Hepatic CYP2B6 Expression: Gender and Ethnic Differences and

Relationship to *CYP2B6* Genotype and CAR Expression

Vishal Lamba, [1] Jatinder Lamba, Kazuto Yasuda, Stephen Strom, Julio Davila, Michael
L. Hancock, James D. Fackenthal, Peter K. Rogan, Barbara Ring, Steven A. Wrighton and
Erin G. Schuetz

Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital (V.L.,
J.L., K.Y., M.L.H., E.G.S.), Memphis, Tennessee; Department of Pathology,
University of Pittsburgh (S.S.), Pittsburgh, Pennsylvania; Pfizer Corp (J.D.), St. Louis,
Missouri; Department of Medicine, University of Chicago Medical Center (J.D.F.),
Chicago, Illinois; Schools of Medicine and Interdisciplinary Computing and Engineering,
University of Missouri (P.K.R.), Kansas City, Missouri; and Eli Lilly & Company
(B.R., S.A.W.), Indianapolis, Indiana

b) Corresponding Author: Erin G. Schuetz, Ph.D., Department of Pharmaceutical

Sciences, Mail Stop 313, St. Jude Children's Research Hospital, 332 N. Lauderdale

Street, Memphis, TN 38105; PH: (901) 495-2205; Fax: (901)525-6869; Email:

erin.schuetz@stjude.org

Abstract. CYP2B6 metabolizes many drugs, and its expression varies greatly. CYP2B6 genotype-phenotype associations were determined using human livers that were biochemically phenotyped for CYP2B6 (mRNA, protein, and CYP2B6 activity), and genotyped for *CYP2B6* coding and 5'-flanking regions. CYP2B6 expression differed significantly between sexes. Females had higher amounts of *CYP2B6* mRNA (3.9-fold, P<0.001), protein (1.7-fold, P<0.009), and activity (1.6-fold, P<0.05) than did male subjects. Further, 4.5% of females and 20% of males were CYP2B6 poor metabolizers. Striking differences among different ethnic groups were observed: CYP2B6 activity was 3.6- and 5.0-fold higher in Hispanic females than in Caucasian (P< 0.022) or African American females (P<0.038). Ten single nucleotide polymorphisms (SNPs) in the *CYP2B6* promoter and seven in the coding region were found; including a newly identified 13072A>G substitution that resulted in a Lys139Glu change. Many *CYP2B6* splice variants (SV) were observed, and the most common variant lacked exons 4-6. A non-synonymous SNP in exon 4 (15631G>T), which disrupted an exonic splicing enhancer, and a SNP 15582C>T in a intron-3 branch site were correlated with this SV. The extent to which *CYP2B6* variation was a predictor of CYP2B6 activity varied according to sex and ethnicity. The 1459C>T, which resulted in the Arg487Cys substitution was associated with the lowest level of CYP2B6 activity in livers of females. The intron-3 15582C>T SNP (in significant linkage disequilibrium with a SNP in a putative HNF4-binding site) was correlated with lower CYP2B6 expression in females. In conclusion, we found several common SNPs that are associated with polymorphic CYP2B6 expression.

Cytochrome P4502B6 (CYP2B6) is expressed in human liver and some extrahepatic tissues. CYP2B6 activity in liver microsomes varies more than 100-fold among different persons (Yamano et al., 1989; Ekins et al., 1998). This finding suggests that there are significant interindividual differences in the systemic exposure to a variety of drugs that are metabolized by CYP2B6. The number of substrates that are partially or completely metabolized by CYP2B6 is estimated to be around 70 (Ekins et al., 1998; Ekins et al., 1999). CYP2B6 substrates include clinically prescribed drugs such as alfentanil, ketamine, propofol, cyclophosphamide, ifosfamide, nevirapine, efavirenz, bupropion, and tamoxifen; drugs of abuse such as methylenedioxymethamphethamine (MDMA, "ecstasy") and recreational drugs such as nicotine; and procarcinogens such as the environmental contaminants aflatoxin B1 and dibenzanthracene (Lang et al., 2001). Because differences in systemic exposure to drugs or environmental chemicals metabolically cleared by CYP2B6 may lead to variation in therapeutic and toxic responses to these xenobiotics, it is important to determine all of the genetic variants responsible for altered CYP2B6 expression. This is especially important because there is no validated drug that can currently be used as a probe to identify CYP2B6 activity *in vivo* and to guide dosing of CYP2B6 substrates with narrow therapeutic indices.

