Published in final edited form as:

Pharmacogenet Genomics. 2008 August ; 18(8): 683–697. doi:10.1097/FPC.0b013e3283037fe4.

SNP discovery and functional assessment of variation in the *UDP-glucuronosyltransferase 2B7 (UGT2B7)* gene

Federico Innocenti^{1,3,4}, Wanqing Liu¹, Donna Fackenthal², Jacqueline Ramírez¹, PeiXian Chen², Xin Ye¹, Xiaolin Wu², Wei Zhang¹, Snezana Mirkov¹, Soma Das^{2,3}, Edwin Cook Jr.⁵, and Mark J. Ratain^{1,3,4}

¹Department of Medicine The University of Chicago, Chicago, IL 60637, USA

²Department of Genetics, The University of Chicago, Chicago, IL 60637, USA

³Committee on Clinical Pharmacology and Pharmacogenomics, The University of Chicago, Chicago, IL 60637, USA

⁴Cancer Research Center, The University of Chicago, Chicago, IL 60637, USA

⁵Department of Psychiatry, The University of Illinois at Chicago, Chicago, IL 60612, USA

Abstract

Objective—UGT2B7 plays a central role in the liver-mediated biotransformation of endogenous and exogenous compounds. The genetic basis of interindividual variability in UGT2B7 function is unknown. This study aimed to discover novel gene variants of functional significance.

Methods—Caucasian human livers (n=54) were used. *UGT2B7* was resequenced in 12 samples (6 highest and 6 lowest for the formation of morphine-3-glucuronide, M3G). Haplotype-tagging single nucleotide polymorphisms (tSNPs) were genotyped in the entire sample set. Samples were phenotyped for mRNA expression.

Results—10 tSNPs were identified and their haplotypes were inferred. Haplotype 4 (-45597G;-6682_-6683A;372A;IVS1+9_IVS1+10A;IVS1+829T;IVS1+985G;IVS1+999C;IVS1+1250G;80 1T;IVS4+185C) (frequency of 0.12) was associated with an increase in enzyme activity and gene expression. The 1/4 and 4/6 diplotypes had higher M3G formation compared to 1/1 (p<0.05) and 2/3 (p<0.01) diplotypes. Diplotypes containing haplotype 4 resulted in a significant 45% average increase in the formation of M3G compared to diplotypes without haplotype 4 (p=0.002). There was also an association between haplotype 4 and increased mRNA expression. IVS1+985A>G, 735A>G and 1062C>T are the putative functional variants of haplotype 4. We also identified two mRNA splicing variants (*UGT2B7_v2* and *UGT2B7_v3*) splicing out exon 1, 4, 5 and 6 but sharing exons 2 and 3 with the involvement of additional 5' exons. *UGT2B7_v2* was detected in all livers tested, but *UGT2B7_v3* was present at much lower levels compared to *UGT2B7_v2*. The *UGT2B7* reference sequence mRNA is now named *UGT2B7_v1*.

Conclusions—*UGT2B7* haplotype 4 is functional and its effects on the biotransformation of *UGT2B7* substrates should be tested in controlled clinical trials. Biochemical studies should investigate the functional role of the newly discovered mRNA splicing variants.

Correspondence to: Federico Innocenti.

Contact information for correspondence including requests for reprints: Federico Innocenti, M.D., Ph.D., The University of Chicago, 5841 South Maryland Ave., MC 2115, Chicago, Illinois 60637, U.S.A. Phone: (773) 834-2452. Fax: (773) 702-9698. E-mail: E-mail: finnocen@medicine.bsd.uchicago.edu.

Keywords

pharmacogenetics; haplotype; SNP; splicing variant; UGT2B7

Introduction

UDP-glucuronosyltransferases (UGTs) are major phase II conjugative metabolic enzymes. The liver is the main organ of drug biotransformation, and the reaction with UDPGA as the cosubstrate results in the production of glucuronidated metabolites that are more water-soluble than the parent compound and readily eliminated through the biliary and renal systems [1]. Similar to the 18 members of the human UGT superfamily of enzymes, the C-terminal domain of UGT2B7 is the site of binding for UDPGA, and the N-terminal domain is the site of binding for the various substrates. The elucidation of the biochemical properties of this enzyme is an area of active research, and the crystal structure of the UDPGA-binding domain has been recently obtained [2].

UGT2B7 is highly expressed in the human liver, with lower levels of expression in other extrahepatic organs [3]. UGT2B7 is able to glucuronidate various steroid hormones (androsterone, epitestosterone) and fatty acids. UGT2B7 therefore plays a key role in the detoxification of cholestatic bile acids and may prevent the formation of proximal carcinogens such as quinone estrogens [3]. In addition, UGT2B7 is also able to conjugate major classes of drugs such as analgesics, carboxylic nonsteroidal anti-inflammatory drugs (ketoprofen), anticarcinogens (all-trans retinoic acid) and anticancer drugs (epirubicin) [3].

The prototypical UGT2B7 substrate is morphine. In humans, morphine is primarily metabolized through glucuronidation via UGT2B7, and the generated metabolites have different pharmacological properties. The formation of the inactive morphine-3-glucuronide (M3G) is an important elimination pathway of morphine. The analgesic potency of morphine-6-glucuronide (M6G) is higher than that of morphine [4]. Interpatient variability in the formation of M3G and M6G exists, but its genetic component remains to be established [5].

The variant in the *UGT2B7* gene that has been studied the most is 802T>C in exon 2 (H268Y) (*UGT2B7**2) [6]. In established HEK293 cell lines expressing the two isoforms, many substrates were tested, with the result that the substrate specificity was not changed for the vast majority of the substrates [6-8]. In another study using a small series of human liver microsomes, morphine glucuronidation rates were not significantly affected by the 802T>C variant [9]. More recently, we conducted a resequencing study in acute pain patients treated with morphine, and a -161T>C variant (relative to the ATG) in the *UGT2B7* promoter [in complete linkage disequilibrium (LD) with 802T>C] was found to be associated with increased M6G/morphine plasma ratios [10]. However, recent studies of morphine in cancer patients did not detect any significant association between these two *UGT2B7* variants and either morphine glucuronidation [11] or analgesic response [12]. Additional clinical studies investigating other *UGT2B7* variants did not detect an effect on morphine disposition and response [11,13,14]. For example, in vitro luciferase assays of promoter variants show significant effects on gene transcription but no effect on morphine metabolism in patients [11,13]. A -840G>A variant has been recently shown to affect hepatic clearance of morphine in 20 sickle cell disease patients [15] but no effect was detected in a larger study of morphine treated patients [11].

From the studies conducted so far, there is no compelling evidence for the existence of functional *UGT2B7* variants. The genetic variability that influences the activity of the UGT2B7 enzyme remains to be established [5]. In addition, mRNA splicing variation can add another

level of complexity to the biological functions of *UGT* genes in the liver, as previously demonstrated [16,17]. Due to the central role of *UGT2B7* in the liver-mediated biotransformation of endogenous and exogenous compounds, we performed extensive resequencing of the *UGT2B7* gene in human livers aiming to 1) discover novel gene variants, 2) characterize their haplotypic structure, 3) define the effect of genotypes and haplotypes on gene expression and enzyme activity, and 4) identify mRNA splicing variants. Our ultimate goal is to provide the knowledge base for testing the clinical impact of *UGT2B7* variation in humans.

Methods

Morphine glucuronidation in human liver microsomes

Caucasian human livers (n=54) were processed through Dr. Mary Relling's laboratory at St. Jude Children's Research Hospital and were provided by the Liver Tissue Procurement and Distribution System and by the Cooperative Human Tissue Network, funded through NIH contract #N01-DK-9-2310. Samples were collected with approval of institutional review boards. Morphine and epirubicin glucuronidation were measured in liver microsomes following the methods previously published [8]. Briefly, incubations with morphine lasted 20 min and contained 1.4 mM morphine, 5 mM MgCl₂, 2 mg/ml microsomal protein, 0.1 M Tris-HCl buffer (pH 7.4) and 5 mM UDPGA. The reactions were stopped with cold acetonitrile and spiked with internal standard (42 nmol 10,11-dihydrocarbamazepine). Formation of M3G and M6G was measured by HPLC [8]. The concentration of morphine used (1.4 mM) resembles the *K_m* of M3G (2 mM) and M6G (1.9 mM) formation [8]. Incubations with epirubicin contained 600 μM epirubicin (similar to the *K_m* for the formation of epirubicin-G=568 μM [8]), 10 mM MgCl₂, 3 mg/ml microsomal protein, 0.1 M Tris-HCl buffer (pH 7.4) and 5 mM UDPGA. Reactions were stopped after 4 h with cold methanol followed by addition of internal standard (1 nmol daunorubicin), and analyzed by HPLC [8].

