

Sequence variations in the UDP-glucuronosyltransferase 2B7 (UGT2B7) gene: identification of 10 novel single nucleotide polymorphisms (SNPs) and analysis of their relevance to morphine glucuronidation in cancer patients. *Pharmacogenomics J*, 3, 17-26.

Hu, D. G. & P. I. Mackenzie (2009) Estrogen receptor alpha, fos-related antigen-2, and c-Jun coordinately regulate human UDP glucuronosyltransferase 2B15 and 2B17 expression in response to 17beta-estradiol in MCF-7 cells. *Mol Pharmacol*, 76, 425-439.

Hu, D. G. & P. I. Mackenzie (2010) Forkhead box protein A1 regulates UDP-glucuronosyltransferase 2B15 gene transcription in LNCaP prostate cancer cells. *Drug Metab Dispos*, 38, 2105-2109.

Hwang, M. S., S. J. Lee, H. E. Jeong, S. Lee, M. A. Yoo & J. G. Shin (2010) Genetic variations in UDP-glucuronosyltransferase 2B7 gene (UGT2B7) in a Korean population. *Drug Metab Pharmacokinet*, 25, 398-402.

Innocenti, F., W. Liu, D. Fackenthal, J. Ramirez, P. Chen, X. Ye, X. Wu, W. Zhang, S.

Mirkov, S. Das, E. Cook, Jr. & M. J. Ratain (2008) Single nucleotide polymorphism discovery and functional assessment of variation in the UDP-glucuronosyltransferase 2B7 gene. *Pharmacogenet Genomics*, 18, 683-697.

Ishii, Y., A. J. Hansen & P. I. Mackenzie (2000) Octamer transcription factor-1 enhances hepatic nuclear factor-1alpha-mediated activation of the human UDP glucuronosyltransferase 2B7 promoter. *Mol Pharmacol*, 57, 940-947.

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

DMD #56630

- Joly, P., M. C. Gagnieu, C. Bardel, A. Francina, C. Pondarre & C. Martin (2012) Genotypic screening of the main opiate-related polymorphisms in a cohort of 139 sickle cell disease patients. *Am J Hematol*, 87, 534-536.
- Kaer, K. & M. Speek (2013) Retroelements in human disease. *Gene*, 518, 231-241.
- Kerdpin, O., P. I. Mackenzie, K. Bowalgaha, M. Finel & J. O. Miners (2009) Influence of N-terminal domain histidine and proline residues on the substrate selectivities of human UDP-glucuronosyltransferase 1A1, 1A6, 1A9, 2B7, and 2B10. *Drug Metab Dispos*, 37, 1948-1955.
- Klepstad, P., S. Kaasa & P. C. Borchgrevink (2000) Start of oral morphine to cancer patients: effective serum morphine concentrations and contribution from morphine-6-glucuronide to the analgesia produced by morphine. *Eur J Clin Pharmacol*, 55, 713-719.
- Le Goff, W., M. Guerin, M. J. Chapman & J. Thillet (2003) A CYP7A promoter binding factor site and Alu repeat in the distal promoter region are implicated in regulation of human CETP gene expression. *J Lipid Res*, 44, 902-910.
- Li, Y., C. Bolten, B. G. Bhat, J. Woodring-Dietz, S. Li, S. K. Prayaga, C. Xia & D. S. Lala (2002) Induction of human liver X receptor alpha gene expression via an autoregulatory loop mechanism. *Mol Endocrinol*, 16, 506-514.
- Lin, G. F., W. C. Guo, J. G. Chen, Y. Q. Qin, K. Golka, C. Q. Xiang, Q. W. Ma, D. R. Lu & J. H. Shen (2005) An association of UDP-glucuronosyltransferase 2B7 C802T (His268Tyr) polymorphism with bladder cancer in benzidine-exposed workers in China. *Toxicol Sci*, 85, 502-506.
- McQuay, H. J., D. Carroll, C. C. Faura, D. J. Gavaghan, C. W. Hand & R. A. Moore (1990) Oral morphine in cancer pain: influences on morphine and metabolite concentration. *Clin Pharmacol Ther*, 48, 236-244.

DMD #56630

Menard, V., O. Eap, J. Roberge, M. Harvey, E. Levesque & C. Guillemette (2011)

Transcriptional diversity at the UGT2B7 locus is dictated by extensive pre-mRNA splicing mechanisms that give rise to multiple mRNA splice variants. *Pharmacogenet Genomics*, 21, 631-641.

Menard, V., E. Levesque, S. Chen, O. Eap, M. S. Joy, L. Ekstrom, A. Rane & C. Guillemette

(2013) Expression of UGT2B7 is driven by two mutually exclusive promoters and alternative splicing in human tissues: changes from prenatal life to adulthood and in kidney cancer. *Pharmacogenet Genomics*, 23, 684-696.