To determine the contribution of *CYP2B6* genetic variants to polymorphic CYP2B6 expression, a recent study resequenced *CYP2B6* exons in cDNA derived from 35 German Caucasians. Nine single nucleotide polymorphisms (SNPs) including five that caused non-synonymous amino acid changes, were found (Lang et al., 2001). A 1459C>T substitution in exon 9, which resulted in a Arg487Cys substitution, was associated with CYP2B6 activity

that was 8-fold lower in Cys^{487} homozygotes than in Arg^{487} homozygotes. However, the frequency of this variant (14%) in a German Caucasian population cannot explain the large variability in CYP2B6 activity. Using expressed cDNAs, another group found that 7-ethoxycoumarin *O*-deethylase activity of CYP2B6 was greater when the enzyme contained His^{172} (encoded by exon 4) rather than Gln^{172} (Ariyoshi et al., 2001). Further, it has been reported that compared with *CYP2B6*1*, the alleles *CYP2B6*4* (Lys262Arg), *2B6*5* (Arg487Cys), *2B6*6* (Gln172His; Lys262Arg), and *2B6*7* (Gln172His; Lys262Arg; Arg487Cys) are associated with a higher intrinsic clearance (V_{max}/K_m) of 7-ethoxy-4-trifluoromethylcoumarin. Moreover, *CYP2B6*2* (Arg22Cys) reportedly does not affect CYP2B6 activity (Jinno et al., 2003).

Although alternatively spliced transcripts of *CYP2B6* have been reported, the early splicing studies (Miles et al., 1988; Yamano et al., 1989) did not attempt to fully characterize the splicing of *CYP2B6* transcripts, and these studies contained a relatively small number of samples. Neither of the previous genotyping studies of *CYP2B6* (Lang et al., 2001; Ariyoshi et al., 2001) identified splice site variations that could lead to alternatively spliced mRNA. However, both studies followed the exon-by-exon sequencing approach of genomic DNA analysis, and splice variants could only be inferred on the basis of canonical splice acceptor and donor nucleotides.

There is a growing awareness that some variation in human gene expression is dramatically influenced by variation in regulation of that expression. CYP2B expression is regulated by the transcription factor CAR (constitutive androstane receptor) (Wei et al., 2000), whose hepatic expression was recently reported to be highly correlated with that of

5

hepatic CYP2B6 (Chang et al., 2003). However, this association needs to be validated in a larger study because liver samples from only 12 human subjects were analyzed.

To further investigate the polymorphic behavior of CYP2B6 and its causes, we conducted the study described in the present report. Our objectives were to determine whether alternative splicing is an important determinant of polymorphic expression of CYP2B6 in the liver; to determine whether SNPs in the coding region, introns, or 5'-flanking sequences of *CYP2B6* are correlated with altered CYP2B6 expression in the livers of Caucasian, African, or Hispanic Americans; and to determine whether expression of CAR is correlated with CYP2B6 expression in a large number of human livers. Here we report a striking difference in CYP2B6 expression between sexes, newly identified alternatively spliced mRNA variants, several newly found SNPs whose presence is correlated with altered CYP2B6 expression, and an association between CYP2B6 expression and CAR expression.

Materials and Methods

Materials. S-Mephenytoin was purchased from Toronto Research Chemicals (North York, Ontario, Canada) and purified at Eli Lilly & Company (Indianapolis, IN). We bought nirvanol from Ultrafine (Manchester, United Kingdom) and phenobarbital and NADPH from Sigma Chemical Co. (St. Louis, MO).

Study Subjects and Liver Samples. Genomic DNA, total RNA, microsomes, or a combination of these materials was prepared from liver samples (University of Pittsburgh/St Jude Liver Resource). The majority of the livers used in this study were from organ donors and additional relevant information on the donor history is listed in legends to the Tables. The liver samples had been isolated from 51 male subjects, 30 female subjects, and 2 subjects whose sex was unknown. Of the 80 subjects whose ethnicity was reported 43 were Caucasian (15 females, 28 males), 29 were African Americans (9 Females and 18 Males), 7 were Hispanic (3 females, 4 males), and 1 was Asian (female). The ethnicity of three of the samples (2 Females and 1 male) and gender for two (both African Americans) was unknown.

GenBank Accession Numbers. *CYP2B6* promoter sequence, AF 081569 (Sueyoshi et al., 1999; Wang et al., 2003); *CYP2B6* gene sequence, NG_000008 and AC023172; *CYP2B6* mRNA, NM_000767; *CYP2B6* exon 3A, X16864 (Miles et al., 1990); and *CYP2B7* gene, AC008537. Other *CYP2B6* sequences used were those with the accession numbers X06399 (Miles et al., 1988), X06400 (Miles et al., 1988), X13494 (Miles et al., 1989), M29873, M29874, and J02864 (Yamano et al., 1989).