Activities determined at a single substrate concentration are not ideal for the phenotypic assessment of enzyme activities. The *K_m* and *V_{max}* for the formation of the glucuronides in each liver was not assessed due to the low amount of microsomes available from each liver, and this should be regarded as a limitation of this study. However, the phenotypic assessment of the *UGT2B7* enzyme served the purpose of a rapid screening to identify the extremes of the phenotypic distribution and select those samples for resequencing, increasing the likelihood of identifying functional variants.

Phenotypic glucuronidating activity in each liver was expressed as the ratio between the peak heights of glucuronides to the internal standard. Although the amount of formed metabolite has not been assessed by using a standard curve, the chromatographic peak heights ratios of M3G, M6G, and epirubicin-G vs. their internal standards reported were in the same range of those reported in our previous publication [8] (data not shown).

Sequencing of M3G phenotypic extremes

Twelve samples in the phenotypic extremes of the frequency distribution of the M3G formation (6 with the highest and 6 the lowest formation) were chosen for resequencing. All six coding exons with flanking intron sequence, 3'UTR, promoter/5'UTR and the 5'-upstream region containing alternatively spliced exons were resequenced (Fig. 1). Sequence coverage includes approximately 2.4 kb of the 5' flanking region which contains 0.4 kb of exon 1A, 0.25 kb of exon 1B and flanking regions, 2.2 kb of exon 1 and flanking regions (including 1.2 kb upstream and 0.25 kb downstream flanking sequence), 0.78 kb of exon 2 and flanking intron sequence (including 0.48 kb upstream and 0.15 kb downstream flanking intron sequence), 0.4 kb of exon 3 and flanking intron sequence, 1.4 kb of exons 4 and 5 which includes all of intron 4 (0.84

kb), and approximately 0.9 kb of exon 6 including 0.1-0.2 kb of flanking intron and 3' flanking regions. Primers used for PCR amplification are listed in Table 1.

PCRs were set up in a 25 μ l volume containing 2.5 mM $MgCl_2$, 200 μ M each dNTP, 500 nM forward and reverse primers, 0.5 units AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA) and 100 ng of DNA. Reactions were cycled at 94°C for 45 s, and annealing was performed at appropriate temperatures for 30 s and 72°C for 1 min for 35 cycles. This was followed by double-stranded sequencing. In preparation for sequencing, PCR products were purified using the QIAquick® PCR Purification Kit (Qiagen, Inc., Valencia, CA, USA). Purified products were eluted in 30 μ l of elution buffer. DNA cycle sequencing reactions were carried out in 10 μ l reactions using 4 μ l of purified PCR product, 400 nM forward or reverse primer, BigDye® Terminator Version 3.0 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and HalfBD Dye Terminator Sequencing Reagent (Sigma-Aldrich Co., St. Louis, MO, USA) per manufacturer's instructions. The sequencing primers are shown in Table 1. In many instances the same primers used for PCR were also used for sequencing. Cycle sequencing was performed using standard conditions and reactions were run on a 3100 DNA Sequencer (Applied Biosystems, Foster City, CA, USA). Sequence analysis was performed using Polyphred, Version 4.05.

Linkage disequilibrium (LD) analysis, selection of haplotype-tagging single nucleotide polymorphisms (tSNPs), and estimation of haplotypes and diplotypes

LD analysis of SNPs (single nucleotide polymorphisms) among 12 Caucasians was performed by using the VG2 program (<http://pga.gs.washington.edu/VG2.html>). For selection of the tSNPs, SNPs were clustered using LDSelect [18] integrated in the VG2 program. The parameters for the selection of tSNPs were $r^2 \geq 0.9$ and a minor allele frequency of ≥ 0.05 . Haplotypes and diplotypes composed of 10 tSNPs among 54 samples were estimated by using the Phase 2.0 program. A phase probability cut off of ≥ 0.9 was used for diplotype assignment. Diplotypes assigned from 4 out of the 54 individuals were ambiguous and not used in the association with phenotypic data.

Genotyping of *UGT2B7* tSNPs and -161T>C and 802T>C

Ten tSNPs (eight SNPs and two indels) were genotyped in 54 Caucasian liver DNA samples (Table 2, in bold). All SNPs were genotyped by single base extension (SBE) except for 644T>A which was genotyped by primer extension. PCR and extension primers are listed in Table 1. PCRs were set up in a 15 μ l volume containing 1.5 mM $MgCl_2$, 100 μ M each dNTP, 125 nM forward and reverse primers, 0.375 units AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA) and 25 ng of DNA. Reactions were denatured initially at 95°C for 15 min then cycled at 95°C for 15 s, annealing was performed at appropriate temperatures for 15 s and 72°C for 45 s for 40 cycles.

For the SBE reactions, PCR amplified products were treated with shrimp alkaline phosphatase and exonuclease I to remove excess dNTPs and primers. Six μ l of the purified PCR product was combined with 1 μ M of extension primer, 250 μ M each ddNTP and 1.25 U Thermo Sequenase™ (Amersham Pharmacia Biotech, Piscataway, NJ, USA) in a 10 μ l volume, and denatured initially at 96°C for two min then cycled at 96°C for 30 s, 55°C for 30 s, and 60°C for 30 s for 60 cycles. Samples were denatured at 96°C for 4 min prior to separation of the extension products. Extension products were separated by denaturing high performance liquid chromatography using the WAVE 3500HT Nucleic Acid Fragment Analysis System (Transgenomic, Inc., Omaha, NE, USA). Separation on the column was performed at 70°C with 76% 0.1 M triethylammonium acetate and 24% acetonitrile/0.1 M triethylammonium acetate for 2.7 min.

SNPs 137T>C and IVS1+985A>G were genotyped in a duplex reaction for both the PCR and SBE reactions. In the PCR, the primers were adjusted to 200 nM and 65 nM respectively to obtain approximately equal amounts of both amplicons. One μ M of each extension primer was added to the duplex SBE reaction.

SNP 644T>A was genotyped by primer extension as opposed to SBE as resolution of the T and A alleles was not optimal by SBE. For the primer extension reaction, 10 μ l of purified PCR product was combined with 1 μ M of extension primer, 250 μ M ddATP, 125 μ M dCTP, dGTP, dTTP and 2.5 U Thermo Sequenase™ (Amersham Pharmacia Biotech, Piscataway, NJ, USA) in a 20 μ l volume, and denatured initially at 96°C for two min then cycled at 96°C for 30 s, 52°C for 30 s, and 60°C for 30 s for 60 cycles.

The indel variants -6682_-6683insGCAAAT and IVS1+9_IVS1+10insT were genotyped by PCR and sizing by running of amplified products on an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA, USA) at the University of Chicago DNA Sequencing and Genotyping Facility and scored using GeneMapper software (Applied Biosystems, Foster City, CA, USA).

We also genotyped all the samples for the -161T>C and 802T>C variants using methods established in one of our previous studies [10].

PCR analysis of *UGT2B7* splicing variants *UGT2B7_v2* and *UGT2B7_v3* in human livers

Analysis of the presence of mRNA expression of the *UGT2B7* splicing variants in the 54 human livers (with the exception of 2 samples due to mRNA degradation) was performed by PCR using primers specific for each splicing variant. Reference sequence *UGT2B7* mRNA (now classified as *UGT2B7_v1*) and β -actin mRNA were also amplified. Sequences for forward primers used to amplify splicing variants were 5'-

GGCCCAGATTCCACAGAAGGAA-3' (junction of exon 1B and exon 2 for *UGT2B7_v2*) and 5'- GTCCAGACATCTAAAATTTGGAAGA-3' (junction of exon 1A and exon 2 for *UGT2B7_v3*) (Fig. 1A). The sequence of the reverse primers for *UGT2B7_v2* and *UGT2B7_v3* was the same of that of *UGT2B7_v1* (see below).