Nakamura, A., M. Nakajima, E. Higashi, H. Yamanaka & T. Yokoi (2008) Genetic

polymorphisms in the 5'-flanking region of human UDP-glucuronosyltransferase 2B7 affect the Nrf2-dependent transcriptional regulation. *Pharmacogenet Genomics*, 18, 709-720.

Piedrafita, F. J., R. B. Molander, G. Vansant, E. A. Orlova, M. Pfahl & W. F. Reynolds (1996)

An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. *J Biol Chem*, 271, 14412-14420.

Polak, P. & E. Domany (2006) Alu elements contain many binding sites for transcription

factors and may play a role in regulation of developmental processes. *BMC Genomics*, 7, 133.

Price, A. L., E. Eskin & P. A. Pevzner (2004) Whole-genome analysis of Alu repeat elements

reveals complex evolutionary history. *Genome Res*, 14, 2245-2252.

Radomska-Pandya, A., J. M. Little & P. J. Czernik (2001) Human UDP-

glucuronosyltransferase 2B7. *Curr Drug Metab*, 2, 283-298.

Roy-Engel, A. M., M. L. Carroll, E. Vogel, R. K. Garber, S. V. Nguyen, A. H. Salem, M. A.

Batzer & P. L. Deininger (2001) Alu insertion polymorphisms for the study of human genomic diversity. *Genetics*, 159, 279-290.

DMD #56630

- Saeki, M., Y. Saito, H. Jinno, T. Tanaka-Kagawa, A. Ohno, S. Ozawa, K. Ueno, S. Kamakura, N. Kamatani, K. Komamura, M. Kitakaze & J. Sawada (2004) Single nucleotide polymorphisms and haplotype frequencies of UGT2B4 and UGT2B7 in a Japanese population. *Drug Metab Dispos*, 32, 1048-1054.
- Saito, K., H. Moriya, T. Sawaguchi, T. Hayakawa, S. Nakahara, A. Goto, Y. Arimura, K. Imai, N. Kurosawa, E. Owada & A. Miyamoto (2006) Haplotype analysis of UDP-glucuronocyltransferase 2B7 gene (UGT2B7) polymorphisms in healthy Japanese subjects. *Clin Biochem*, 39, 303-308.
- Sawyer, M. B., F. Innocenti, S. Das, C. Cheng, J. Ramirez, F. H. Pantle-Fisher, C. Wright, J. Badner, D. Pei, J. M. Boyett, E. Cook, Jr. & M. J. Ratain (2003) A pharmacogenetic study of uridine diphosphate-glucuronosyltransferase 2B7 in patients receiving morphine. *Clin Pharmacol Ther*, 73, 566-574.
- Thibaudeau, J., J. Lepine, J. Tojcic, Y. Duguay, G. Pelletier, M. Plante, J. Brisson, B. Tetu, S. Jacob, L. Perusse, A. Belanger & C. Guillemette (2006) Characterization of common UGT1A8, UGT1A9, and UGT2B7 variants with different capacities to inactivate mutagenic 4-hydroxylated metabolites of estradiol and estrone. *Cancer Res*, 66, 125-133.
- Tian, J. N., I. K. Ho, H. H. Tsou, C. P. Fang, C. F. Hsiao, C. H. Chen, H. K. Tan, L. Lin, C. S. Wu, L. W. Su, C. L. Huang, Y. H. Yang, M. L. Liu, Y. T. Chen, S. C. Liu, Y. T. Hsu, H. W. Kuo, C. T. Liu, Y. T. Yang, A. Chen, Y. H. Shih & Y. L. Liu (2012) UGT2B7 genetic polymorphisms are associated with the withdrawal symptoms in methadone maintenance patients. *Pharmacogenomics*, 13, 879-88.
- Tiseo, P. J., H. T. Thaler, J. Lapin, C. E. Inturrisi, R. K. Portenoy & K. M. Foley (1995) Morphine-6-glucuronide concentrations and opioid-related side effects: a survey in cancer patients. *Pain*, 61, 47-54.

DMD #56630

- Turker, M. S. & T. H. Bestor (1997) Formation of methylation patterns in the mammalian genome. *Mutat Res*, 386, 119-30.
- Vansant, G. & W. F. Reynolds (1995) The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc Natl Acad Sci U S A*, 92, 8229-33.
- Wiener, D., D. R. Doerge, J. L. Fang, P. Upadhyaya & P. Lazarus (2004) Characterization of N-glucuronidation of the lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) in human liver: importance of UDP-glucuronosyltransferase 1A4. *Drug Metab Dispos*, 32, 72-9.

DMD #56630

## **Footnotes**

This work was supported by funding from the National Health and Medical Research Council of Australia [ID 1020931].