Western Blot Analysis. Quantitative immunoblotting of CYP2B6 from human liver

microsomes was performed as described on 60 human livers (Ekins et al., 1997). Twenty

micrograms of liver microsomes prepared from donor tissue were separated on 7.5% slab

polyacrylamide gels, and 0.05 to 5 pmol of human CYP2B6 (supersomes from GENTEST,

Woburn, MA) were included so that a standard curve could be developed as we have done

previously (Ekins et al., 1997). Blots were incubated with an anti-CYP2B6 antibody

(GENTEST) and subsequently with a biotinylated rabbit anti–mouse antibody and

streptavidin alkaline phosphatase. Bound antibody was detected by treatment with 5-bromo-

4-chloro-3-indoyl-phosphate/nitroblue tetrazolium (BCIP/NBT) (Kirkegaard & Perry

Laboratories, Gaithersburg, MD). The integrated optical density of each band was determined

by computer using the public domain NIH Image program (developed at the U.S. National

Institutes of Health and available at http://rsb.info.nih.gov/nih-image/).

CYP2B6-Mediated Biotransformation of S-Mephenytoin to Nirvanol. Microsomes

from 73 human liver samples were characterized for the CYP2B6 form-selective formation of

nirvanol after incubation with *S*-mephenytoin (Heyn et al., 1996). Duplicate mixtures of 100

µM *S*-mephenytoin and human liver microsomes (0.5 mg protein/ml) in 0.1 M sodium

phosphate, pH 7.4, were preincubated at 37^oC for approximately 3 minutes before the

reaction was initiated by the addition of 0.6 mM NADPH (total volume, 500 µL). After a 45-

minute incubation period (we used linear rate conditions for nirvanol formation with respect

to protein concentration and time of incubation), the reaction was stopped with 0.1 mL of

2% sodium azide containing phenobarbital as an internal standard. Liquid extraction was then undertaken with the addition of 5 mL dichloromethane and a subsequent 10-minute mixing period. The sample was centrifuged, the aqueous layer discarded, and the dichloromethane layer transferred to a clean tube. This sample was dried under nitrogen for 20 to 30 minutes at 37°C.

HPLC was performed to determine the amount of nirvanol that had formed (Heyn et al., 1996). The sample was resuspended in 160 µL of mobile phase (20 mM perchlorate, pH 2.5; methanol; acetonitrile [ratio of volume of each component: 69:25:6, respectively]). Forty microliters of the sample was injected into a Nucleosil C18 column (5 µm, 150 x 4.6 mm; MetaChem, Torrance CA), and the gradient that was used consisted of 100% mobile phase for 0 to 11 min, 85% methanol for 11 to 13 min, and 100% mobile phase for 13 to 24 min. The flow rate was 1.0 mL/min, and UV detection (204 nm) was conducted to monitor the eluants. The retention times for nirvanol and for phenobarbital were 12 min and 13.5 min, respectively. The rate of nirvanol formation was expressed as pmol/min per milligram of microsomal protein. The lower limit of quantitation was 20 pmol nirvanol/mL incubation and the upper limit of quantitation was 2000 pmol nirvanol /mL incubation.

Quantitation of mRNA by Real-Time PCR. Total RNA was isolated from human liver samples, and the amount of *CYP2B6* mRNA present was quantified by real-time PCR using TaqMan fluorogenic probes as part of an earlier study (Davila and Strom, *in preparation*). The *CYP2B6* primers were 5'- CACTCATCAGCTCTGTATTCG-3'(sense) and 5'- GTAGACTCTCTCTGCAACATGAG- 3'(antisense). Real-time PCR quantitation of CAR

and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNAs was carried out by using

the QuantiTect SYBR Green PCR kit (QIAGEN, Chatsworth, CA, USA) according to the

manufacturer's instructions. Amplification was done with the ABI PRISM 7900HT

Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). CAR primers

(Chang et al., 2003) (forward 5'-CCAGCTCATCTGTTCATCCA-3' and reverse 5'-

GGTAACTCCAGGTCGGTCAG-3') were used in PCR in which the initial activation step

was conducted at 95°C for 15 min and was followed by 40 cycles in which each cycle

consisted of denaturation at 94°C for 30 s, annealing at 57.5^{o}C for 30 s, and synthesis at 72°C

for 30 s. The relative amount of CAR mRNA in each human liver sample was normalized to

the GAPDH values. Forward and reverse primers used for GAPDH amplification were 5'-