PCRs were performed using HotStar Taq DNA Polymerase (Qiagen, Valencia, CA, USA), with thermal profile for both splicing variants as follows: initial activation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 10 s, annealing at 62°C for 10 s and extension at 72°C for 30 s and a final extension cycle at 72°C for 7 min. The PCR conditions for the *UGT2B7_v1* and β -actin were the same of those used for the splicing variants except for annealing at 58°C and number of cycles (29 and 28, respectively). PCR products were separated on a 1.5% agarose gel and visualized by ethidium bromide staining using a ChemiDoc (Bio-Rad Laboratories, Hercules, CA, USA). PCR products of the two splicing variants were also cut out from the gel, purified and confirmed by direct sequencing.

To discover the full sequences of *UGT2B7_v2* and *UGT2B7_v3*, 5'- and 3'-RACE (Rapid Amplification of cDNA End) were conducted. Liver cDNA library was constructed by using the SMART™ RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA) according to the manufacturer's instruction. Nested PCR was used to generate the 5'- and 3'-ends of the cDNA. Briefly, gene specific primers 5'-

GGAATCTGGGCCAAGTCTGAAGC-3' (5'-RACE) and 5'-

GAAGCAGGTAGCAGGCCCTCGCAG-3' (3'-RACE) were used to perform the first RACE PCR with the universal primer AP1 (Clontech), respectively; Nested PCR was performed using gene specific primers 5'-CTCAGATAATGTAGTGGGTCTTCCAAAT-3' (5'RACE) 5'-GTCCAGATCTCTGGCAAAGATGG-3' (3'-RACE) and universal nested primer AP2

(Clontech), respectively. After amplification, PCR products were separated in 2% agarose gel; PCR bands were cut out from the gel, purified and directly sequenced from both ends.

mRNA levels of *UGT2B7* reference sequence (*UGT2B7_v1*) and *UGT2B7_v2* in human livers

The mRNA levels of *UGT2B7_v1* and *UGT2B7_v2* were measured in the same set of 54 livers (with the exception of 2 samples due to mRNA degradation) by two-step real time PCR using IQ™ SYBR Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA) on a Mx3000P system (Stratagene, Cedar Creek, TX, USA). β -actin mRNA expression was also measured. We were not able to quantify the expression level of *UGT2B7_v3* with accuracy.

cDNA was synthesized in a 20 μ l reaction volume using the iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA, USA) and 2 μ g of total RNA. Real time PCRs for *UGT2B7_v1* and β -actin were then performed with the thermal cycles as follows: 40 cycles with denaturing at 95°C for 30 s, annealing at 55°C for 1 min and extension at 72 °C for 30 s, after preheating at 95°C for 10 min. Real time PCR for *UGT2B7_v2* was performed as follows: 45 cycles with denaturing at 95°C for 15 s, annealing and extension at 68°C for 1 min, after preheating at 95 °C for 10 min. The oligonucleotide sequences of the primers were: 5'-CAGCTTCTCTCTGGCTACACTT-3' (forward primer) for *UGT2B7_v1*, 5'-GGCCCAGATTCCACAGAAGGAA (forward primer) for *UGT2B7_v2* and the same reverse primer 5'-CAGGAGTTTCGAATAAGCCATAC-3' for both. The β -actin gene was amplified using 5'-ACGTGGACATCCGCAAGAC-3' (forward primer) and 5'-CAAGAAAGGGTGTAACGCAACTA-3' (reverse primer). Reactions were performed in triplicate in 96-well plates including standard curves. Gene expression was normalized to the expression level of β -actin.

Statistics

The frequency distribution of M3G, M6G, epirubicin glucuronide (epirubicin-G) was normal according to the Kolmogorov-Smirnov test ($p>0.05$). The mRNA levels of *UGT2B7_v1* (ratio to β -actin) were log transformed, as they did not pass the normality test ($p<0.05$); after log transformation, they were normally distributed ($p>0.05$). ANOVA test with post-hoc pair-wise comparison was applied to the mRNA levels, M3G, M6G and epirubicin-G formation in each diplotype. Data are expressed as mean $\pm$ SD, unless specified otherwise. The results of this study should be regarded as hypothesis generating, as this study was not prospectively powered to detect statistically significant associations between SNPs, haplotypes, diplotypes and phenotypes. Moreover, the p values have not been adjusted for multiple comparisons.

Linear regression analysis of each tSNP genotype vs. the phenotypic data (M3G and M6G formation and expression of reference sequence mRNA) was also performed. For the purpose of this analysis, three univariate models (additive, dominant and recessive) were tested, as the modes of inheritance were either unknown or not clearly identifiable from the plotted data. For example, for the hypothetical variant A>G, we have genotypes A/A, A/G and G/G. The additive model would be: A/A=0, A/G=1, G/G=2. The dominant model would be: A/A=0, A/G=1, G/G=1. The recessive model would be: A/A=0, A/G=0, G/G=1. Hardy-Weinberg equilibrium for each variant (both after sequencing and genotyping) was also tested. Statistical significance of the data was set at $p<0.05$. We used ESEfinder 2.0 [<http://rulai.cshl.edu/tools/ESE>, 18] for predicting the effect of the gene variants comprised in haplotype 4 on exonic splicing enhancer binding sites for specific serine/arginine-rich (SR) proteins (SF2/ASF, SC35, SRp40, SRp55).

Results

***UGT2B7* resequencing in phenotypic extremes**

The frequency distribution of the M3G formation in 49 samples appeared normal ($p>0.1$) (Fig. 2). Resequencing of approximately 7.8 kb in 12 phenotypic extremes revealed the presence of 50 variants in the *UGT2B7* gene, comprising 6 indels and 44 single nucleotide substitutions (Table 2). The former also included a 6 bp indel (-6682_-6683insGCAAAT) and a 324bp Alu repeat at position -2065. With regard to the location of the variants, 14 SNPs were found in exons (including 3 variants in the newly discovered exon 1A) and 2 nonsynonymous variants were identified in exon 1 (644T>A, I215N) and exon 2 (802T>C, Y268H) (Table 2).

Gene haplotypes

Using LDSelect ($r^2\geq 0.9$ and a minor allele frequency cut off of ≥ 0.05), 10 bins were identified among 45 out of the 50 variants (5 variants had a frequency <0.05). Ten tSNPs were genotyped in the 54 samples (tSNPs indicated by arrows, Fig. 3). No deviation from Hardy-Weinberg equilibrium was observed ($p>0.01$). Haplotypes comprising the 10 tSNPs are summarized in Table 3. In the 54 samples, haplotypes 1-6 account for approximately 90% of all haplotypes. Haplotypes 1-6 can be also detected by using only the data from 12 resequenced samples (Table 3). The following common diplotypes were identified: 1/4 (frequency of 0.13), 1/1 (0.11), 1/2 (0.09), 1/3 (0.07), 2/3 (0.07), 3/5 (0.06), and 4/6 (0.06).

M3G formation: haplotype-phenotype correlation

We selected M3G formation as the primary phenotype used to select the samples for resequencing. M3G formation data are available from 49 samples out of 54 due to insufficient availability of microsomes for some samples. The 1/4 diplotype had higher M3G formation compared to 1/1 ($p<0.05$) and 2/3 ($p<0.01$), similar to the 4/6 diplotype ($p<0.05$ and $p<0.01$, respectively) (Fig. 4A). Diplotypes containing haplotype 4 resulted in a significant 45% average increase in the formation of M3G compared to diplotypes without haplotype 4 (1.74 ± 0.21 vs. 1.08 ± 0.09 , respectively, $p=0.002$). This comparison retained statistical significance when the phenotypic extremes were excluded (1.64 ± 0.39 vs. 1.11 ± 0.41 , $p=0.002$).

M6G formation: haplotype-phenotype correlation

In addition to M3G, we sought to investigate the role of *UGT2B7* haplotypes on the formation of M6G, the active metabolite of morphine. M6G formation data are available from 49 samples out of 54 due to insufficient availability of microsomes for some samples. The 1/4 and 4/6 diplotypes had higher M6G formation compared to 2/3 ($p<0.01$ and $p<0.05$, respectively) (Fig. 4B). Diplotypes containing haplotype 4 resulted in a significant 56% average increase in the formation of M6G compared to diplotypes without haplotype 4 (0.13 ± 0.02 vs. 0.09 ± 0.01 , respectively, $p=0.002$). This comparison retained statistical significance when the phenotypic extremes were excluded (0.12 ± 0.03 vs. 0.09 ± 0.03 , $p=0.002$).