DMD #56630

## Legends for Figures

**Fig. 1.** Chromatogram sequencing results of three novel SNPs and the INDEL (rs61641343) at the *UGT2B7* promoter. Three novel single nucleotide polymorphisms (SNPs) were found in a single cloned 5 kb *UGT2B7* promoter, including (A): -788C>T; (B): -2036G>A, and (C): -2655C>T. Shown are both the major alleles (top panel) and minor alleles (bottom panel) of these three SNPs, which are boxed. (D) Shown are four INDELs (rs35014193, rs61641343, rs33937977, and rs113605769) which are present in the *UGT2B7* promoter between -3965 and -3970 (bottom chromatogram). Of these four INDELs, only the minor allele of the INDEL rs61641343, an insertion of 5'TAACAGAC3' between nucleotides -3966 and -3967, was found in this study (top chromatogram). Nucleotide positions are numbered relative to the translation start site with the "A" of the initiation ATG codon positioned as + 1.

**Fig. 2.** Polymorphisms are in tight linkage disequilibrium (LD) at the 4 kb *UGT2B7* promoter. The *UGT2B7* promoter region (4 kb or 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 8 haplotypes that were found in 20 cloned promoter constructs. Both minor and major alleles (bold) of these polymorphisms are indicated by arrows. 23 prevalent alleles with a similar global minor allele frequency of around 0.35 are indicated by \*. Due to LD in the promoter, 8 sub-haplotypes 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 construct are highlighted in bold. Previously reported haplotypes of the *UGT2B7* proximal promoter are shown for comparison. a, indicating the INDEL of a 324 bp Alu element between nucleotides

DMD #56630

-2067 and -2390. b, indicating the INDEL of TAACAGAC between nucleotides -3966 and -3967. Ref1, Holthe et al, 2003. Ref2, Duguay et al, 2004. Ref3, Nakamura et al., 2008.

**Fig. 3.** Higher *UGT2B7* promoter activity of Haplotype II over Haplotype I is mediated by the SNP -900A>G in HepG2 cells. HepG2 cells were transfected in 96-well plates with *UGT2B7* promoter-luciferase reporter constructs as described in the Methods. 48 h post-transfection, cells were harvested to quantify the Firefly and Renilla luciferase activities using the Dual-Luciferase Reporter Assay system (Promega). 8 reporter constructs were transfected, including the empty pGL3-basic vector, 6 pGL3 constructs carrying the 5 kb *UGT2B7* promoter with different haplotypes (I, I-1, I-2, II, II-2, and II-3), and the construct carrying the 5 kb *UGT2B7* haplotype II promoter with an experimentally introduced A-to-G substitution at -900 (MT). After normalization to the internal control pRL-null vector, the *UGT2B7* promoter activity was presented as -fold change relative to that of the empty pGL3-basic vector. Data shown are the average fold changes from three independent experiments (in triplicate for each transfection), error bars representing 1 S.D. ¥:  $p > 0.05$ ; \*  $p < 0.05$ .

**Fig. 4.** An INDEL of an AluY element in the *UGT2B7* promoter. An AluY element is inserted in an antisense orientation in the promoter carrying Haplotype I between nucleotides -2066 and -2390. Shown is the alignment between the consensus AluY sequence and the 324 bp *UGT2B7* AluY element sequence (antisense sequence). This AluY element has all of the features (boxed or underlined) of a classic Alu element, including (1) the highly conserved left and right arms, a central polyA tract, an internal RNA polymerase III promoter consensus sequence (Boxes A and B), two AluI restriction sites, a polyA tail, a target site duplication signal sequence of 5'AGAAACC3' flanking the element. This AluY element differs from the AluY consensus sequence by only two single nucleotide substitutions in the right arm and an



DMD #56630

insertion of nine “A” in the central polyA tract. In addition, this AluY element has three hormone-response element (HREs) half-sites that are arranged as direct repeats separated by 2 bp (DR2). Of note, HRE3 is contained within the internal RNA polymerase III promoter consensus sequence (Box B).

**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 9 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 ul of each of the amplicons was analysed on agarose gels as shown in (A) for cell lines and (B, C, 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 was PCR products from sample heterozygous for the Alu Y polymorphism. M was DNA marker (100 bp ladder).

**Fig. 6.** Haplotype II is associated with higher *UGT2B7* protein levels than haplotype I in human liver microsomes. Human liver microsomes analysed included 4 samples homozygous for haplotype I, 4 samples homozygous for Haplotype II, and 4 samples heterozygous for Haplotype I and II. Whole cell lysates from HEK293 cells stably overexpressing *UGT2B7* or

DMD #56630

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 analyser. (B): Quantification of the relative UGT2B7 protein levels in samples of the three genotypes (I/I, I/II, and II/II); after background subtraction, data were presented as the means of each genotype from 4 samples (arbitrary units), error bars representing 1 S.D. A p value less than 0.05 was considered statistically significant (Student's t-test).