Epirubicin-G formation: haplotype-phenotype correlation

Data for the formation of epirubicin-glucuronide (epirubicin-G), a reaction mediated by *UGT2B7*, are available from 53 samples out of 54 due to insufficient availability of microsomes for one sample. The 1/4 and 4/6 diplotypes had higher epirubicin-G formation than the 2/3 diplotype ($p<0.05$, Fig. 4C). The diplotypes containing haplotype 4 resulted in a significant 27% average increase in the formation of epirubicin-G compared to the diplotypes without haplotype 4 (1.08 ± 0.11 vs. 0.85 ± 0.05 , $p=0.04$). This comparison retained statistical significance when the phenotypic extremes were excluded (1.10 ± 0.27 vs. 0.87 ± 0.27 , $p=0.012$).

Expression of *UGT2B7* mRNA reference sequence (*UGT2B7_v1*): haplotype-phenotype correlation

Due to the observed associations with glucuronide formation (see above), we sought to investigate if they were possibly due to an effect of haplotype 4 on *UGT2B7_v1* expression. The diplotype 1/4 had increased *UGT2B7_v1* mRNA expression compared to diplotypes 1/1 ($p<0.05$), 2/3 ($p<0.01$) and 3/5 ($p<0.05$) (Fig. 4D). This was reflected by more than a two-fold increase in mRNA expression in diplotypes containing haplotype 4 [*UGT2B7_v1*/ β -actin, median (range): 0.44 (-1.20, 1.16)] compared to diplotypes without haplotype 4 [0.088 (-1.01, 0.91)] ($p=0.028$, t-test). This comparison retained statistical significance when the phenotypic extremes were excluded [median (range): 0.47 (-0.19, 1.16) vs. 0.21 (-1.01, 0.91) ($p=0.017$, t-test)].

Genotype-phenotype correlation

In this analysis, we aimed to investigate which of the tSNPs in haplotype 4 was likely to be responsible for the observed effect of haplotype 4. Linear regression analysis using three different genetic models indicated that IVS1+985A>G was significantly associated with mRNA levels (for *UGT2B7_v1*), M3G and M6G formation (additive and dominant models, see Tables 4 and 5 and Fig. 5). The IVS1+985A>G variant is included in haplotype 4 and was associated with an increase in M3G and M6G formation and *UGT2B7_v1* mRNA. The IVS1+985A>G variant (ID number 56721) is located in intron 1 and has a frequency of 0.17. The other two variants of haplotype 4 that differ from haplotype 1 were IVS1+1250A>G and IVS1+829T>C, but no effect on phenotype was detected (Tables 4 and 5).

Moreover, in this analysis, we intended to see whether variants not included in haplotype 4 might show a phenotypic effect that was not detected by the haplotype analysis. The IVS1+9_IVS1+10insT variant was significantly associated with M3G and M6G formation in both models. We also evaluated whether the -161T>C and 802T>C variants were correlated with M3G and M6G formation and with mRNA levels: no significant correlation was found in any of the models tested ($p>0.05$, linear regression, data not shown).

ESEfinder analysis of IVS1+985A>G, 735A>G, and 1062C>T

The IVS1+985A>G (ID number 56721) is in high LD with 735A>G (ID number 57049) and 1062C>T (ID number 65704) (Fig. 3). Hence, due to the association between IVS1+985A>G and the phenotypes mentioned above, we investigated the effect of IVS1+985A>G, 735A>G, and 1062C>T on the binding sites for SR proteins. Concerning IVS1+985A>G, the presence of a G allele creates a SRp40 binding site (motif TTTAAG, score 3.40, cut-off 2.68). Similarly, the presence of the G allele in 735A>G creates a SRp55 binding site (motif TAACGTT, score 2.97, cut-off 2.68). Concerning 1062C>T, SRp40 (motif GTACAAG, score 2.94, cut-off 2.68) and SRp55 (motif TGTATA, score 3.61, cut-off 2.68) binding sites are present for each allele (C and T, respectively).

UGT2B7 splicing variants *UGT2B7_v2* and *UGT2B7_v3*

By searching the AceView database (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>), we found two putative alternative splicing variants supported by more than one clone. With PCR amplification, we confirmed the existence of the two variants in human liver mRNA. We designated the newly discovered splicing variants as *UGT2B7_v2* and *UGT2B7_v3*; and the original *UGT2B7* sequence is therefore designated as *UGT2B7_v1* according to the Human Gene Nomenclature Committee (<http://www.genenames.org/>). Further validation studies were conducted to amplify the 5'- and 3'-ends of the two variants by RACE PCR. As a result, we discovered that both variants splice out the exon 1, 4, 5, and 6, but share the exon 2 and 3 of *UGT2B7_v1*. The 3'-end extends from exon 3 to intron 3 and stops 244bp from the end of exon

3. For the 5'-ends, *UGT2B7_v2* comprises exons 1A and 1B, while *UGT2B7_v3* comprises exons 1A alone. The 1A and 1B exons are 400 bp and 132 bp long and located 45 kb and 6.9 kb 5' to exon 1, respectively (Fig. 1A). The splicing out of intron 1A and 1B fits the rule of GT-AG on the exon-intron boundaries (Fig. 1B).

Expression of *UGT2B7_v2* and *UGT2B7_v3* mRNA was observed in human livers (Fig. 6). Their expression was also found in other human tissues (placenta and kidney) but not brain (data not shown). Sequencing of PCR products confirmed the DNA sequences of the two variants (data not shown). *UGT2B7_v2* was detectable in all samples (Fig. 6). However, under the present PCR conditions, *UGT2B7_v3* was undetectable in some samples (Fig. 6). Real time PCR analysis indicated that the mRNA levels of *UGT2B7_v2* were lower than those of *UGT2B7_v1* (Ct value, mean $\pm$ SD, 33.30 $\pm$ 1.84 for *UGT2B7_v2* vs. 22.19 $\pm$ 1.84 for *UGT2B7_v1*, n=52, p<0.0001). In addition, a modest correlation was found between the expression of the reference sequence mRNA and *UGT2B7_v2* mRNA (Pearson r=0.53, p<0.0001).

DISCUSSION

UGT2B7 is an important enzyme for phase II metabolism. This study identified probable functional variants and haplotypes in the *UGT2B7* gene using the phenotype-to-genotype approach of resequencing the tails of the distribution of M3G formation in human livers. In addition to discovering several novel *UGT2B7* variants that have not been deposited in the dbSNP database, our study identified haplotype 4 as a common and functional haplotype increasing enzyme activity by regulating mRNA expression of *UGT2B7*. Haplotype 4 is comprised of 10 tSNPs, and the investigation of which SNP is more likely to be driving the effect of haplotype 4 indicated that IVS1+985A>G (intron 1) is the only tSNP consistently associated with an increase in glucuronide formation and mRNA expression. This is consistent with the fact that IVS1+985A>G is the only variant present only in haplotype 4 and absent in all the other common haplotypes (i.e., haplotypes 1-6). Hence, the effect of haplotype 4 is likely to reside in one of the 3 SNPs of the bin tagged by IVS1+985A>G (Fig. 3).

In agreement with the observed effect of increased mRNA expression, ESEfinder scan of these three SNPs tagged by IVS1+985A>G is suggestive of their role on constitutive *UGT2B7* splicing. For IVS1+985A>G (intronic) and 735A>G (exonic), the variant allele creates splicing enhancers for SR proteins (SRp40 and SRp55, respectively) that are regulators of both alternative and constitutive gene splicing [20]. Concerning the third variant in the bin (i.e., 1062C>T, exon 4), the T allele determines a stronger score for the SRp40 binding site (3.61) compared to that of the C allele (2.94, cut-off 2.68). These observations require biological validation in experimental models. However, the results of ESEfinder are consistent with the hypothesis that the effect of haplotype 4 is probably mediated by a more efficient splicing of the reference sequence mRNA.

While the IVS1+985A>G has never been identified before, 735A>G and 1062C>T (the other two variants in high LD with IVS1 +985A>G) were previously reported in a population of Norwegian cancer patients [11]. In agreement with our study, these two variants were in complete LD, with an allele frequency of 0.11. In an earlier report of Holthe et al. [11], no association was observed between any of the investigated gene variants (including 735A>G and 1062C>T, that are in high LD with IVS1+985A>G) and morphine pharmacokinetics and glucuronidation in cancer patients. This lack of correlation highlights the importance of clinical validation of our findings in prospective trials. Such trials should control for all the possible confounders, including inadequate phenotyping, concomitant medications, variables known a priori to affect the outcome measures and patient ethnicity. In the study of Holthe et al. [11], differences among patients in morphine doses and blood sampling time (during a 1-2 h interval

post-dosing), and concomitant medications (potentially altering morphine metabolism and excretion) might have hampered the detection of significant associations with *UGT2B7* genotypes. Concerning the effect of ethnicity of patients in prospective trials, our study focused on Caucasians only, and the functional role on common haplotypes should be tested in patients of African and Asian background to account for ethnic differences in a study population including different ethnicities. A detailed haplotype analysis has been recently described in Japanese [21]. Ethnic variability in *UGT2B7* allele frequencies has been demonstrated among the three main ethnic groups [9,22,23].