DMD #56630

## Tables

Table 1

Primers used in this study for cloning, sequencing, genotyping and mutagenesis (5'-3')

cloning	
2B7-1046F	ATAG <u>ACGCGT</u> GACCTGAAGGAATACCTCAA
2B7-2076F	ATGT <u>ACGCGT</u> GCCGGTTTCTAAGGCATATA
2B7-3026F	ATAG <u>ACGCGT</u> TCAGCCTCCCACCTTTGCAAT
2B7-4026F	ATCAACGCGTAGTGATGCTTCAGGAACTT
2B7-4925F	TACT <u>ACGCGT</u> GAGCCCTGTAACTCTTAAT
2B7-1R	ATAT <u>CTCGAG</u> CCTGGTGCAATGCAAT
Sequencing	
2B7-917R	GTTATTTGGCTGTTGTCCTA
2B7-1947R	TCATTAGCACTTTTCTAGCAA
2B7-2907R	GCCACTCAGGCACGTGTAGA
2B7-3867R	CTGGATACAAAAGGTAAAGA
Genotyping	
2B7-2579F	GCATATTACCTACTTCCGGTC
2B7-1862R	CCACCACCAAAATGAGTTAGCAA
Mutagenesis	
2B7-900 A to G MTF	GTGAGGAAGTGAGTCGGAGAACAAGCTAACC
2B7-900 A to G MTR	GGTTAGCTTGTTCTCCGACTCACTTCCTCAC

Primer nucleotides are numbered relative to the UGT2B7 translation start site with the "A" of the initiator codon (ATG) positioned as +1. Incorporated restriction sites of both *MulI* (ACGCGT) and *XhoI* (CTCGAG) are underlined.

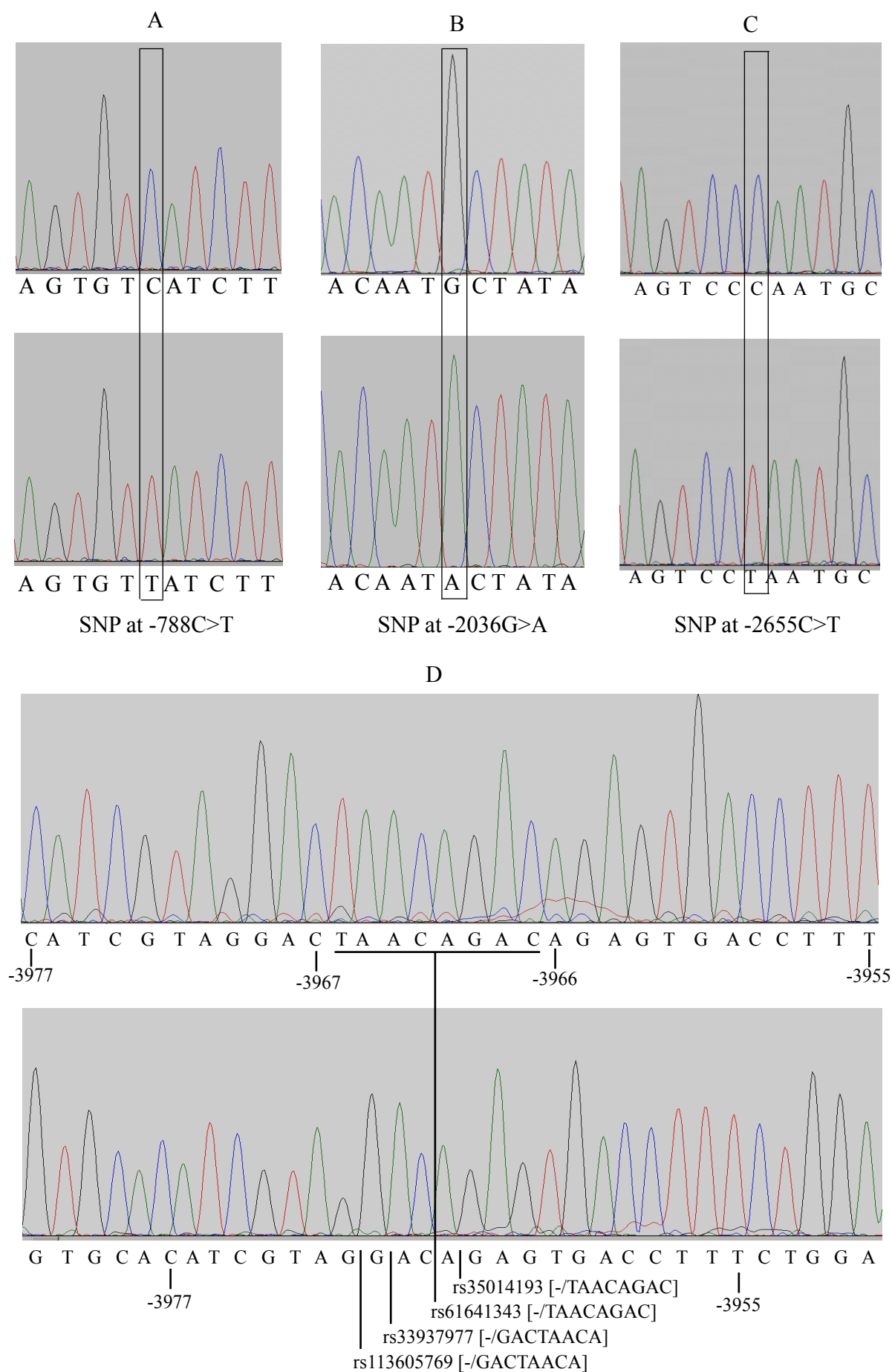


Figure 2

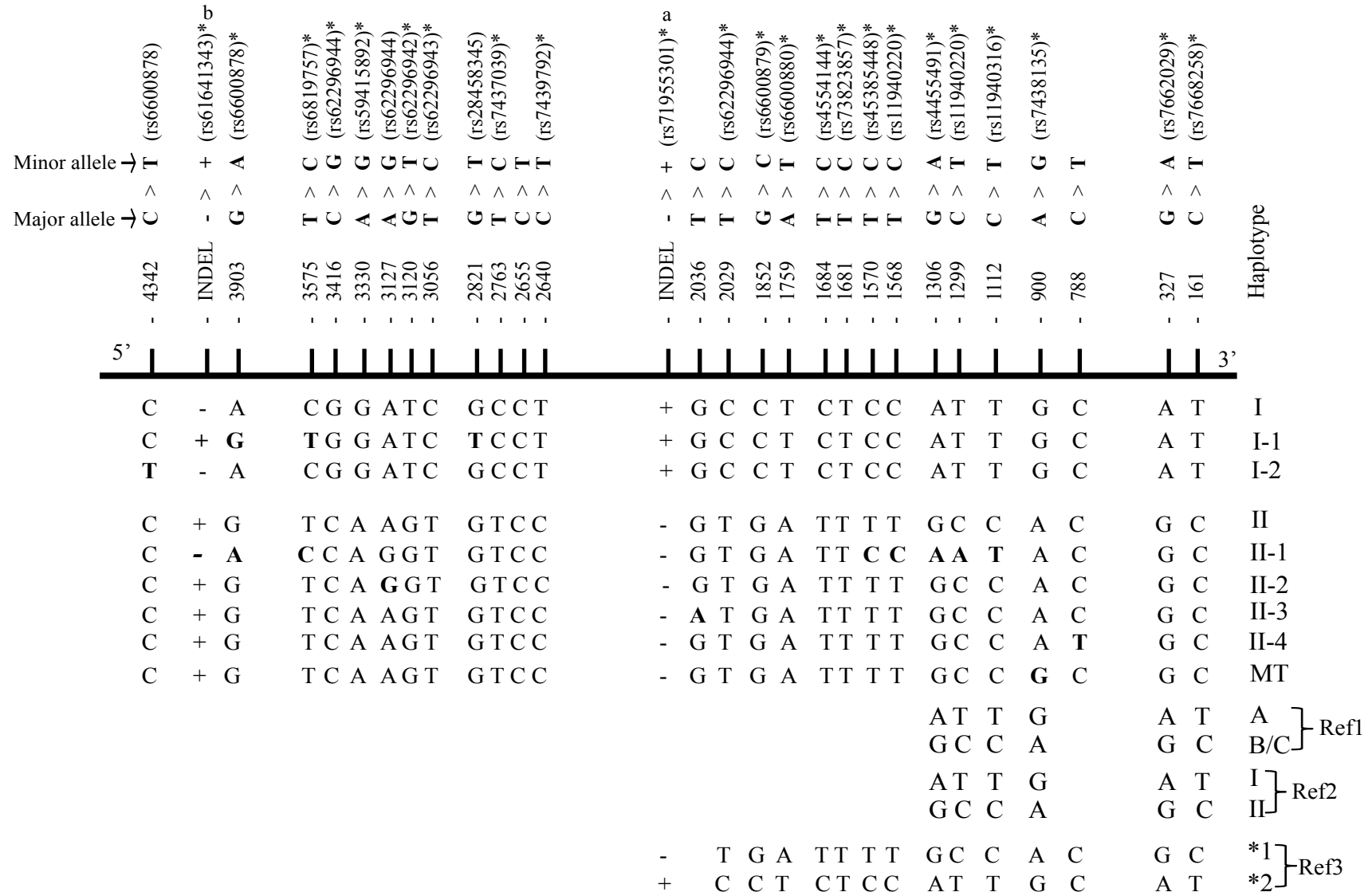


Figure 3

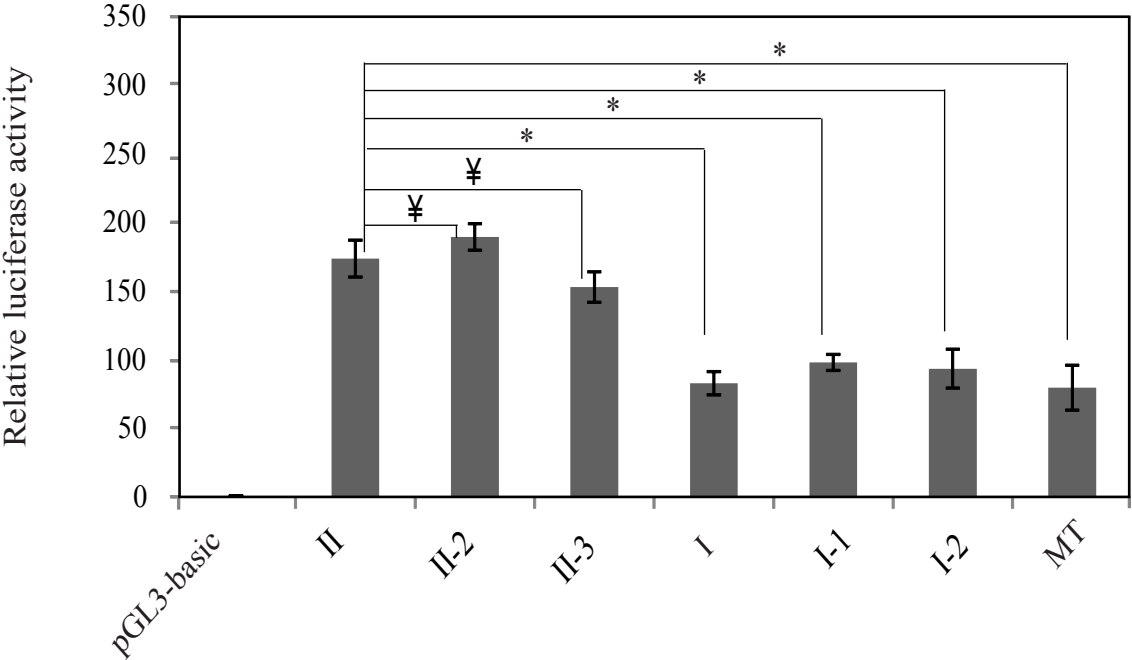


Figure 4

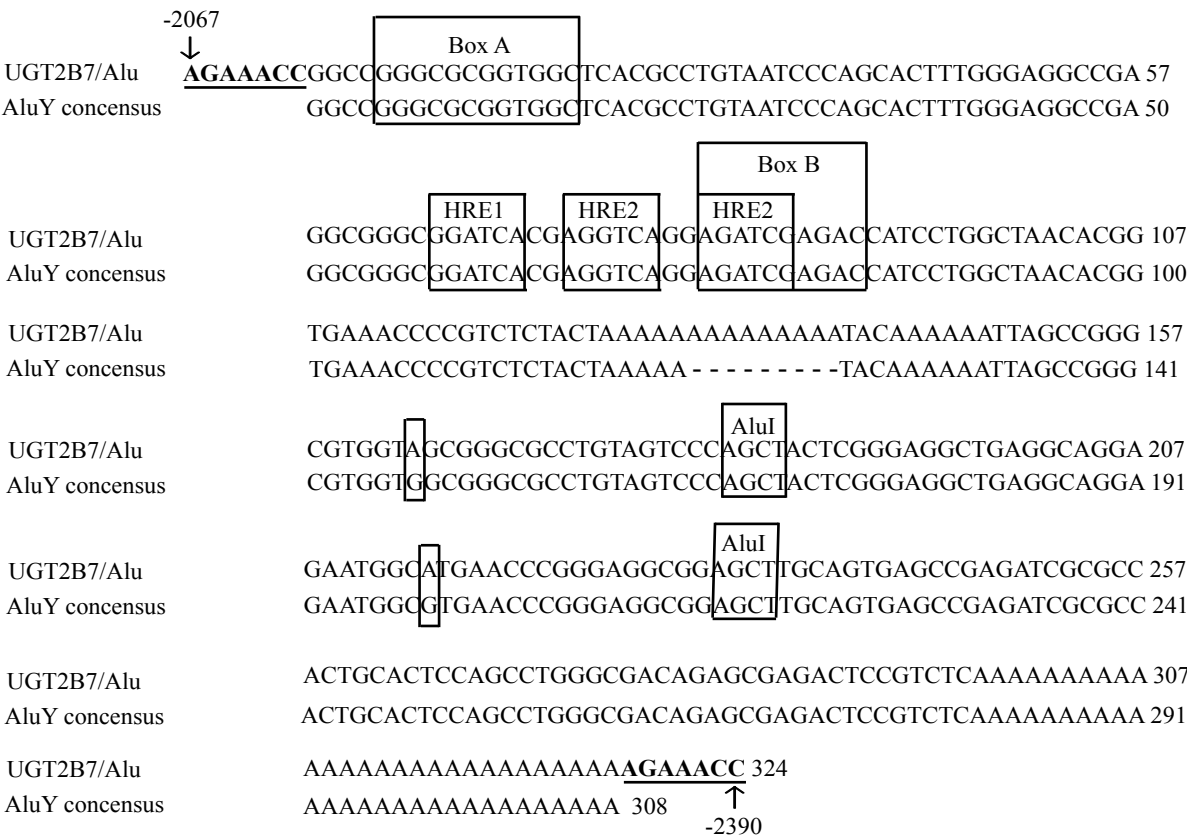