Performing a SNP discovery and biochemical characterization of SNP function is necessary for understanding the role of genetic variation that occurs in the population. In our study, approximately 30% of the variants found in this study have not been reported yet in the dbSNP database (Table 2). Our study led to the discovery of a significant proportion of novel variants using a relatively small sample size (n=12) for resequencing. Using a small sample size for SNP discovery might hamper a detailed characterization of the LD structure of a certain gene. However, when we compared the LD pattern of the *UGT2B7* variants reported in the HapMap database (Caucasian CEPH samples), we observed a close similarity in the pattern of LD between our data and the HapMap data (Fig. 7). Moreover, the most common haplotypes (i.e., haplotypes 1-6) found in the 54 samples using 10 tSNPs could be also identified by using the resequencing data of 12 phenotypic extremes (Table 3). Taken together, these observations indicate that the inferred haplotype analysis is likely to be accurate for the most common haplotypes, despite the limited sample size of the resequenced samples.

Using an *in vitro* model for SNP discovery has both advantages and limitations. The main advantage is that putative variants of functional significance are easier to detect in an *in vitro* study than in a clinical trial. However, the liver system is not devoid of sources of environmental variability, and differences may exist in microsomal yield and storage of fractions. Another source of variation might derive from the induction state of the livers from the previous exposure of the donors to inducers. We tried to account for the induction state of livers by normalizing the phenotypic data (specifically the M3G formation) by the formation of SN-38G (the glucuronide of SN-38, a *UGT1A1* probe); this approach did not lead to an improvement of the observed correlations (data not shown). The *UGT2B7* gene seems to be refractory to upregulation by classical inducers [24-26]. Despite its limitations, the *in vitro* system has the advantage of providing evidence for the functional role of gene variants that should be confirmed in further studies.

In addition to discovering new functional genetic variants of *UGT2B7*, this study demonstrates the existence of alternative splicing variants of the gene. The two newly found *UGT2B7_v2* and *UGT2B7_v3* variants comprise two additional exons and splice out exon 1, 4, 5 and 6 (Fig. 1). Both variants were confirmed with PCR amplification and the 5'- and 3'-ends were validated using RACE PCR. In addition to the liver, their existence has been verified in placenta and kidney but not in brain (data not shown). Splicing variants of *UGT2B7* have never been demonstrated before, and they appear to exist in all samples tested, although *UGT2B7_v3* seems to be expressed at a much lower level than both *UGT2B7_v2* and the reference sequence mRNA. In addition, the expression of *UGT2B7_v2* mRNA seems to share, at least in part, common regulatory mechanisms with the reference sequence *UGT2B7*, as shown by a significant correlation with the expression of the reference sequence mRNA.

The present data on the two newly identified mRNA splicing variants do not elucidate the molecular mechanism behind the alteration of the splicing machinery. Regarding their potential function, our data suggest that exon 1 skipping in *UGT2B7_v2* and *UGT2B7_v3* might alter the substrate binding to the enzyme. Although the expression of the truncated proteins has not been performed, translation analysis showed that the two additional exons 1A and 1B may not

be able to be translated because of the existence of stop codons. Therefore, they may serve as putative 5'-UTRs. Arbitrary open reading frames in both variants, if any, may start from the ATG codon at position 477 of *UGT2B7_v2* or position 369 of *UGT2B7_v3*. If this is the case, both variants would produce the same putative protein equivalent to a truncated short form of *UGT2B7* beginning from the 250Met to 336Val. The 3'-end extend into the intron 3, but only one putative amino acid (Arg) after the 336Val may be translated. This putative N, C-terminal-less protein seems to share only 87 amino acid in the endoplasmic reticulum transmembrane domain and the UDPGA binding domain but may lack the substrate binding domain and the catalytic domain [27-29].

An alternative splicing pattern similar to that of *UGT2B7* can be found in other *UGT2B* genes. *UGT2B* genes and pseudogenes are clustered in chromosome 4q13-q21.1 probably due to the results of gene duplication during human evolution. *UGT2B28* mRNA variants exist, with type III mRNA altering the putative substrate binding domain of the enzyme and retaining its homodimerization capacity [16]. It has been hypothesized that shorter UGT isoforms might be involved in cofactor transport from the cytosol to the cisternal lumen or within membranes that occurs during homo and heterodimerization (30). As the interpretation of the possible functions of these two variants is not based upon any experimental evidence, further biochemical studies are needed to evaluate the presence of cryptic splicing sites and the function of the protein products of *UGT2B7_v2* and *UGT2B7_v3* on *UGT2B7* dimerization.

To conclude, our study supports the hypothesis that a common haplotype of the *UGT2B7* gene might explain part of the phenotypic variability in the pharmacokinetics and pharmacodynamics of *UGT2B7* substrates. As our statistical analyses have not been adjusted for multiple testing, they served the purpose of providing hypotheses for further molecular, biochemical and clinical studies. Studies are planned in postoperative patients treated with morphine to test the effect of haplotype 4 on morphine clearance and acute pain control. Other studies should evaluate the importance of haplotype 4 for *UGT2B7*-mediated detoxification of drugs and inactivation of carcinogens.

ACKNOWLEDGEMENTS

We thank Kathy Hennessy for providing expert technical assistance.

Sponsorship This work was supported by the Pharmacogenetics of Anticancer Agents Research (PAAR) Group (<http://pharmacogenetics.org>) (NIH/NIGMS grant U01GM61393). Data will be deposited into PharmGKB (supported by NIH/NIGMS U01GM61374, <http://pharmgkb.org/>).

REFERENCES

1. Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, et al. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics* 2005;15:677–685. [PubMed: 16141793]Review
2. Miley MJ, Zielinska AK, Keenan JE, Bratton SM, Radomska-Pandya A, Redinbo MR. Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7. *J Mol Biol* 2007;369:498–511. [PubMed: 17442341]
3. Radomska-Pandya A, Little JM, Czernik PJ. Human UDP-glucuronosyltransferase 2B7. *Curr Drug Metab* 2001;2:283–298. [PubMed: 11513331]
4. Lotsch J. Opioid metabolites. *J Pain Symptom Manage* 2005;29(Suppl):S10–24. [PubMed: 15907643] Review
5. Klestad P, Dale O, Skorpen F, Borchgrevink PC, Kaasa S. Genetic variability and clinical efficacy of morphine. *Acta Anaesthesiol Scand* 2005;49:902–908. [PubMed: 16045647]

6. Jin C, Miners JO, Lillywhite KJ, Mackenzie PI. Complementary deoxyribonucleic acid cloning and expression of a human liver urine diphosphate-glucuronosyltransferase glucuronidating carboxylic acid-containing drugs. *J Pharmacol Exp Ther* 1993;264:475–479. [PubMed: 8423545]
7. Coffman BL, King CD, Rios GR, Tephly TR. The glucuronidation of opioids, other xenobiotics, and androgens by human UGT2B7Y(268) and UGT2B7H(268). *Drug Metab Dispos* 1998;26:73–77. [PubMed: 9443856]
8. Innocenti F, Iyer L, Ramirez J, Green MD, Ratain MJ. Epirubicin glucuronidation is catalyzed by human UDP-glucuronosyltransferase 2B7. *Drug Metab Dispos* 2001;29:686–692. [PubMed: 11302935]
9. Bhasker CR, McKinnon W, Stone A, Lo AC, Kubota T, Ishizaki T, et al. Genetic polymorphism of UDP-glucuronosyltransferase 2B7 (UGT2B7) at amino acid 268: ethnic diversity of alleles and potential clinical significance. *Pharmacogenetics* 2000;10:679–685. [PubMed: 11186130]
10. Sawyer MB, Innocenti F, Das S, Cheng C, Ramirez J, Pantle-Fisher FH, et al. A pharmacogenetic study of uridine diphosphate-glucuronosyltransferase 2B7 in patients receiving morphine. *Clin Pharmacol Ther* 2003;73:566–574. [PubMed: 12811366]
11. Holthe M, Rakvag TN, Klepstad P, Idle JR, Kaasa S, Krokan HE, et al. 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* 2003;3:17–26. [PubMed: 12629580]
12. Ross JR, Rutter D, Welsh K, Joel SP, Goller K, Wells AU, et al. Clinical response to morphine in cancer patients and genetic variation in candidate genes. *Pharmacogenomics J* 2005;5:324–336. [PubMed: 16103897]
13. Duguay Y, Baar C, Skorpen F, Guillemette C. A novel functional polymorphism in the uridine diphosphate-glucuronosyltransferase 2B7 promoter with significant impact on promoter activity. *Clin Pharmacol Ther* 2004;75:223–233. [PubMed: 15001974]
14. Coulbault L, Beaussier M, Verstuyft C, Weickmans H, Dubert L, Tregouet D, et al. Environmental and genetic factors associated with morphine response in the postoperative period. *Clin Pharmacol Ther* 2006;79:316–324. [PubMed: 16580900]
15. Darbari DS, van Schaik RH, Capparelli EV, Rana S, McCarter R, van den Anker J. UGT2B7 promoter variant -840G>A contributes to the variability in hepatic clearance of morphine in patients with sickle cell disease. *Am J Hematol* 2008;83:200–202. [PubMed: 17724700]
16. Levesque E, Turgeon D, Carrier JS, Montminy V, Beaulieu M, Belanger A. Isolation and characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. *Biochemistry* 2001;40:3869–3881. [PubMed: 11300766]
17. Levesque E, Girard H, Journault K, Lepine J, Guillemette C. Regulation of UGT1A1 bilirubin-conjugating pathway: role of a new splicing event at the UGT1A locus. *Hepatology* 2007;45:128–138. [PubMed: 17187418]
18. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–120. [PubMed: 14681826]
19. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003;31:3568–3571. [PubMed: 12824367]
20. Wang J, Smith PJ, Krainer AR, Zhang MQ. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res* 2005;33:5053–5062. [PubMed: 16147989]
21. Saito K, Moriya H, Sawaguchi T, Hayakawa T, Nakahara S, Goto A, et al. Haplotype analysis of UDP-glucuronosyltransferase 2B7 gene (UGT2B7) polymorphisms in healthy Japanese subjects. *Clin Biochem* 2006;39:303–308. [PubMed: 16466707]
22. Mehlotra RK, Bockarie MJ, Zimmerman PA. Prevalence of UGT1A9 and UGT2B7 nonsynonymous single nucleotide polymorphisms in West African, Papua New Guinean, and North American populations. *Eur J Clin Pharmacol* 2007;63:1–8. [PubMed: 17115150]
23. Lampe JW, Bigler J, Bush AC, Potter JD. Prevalence of polymorphisms in the human UDP-glucuronosyltransferase 2B family: UGT2B4(D458E), UGT2B7(H268Y), and UGT2B15(D85Y). *Cancer Epidemiol Biomarkers Prev* 2000;9:329–333. [PubMed: 10750673]

24. Munzel PA, Schmohl S, Heel H, Kalberer K, Bock-Hennig BS, Bock KW. Induction of human UDP glucuronosyltransferases (UGT1A6, UGT1A9, and UGT2B7) by tbutylhydroquinone and 2,3,7,8-tetrachlorodibenzo-p-dioxin in Caco-2 cells. *Drug Metab Dispos* 1999;27:569–573. [PubMed: 10220484]
25. Walle T, Otake Y, Galijatovic A, Ritter JK, Walle UK. Induction of UDP-glucuronosyltransferase UGT1A1 by the flavonoid chrysin in the human hepatoma cell line hep G2. *Drug Metab Dispos* 2000;28:1077–1082. [PubMed: 10950852]
26. Sabolovic N, Magdalou J, Netter P, Abid A. Nonsteroidal anti-inflammatory drugs and phenols glucuronidation in Caco-2 cells: identification of the UDP-glucuronosyltransferases UGT1A6, 1A3 and 2B7. *Life Sci* 2000;67:185–189. [PubMed: 10901286]
27. Senay C, Ouzzine M, Battaglia E, Pless D, Cano V, Burchell B, et al. Arginine 52 and histidine 54 located in a conserved amino-terminal hydrophobic region (LX2-R52-GH54-X3-V-L) are important amino acids for the functional and structural integrity of the human liver UDP-glucuronosyltransferase UGT1*6. *Mol Pharmacol* 1997;51:406–413. [PubMed: 9058595]
28. Meech R, Mackenzie PI. UDP-glucuronosyltransferase, the role of the amino terminus in dimerization. *J Biol Chem* 1997;272:26913–26917. [PubMed: 9341125]
29. Coffman BL, Kearney WR, Green MD, Lowery RG, Tephly TR. Analysis of opioid binding to UDP-glucuronosyltransferase 2B7 fusion proteins using nuclear magnetic resonance spectroscopy. *Mol Pharmacol* 2001;59:1464–1469. [PubMed: 11353807]
30. Ikushiro S, Emi Y, Iyanagi T. Protein-protein interactions between UDP-glucuronosyltransferase isozymes in rat hepatic microsomes. *Biochemistry* 1997;6:7154–7161. [PubMed: 9188715]

Fig. 1.

A) Gene structure and resequencing coverage. The horizontal lines above the genomic structure of the gene indicate the sequencing coverage. The arrows show the position of the translation start sites (putative for the *UGT2B7_v2* and *UGT2B7_v3*). **B)** Exon 1A (400 bp) and 1B (132 bp) sequence of alternative splicing variants. Exons are in lower case and underlined. Exon-intron boundaries are shown, and GT-AG nucleotides are indicated in bold.

Fig. 2. Frequency distribution of M3G formation in human livers

Data are expressed as the chromatographic peak height ratio between M3G and dihydrocarbamazepine (internal standard) (n=49).

Fig. 3. Selection of tSNPs

Ten bins of the 45 polymorphisms (here classified according to the ID number in Table 2) were identified by LDSelect. The top colored panel shows the clustered visual genotype of each polymorphism (blue: homozygote of the common allele; red: heterozygote; yellow: homozygote of the rare allele). Rows 1-6 show samples with high UGT2B7 activity; rows 7-12 show samples with low UGT2B7 activity. The bottom panel shows the pair-wise LD pattern with $r^2 \geq 0.9$. Ten tSNPs (one for each bin) are indicated by arrows.

Fig. 4. Diplotype-phenotype correlation analysis

Data show only diplotypes with $n \geq 3$. Four samples were not used because the diplotype assignment was ambiguous. A) Correlation between estimated diplotypes and M3G formation ($n=32$, ANOVA, $p=0.002$). B) Correlation between estimated diplotypes and M6G formation ($n=32$, ANOVA, $p=0.005$). C) Correlation between estimated diplotypes and epirubicin-G formation ($n=32$, ANOVA, $p=0.008$). D) Correlation between estimated diplotypes and mRNA levels of the *UGT2B7_v1* gene ($n=31$, ANOVA, $p=0.006$). The *UGT2B7_v1* mRNA levels were normalized to $\beta\text{-actin}$ and the ratio log transformed. For the 4/6 diplotypes, only 2 samples ($n=2$) are shown in graph D. In each graph, bars represent the mean values.

Fig. 5. Linear regression analysis of IVS1+985A>G tSNP vs. reference sequence mRNA (A, D), and M3G (B, E) and M6G (C, F) formation
Additive (A, B, C) and dominant (D, E, F) models.

Fig. 6. Expression of *UGT2B7* reference sequence (*UGT2B7_v1*), splicing variants *UGT2B7_v2* and *UGT2B7_v3*, and β -actin using PCR analysis

Eleven representative livers were used: the sample in lane 8 showed partial RNA degradation.

Fig. 7. Pairwise r^2 analysis of the variants that have been found both in our resequencing of 12 samples (left) and in the Caucasian HapMap samples (right)

Table 1

PCR, sequencing, and genotyping primers.