Figure 5

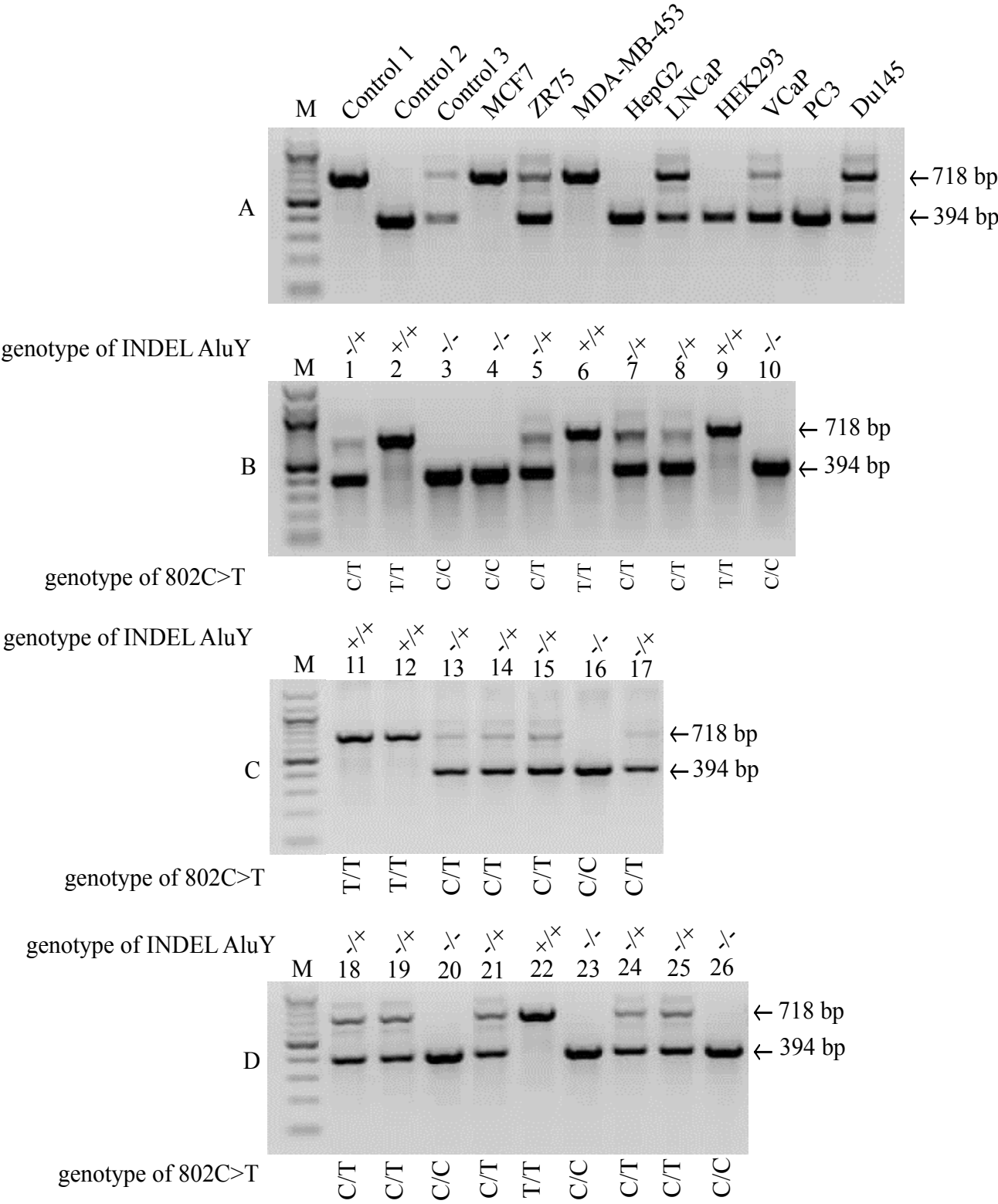
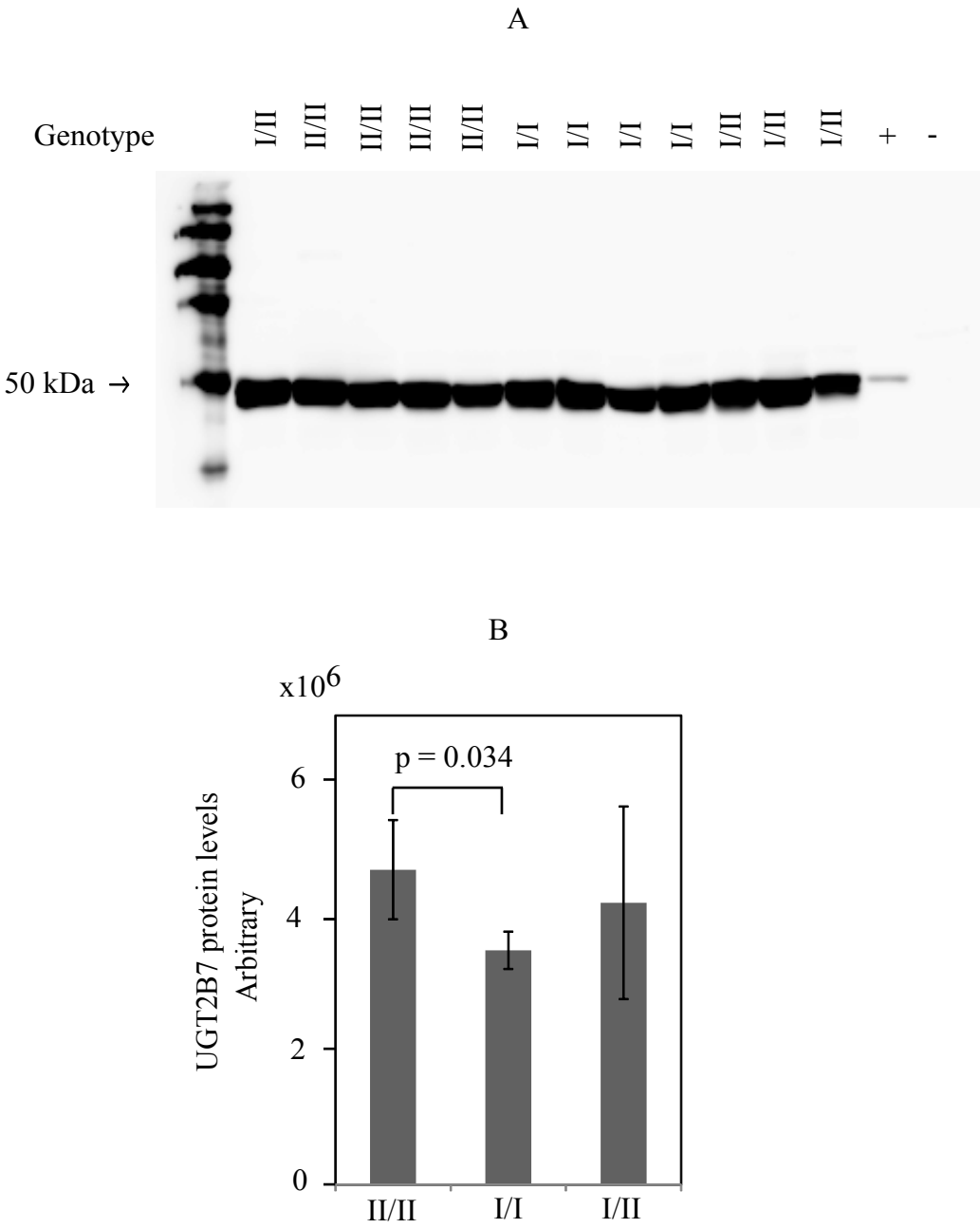


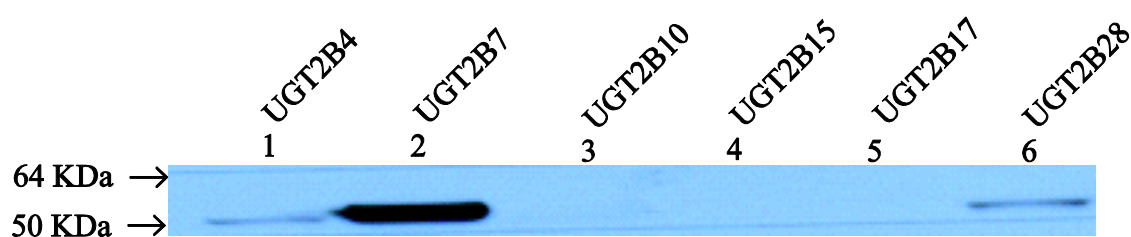


Figure 6



## 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  $\mu$ g/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<sup>®</sup>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  $\mu$ g 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<sup>®</sup>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  $\mu$ g) 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 processed with our custom-made anti-UGT2B7 antibody as described in detail in Supplementary Methods.