Primer name ¹	Sequence 5' -> 3'	Primer position on reference sequence ²
<i>PCR and sequencing primers (promoter, coding and 3'UTR)</i>		
PromF	TGG CAT CAG TGA AGT CTC TC	53887
PromR	ACA CCA GCA CCT TTC CAC AA	55081
PromFSeq	CTG ATT GTT ATG GTA GAT GC	54514
PromRSeq	TAT TTT AGA AGA GTG TCC AGA	54709
Ex1F2	TAC ATT TTA ACT TCT TGG CTA A	54919
Ex1R	ATT TCC TGG GGT GTT CCG T	55943
Ex1FSeq	CAC AAG TAC AGG AAA TCA TGT	55344
Ex1RSeq	GTC AAA TCT TGA CTC TTG TAC	55433
Ex2F2	TGA AAC TAT GTC TCT TTA TTA G	56373
Ex2R2	TTT TTA TAT GAT TTT TAT TAC TTG C	57592
Ex2FSeq	TTT TTT TTT CTA TTC CTG TCA G	57014
Ex2RSeq	AAA AGG TGG ATG ACA CAA GA	56996
Ex3F2	AGA ACC TAT ATT AGT AAC TTT AG	60940
Ex3R2	TTA TGG ACA GGA GGT GAG AA	61694
Ex3FSeq	TAA ATG GTG TTA AGT ATG AAC	61606
Ex3RSeq	ATG CAA CCA CAA TTT TCA AAC	61577
Ex4F	GTG AGT ATT CTA TTT ACA TTA GTC	65502
Ex4R	TGG TAG ATT GCC TCG TAG ATG C	66617
Ex5F2	TTC TTT ATA GTC TGA ATC GGG	65922
Ex5R	TAT TCT TCC TTT CTT CTA TGG	67078
Ex5FSeq	ATT CCT ATG AGT AAT TTT GCT AA	66532
Ex5RSeq	TAT AAA AAG GAT GAA ACT CAC AC	66855
Ex6F2	GCA GAC CCC CTT AGA GTT	70734
Ex6R	GCC AAG CAA TAC TAC TCA TC	71730
Ex6FSeq	GAA GGG AAA AAA TGA TTA GTT A	71184
Ex6RSeq	TTG CTG GAA TAA ACT GAA GTA G	71244
3'UTRF	TTC AGA GAT TTA CCA CCC AGT T	71364
3'UTRR	GAG AAT AAA GTC AAC CAG ATG T	71961
<i>PCR and sequencing primers (5' flanking region and alternatively spliced exons)</i>		
5' FR F	TGG CAT TGG TGG TCA TAG TAT T	8850
5' FR R	CAA GAA ATG TAA CAG GAG GAG C	10096
5' FR Seq F	TTA CCA GGG AAT ATG CAA ATG T	9240
5' FR Seq R	AGG TAC TGA GGA TAA AAC AAT A	9668
Ex 1A F	ATC AAT CAC GAC CCT TTC ATG T	9911
Ex 1A R	AGC CGA GTC AAA ACC AGA ACA A	10460
Ex 1B F	ATA CAT ATC ACA ACT TCT TCA G	48205
Ex 1B R	AGT AAC TTC AAG CAT TCA TAG G	48628

Primer name ¹	Sequence 5' -> 3'	Primer position on reference sequence ²
PCR and single base extension/ primer extension		
Amplicon 1F PCR	TAC ATT TTA ACT TCT TGG CTA A	54919
Amplicon 1R PCR	AGT TAG AGC TTC ATG TTA CTG	55775
55152 Ext F18 SBE	AGA CAA TCC TGG ATG AGC	55152
55387 Ext F18 SBE	ATT TGG TGA CAT AAG TAG	55387
55659 Ext R21 PE	AAG TCA AAG TAA AGC ACA TAG	55659
9419F PCR	TAA CAT CTT CCC AAT CAC AAG T	9319
9419R PCR	TAG CAT CTT TAC AGA GTA GTG	9575
9419 Ext F19 SBE	GCC TGG AAT AGA TAC AAG T	9419
Amplicon 2F PCR	GAT TTC AGC CAT ACT CTC AAC A	56536
Amplicon 2R PCR	GAG TCC TCC AAC AAA ATC AAC	57154
56721 Ext F19 SBE	CTA CTA TTG AAG GTT TAA G	56721
56735 Ext R28 SBE	AAT AAA ATG AGC CAT AGA AAG TTA AAC A	56735
57049 Ext R22 SBE	CTT TCC CCA TTG TCT CAG ATA A	57049
65917F PCR	GTG AGT ATT CTA TTT ACA TTA GTC	65502
65917R PCR	AGA CAG CCC AGG AAG ATA TTG	66005
65917 Ext R22 SBE	GAT TCA GAC TAT AAA GAA TGT G	65917
Sizing for indel genotyping		
48334 FAM-F	ACATGGCAGAATATAGGGACA	48176
48334 R	AAGGTTTTCACAACTGGC	48384
55747 FAM-F	TTTCCCTCCTTCCTACGTAC	55578
55745 R	AAAACCCCACTACCCT	55834

Seq = sequencing, SBE = single-base extension, PE = primer extension

¹ All forward primers are indicated by an F and all reverse primers are indicated by an R.

² Accession number is AC107401.4 for reference sequence.

Table 2

Polymorphisms discovered in the *UGT2B7* gene in 12 resequenced samples.

ID	Variant	Exon-Intron	Amino acid	Position in chromosome 4	5'-flanking	3'-flanking	MAF	dbSNP rs#	Holthe et al. [11]
7959	-47057T>C	-	-	70164854	ATAATTTTCA	ATTTTGTTTT	0.13	rs4618379	
9034	-45982A>C	-	-	70165929	TTGATAAAAT	GGTTTCATCA	0.38	rs7435274	
9400	-45616G>T	-	-	70166295	CAGAAAAAGCT	CCTGGAATAA	0.04	-	
9419	-45597G>T	-	-	70166314	AAATACAAGT	ATTAGCAGA	0.38	rs7438808	
9683	-45333_-45334insA	-	-	70166578	TAAGAAAAAA	CAAAATAAGTG	0.13	rs5859158	
9896	-45120C>T	-	-	70166792	TGTCAAAGTAC	GCCTGCTTGC	0.04	-	
10074	-44942G>A	-	-	70166970	TTTTGTCTGC	GCTCCTCTCTG	0.38	rs4422474	
10107	-44909C>T	-	-	70167003	GCTCCCTGAC	GGGAAGCAAC	0.38	rs4377650	
10351	-44665T>C	-	-	70167504	ACATCTAAAA	TTGTTAAGAC	0.04	-	
48334	-6682_-6683insGCAAAT	-	-	70205487	CCAGATCTCT	GCAAAAGATGG	0.08	rs36053095	
52950	-2065_-2066insALU*	-	-	70210103	TTTGGTTTCT	AGGCATATAA	0.46	-	
53710	-1306A>G	-	-	70210863	AGATAAGAAAG	GAACCCCAAA	0.46	rs4455491	0.35
53717	-1299T>C	-	-	70210870	AAGAGAAACCC	AAATCCAGCT	0.46	rs11940220	0.35
53904	-1112T>C	-	-	70211057	AGTGAAGTCT	TCAGATCAGC	0.46	rs11940316	0.35
54116	-900G>A	-	-	70211269	GAAGTGAGTC	GAGAACCAAGC	0.46	rs7438135	0.35
54689	-327A>G	-	-	70211842	TGCAGAAACA	GATCTGTCAC	0.46	rs7662029	0.39
54855	-161T>C	-	-	70212006	CATTGTCTCT	TTGCCATCCA	0.46	rs7668258	0.44
54878	-138G>A	-	-	70212029	CATGCTCAGA	TGTTGATTTA	0.08	-	
55152	137T>C	Exon 1	L46P	70212305	CTGGATGAGC	TATTCAGAGA	0.04	-	
55387	372A>G	Exon 1	R124R	70212540	ACATAAACTAG	AAGTCTCTGA	0.13	-	0.03
55659	644T>A	Exon 1	I215N	70212812	AAAAATATGA	CTATGTGCTT	0.04	-	
55745	IVS1+9_IVS1+10insT	Intron 1	-	70212898	AGGTAAGTAT	TTTTTCAATC	0.13	-	
55937	IVS1+201A>C	Intron 1	-	70213090	TGTGGCCATC	CTCACACGGA	0.5	rs12647681	
56565	IVS1+829T>C	Intron 1	-	70213718	ACATAGTCCA	AGTTCACTTG	0.46	rs28375964	
56721	IVS1+985A>G	Intron 1	-	70213874	AAGCTTTAAG	GAAAAAGAAAT	0.17	-	
56735	IVS1+999C>A	Intron 1	-	70213888	AAGAAATTTT	TGTTTAACTT	0.17	-	
56854	IVS1+1118C>T	Intron 1	-	70214007	CTTATAAATA	ACATGGGCAA	0.08	-	
56957	IVS1+1221T>C	Intron 1	-	70214110	TTTGCCTACA	TATTTAAACC	0.46	rs7439326	
56986	IVS1+1250A>G	Intron 1	-	70214139	AAATTACCT	AAGTAATTAT	0.5	rs7438244	

ID	Variant	Exon-Intron	Amino acid	Position in chromosome 4	5'-flanking	3'-flanking	MAF	dbSNP rs#	Holthe et al. [11]
57049	735A>G	Exon 2	T245T	70214201	GACCCACTAC	TTATCTGAGA	0.17	rs28365062	0.11
57115	801A>T	Exon 2	P267P	70214267	TTCAGTTTCC	TATCCACTCT	0.46	rs7438284	0.44
57116	802T>C	Exon 2	H268Y	70214268	TCAGTTTCCA	ATCCACTCTT	0.46	rs7439366	0.44
57169	855C>A	Exon 2	A285A	70214321	GCAAACTGCG	AAACCCTGCG	0.08	rs34661811	
57299	IVS2+115G>A	Intron 2	-	70214451	GGAAATGGGT	GGGTAAAGCA	0.46	rs7441750	
57332	IVS2+148G>A	Intron 2	-	70214484	AAATCTATGT	CACGTTAATA	0.46	rs7441774	
65701	1059C>G	Exon 4	L353L	70222879	ATACTCGGCT	TACAAAGTGA	0.46	rs4292394	0.39
65704	1062C>T	Exon 4	Y354Y	70222882	CTCGGCTGTA	AAGTGATAC	0.17	rs4348159	0.14
65796	IVS4+64A>T	Intron 4	-	70222974	TTAATAGTTC	TCATGAAACA	0.46	rs4337789	
65886	IVS4+154_IVS4+155insA	Intron 4	-	70223039	AGAAAGAATA	TTATAATTTT	0.46	rs11302069	
65917	IVS4+185C>A	Intron 4	-	70223096	TGATATACAC	CACATTCCTT	0.17	-	
66276	IVS4+544C>A	Intron 4	-	70223455	AATAAGAGTT	ACTTCATGCC	0.46	rs4521414	
66391	IVS4+659A>G	Intron 4	-	70223570	GAAATCACAC	CACCGTATAG	0.46	rs4314345	
66418	IVS4+686C>G	Intron 4	-	70223597	GTTACATACC	AGTACAAGTA	0.46	rs4588522	
66443	IVS4+711C>T	Intron 4	-	70223622	TTTTTTCCTC	GAAAGTCTGAA	0.46	rs4364327	
66477	IVS4+745G>A	Intron 4	-	70223656	TTAGTTCAGT	TTTTAGCTGG	0.13	rs10022440	
66672	1191C>T	Exon 5	A397A	70223851	CATTGTTTGC	GATCAACCTG	0.08	-	
66976	IVS5+185_IVS5+186insA	Intron 5	-	70224155	ACATACATCT	AAGAATAGCC	0.08	-	
70816	IVS5+4025G>A	Intron 5	-	70227995	GAGGAGTCTT	CCGATGCTCC	0.46	rs7658752	
71498	530+296T>C	3'UTR	-	70228680	AAATTGTGTG	TCTAATTCAC	0.46	rs6851533	
71649	550+447T>C	3'Flanking	-	70228831	TTGAACAGGA	TGTATTGTTT	0.46	rs6600893	

The position of each polymorphism (ID) in the gene structure and on chromosome 4 is shown (*UGT2B7* accession number, AC 107401.4). The variant position is relative to the ATG translation start site on exon 1. Surrounding (5'- and 3'-flanking) sequences of each variant are also indicated. The SNPs are indicated as common/rare allele. Minor allele frequency (MAF) is shown. The variants in bold indicate the 10 tSNPs. The variants in italics were also genotyped in the entire sample set of 54 individuals, in addition to the 10 tSNPs. The ID number is in reference to Fig. 3, for simplicity. The rs# is from the dbSNP database (build 127). The frequency of variants found in Holthe et al. [11] is also provided.

MAF, minor allele frequency.

* Denotes the insertion of a 327 bp ALU element.

Table 3
Haplotypes of 10 tSNPs estimated from 12 resequenced and 54 genotyped Caucasian samples.

Haplotype	tSNP										Frequency in resequenced samples (n=12)	Frequency in all samples (n=54)
	-45597 G>T	-6682_-6683 insGCCAAAT	372 A>G	IVS1+9_IVS1+10 insT	IVS1+829 T>C	IVS1+985 A>G	IVS1+999 C>A	IVS1+1250 A>G	801 A>T	IVS4+185 C>A		
1	G	A	A	A	C	A	C	A	A	C	0.13	0.32
2	T	A	A	A	T	A	C	G	T	A	0.17	0.17
3	G	A	A	A	C	A	A	A	A	C	0.17	0.14
4	G	A	A	A	T	G	C	G	T	C	0.17	0.12
5	T	A	G	A	T	A	C	G	T	C	0.13	0.06
6	G	A	A	C	C	A	C	A	A	C	0.13	0.06
7	T	C	G	A	T	A	C	G	T	C	-	0.04
8	T	A	A	A	T	A	C	A	A	C	-	0.03
9	T	C	A	A	T	A	C	G	T	A	-	0.01
10	T	A	A	C	T	A	C	G	T	A	-	0.01
11	G	A	A	A	C	A	C	G	A	C	0.04	0.01
12	T	C	A	A	T	A	C	A	A	C	0.08	0.01
13	T	A	A	A	T	G	C	G	T	C	-	0.01
14	G	A	A	A	T	A	C	G	T	C	-	0.01
15	T	A	A	A	T	A	C	G	T	C	-	0.01

Table 4
Additive model of linear regression analysis of tSNPs vs. phenotypes.

	UGT2B7 reference sequence mRNA (UGT2B7_v1)			M3G formation			M6G formation		
	n	r	p	n	r	p	n	r	p
-45597G>T	52	-0.22	0.1	50	-0.15	NS	50	-0.17	NS
-6682_-6683insGCAAAT	51	0.03	NS	49	0.00	NS	49	0.01	NS
372A>G	52	-0.27	0.05	50	-0.21	NS	50	-0.21	0.1
IVS1+9_IVS1+10insT	51	0.05	NS	49	0.32	0.024	49	0.28	0.049
IVS1+829T>C	52	-0.05	NS	50	0.08	NS	50	0.07	NS
IVS1+985A>G	52	0.31	0.026	50	0.32	0.025	50	0.31	0.026
IVS1+999C>A	52	0.09	NS	50	0.15	NS	50	0.12	NS
IVS1+1250A>G	52	-0.08	NS	50	0.00	NS	50	-0.03	NS
801A>T	52	-0.01	NS	50	0.06	NS	50	0.03	NS
IVS4+185C>A	52	0.02	NS	50	0.04	NS	50	0.07	NS

Significant regressions ($p \leq 0.05$) are in *italics*.
NS, non significant ($p > 0.05$); n, number of samples; r, Pearson's correlation coefficient; p, p value for Pearson's correlation.

Table 5
Dominant model of linear regression analysis of tSNPs vs. phenotypes.

	UGT2B7 reference sequence mRNA			M3G formation			M6G formation		
	n	r	p	n	r	p	n	r	p
-45597G>T	52	-0.27	0.05	50	-0.16	NS	50	-0.16	NS
-6682_-6683insGCAAAAT	51	0.03	NS	49	0.00	NS	49	0.01	NS
372A>G	52	-0.27	0.05	50	-0.21	0.2	50	-0.21	0.1
IVS1+9_IVS1+10insT	51	0.00	NS	49	0.32	0.023	49	0.29	0.047
IVS1+829T>C	52	-0.08	NS	50	0.10	NS	50	0.10	NS
IVS1+985A>G	52	0.29	0.033	50	0.36	0.010	50	0.35	0.01
IVS1+999C>A	52	0.09	NS	50	0.15	NS	50	0.12	NS
IVS1+1250A>G	52	-0.06	NS	50	0.01	NS	50	-0.02	NS
801A>T	52	-0.06	NS	50	0.01	NS	50	-0.02	NS
IVS4+185C>A	52	0.05	NS	50	-0.16	NS	50	-0.15	NS

Significant regressions ($p \leq 0.05$) are in *italics*.
NS, non significant ($p > 0.05$); n, number of samples; r, Pearson's correlation coefficient; p, p value for Pearson's correlation.