ACCACAGTCCATGCCATCAC-3' and 5'-TCCACCACCCTGTTGCTGTA-3,

respectively. Amplification of GAPDH mRNA consisted of an initial activation step at 95°C

for 15 min and was followed by 40 cycles in which each cycle consisted of denaturation at

94°C for 30 s, annealing at 60°C for 30 s, and synthesis at 72°C for 1 min. The CAR mRNA

and *CYP2B6* mRNAs were each normalized to GAPDH to control for quality of the mRNA.

Although there is always a statistical risk when data are normalized to a single gene, GAPDH

mRNA is the one most frequently used to normalize values of gene expression in liver tissue,

and in one study its expression varied less than that of 18S rRNA in the same set of human

liver samples (K. Thummel, personnel communication).

RT-PCR and Amplification of *CYP2B6* cDNA.   Because the nucleotide sequence of

CYP2B6 is approximately 95% similar to that of *CYP2B7* (Yamano et al., 1989), mismatched

primers were designed to selectively amplify *CYP2B6*. PCR primers were designed with

PRIMER3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi), and their sequence

homology and specificity were checked by using BLASTn (http://www.ncbi.nlm.nih.gov).

To amplify *CYP2B6*, first-strand cDNA was synthesized from 3 to 5 µg of total

RNA from human liver samples according to the manufacturer's instructions (SuperScript

Preamplification System for First-Strand cDNA Synthesis, Invitrogen, Carlsbad, CA, USA).

Input *CYP2B6* cDNA (1 µl) was amplified in a total volume of 50 µl consisting of 5X PCR

buffer with 1.5 mM $MgCl_2$, 50 ng of DNA, 10 pmol of each primer (utrF2 and R2, Table 1),

0.2 mM dNTP (Life Technologies, Gaithersburg, MD, USA), and 2.5 U of *Taq* DNA

polymerase (Expand High-Fidelity PCR system, Boehringer Mannheim, Mannheim,

Germany). After an initial denaturation step at 94 ^{o}C for 1 min, 32 cycles of PCR were

performed; each cycle consisted of denaturation at 94 ^{o}C for 30 s, annealing at 55-57 ^{o}C for

30 s, and synthesis at 72 ^{o}C for 1 min. The last cycle was followed by a final extension step

at 72 ^{o}C for 10 min.

If the amount of first-strand cDNA in the sample was low, 5 and 10 µl of first-strand

cDNA samples were used in the first round of amplification with the primers

CYP2B6/utrF1/R1 (Table 1), and the amplified products then underwent a second round of

nested PCR using the primers CYP2B6/utrF2/R2 (Table 1).

CYP2B6 cDNAs were analyzed by agarose (1%) gel electrophoresis. If a band that

was the expected size of the *CYP2B6* transcript was observed, the amplification products

were incubated with shrimp alkaline phosphatase and exonuclease I (USB Corporation,

Cleveland, OH, USA) for 30 min at 37 ^{o}C to remove unincorporated nucleotides and primers;

1

the enzymes were inactivated by incubation at 80 $^{\text{o}}$C for 15 min before sequencing. If

multiple bands were observed in the agarose gel, individual

sequenced by using the PCR

primers.

To obtain shorter amplicons of *CYP2B6* for further identification of splice variants,

first-round PCR products were amplified by second-round nested PCR in which various

combinations of PCR primers (Table 1) were used. Sequencing was carried out on ABI Prism

3700 Automated Sequencer (PE Applied Biosystems) by using the PCR primers.

PCR Amplification and Sequencing of *CYP2B6* Exons With Flanking Introns. PCR was

designed to amplify the 9 exons of *CYP2B6* and approximately 150 bp of the each intron that

flanked the adjacent exon. PCR of genomic DNA was similar to that of cDNA, except that

the initial denaturation step was at 94 $^{\text{o}}$C for 10 min and the number of cycles was 35.

PCR Amplification and Sequencing of the *CYP2B6* Promoter Region. We used multiple

pairs of primers (CYP2B6/5´UTRF/R, Table 1) in a single reaction to amplify the 2.4-kb

proximal promoter region of *CYP2B6* from genomic DNA. The products were treated with

SAP-exonuclease and sequenced by using 5 pairs of promoter-sequencing primers (Table 1).

The primers CYP2B6/XREMF1/R1 were used to amplify and sequence the region from –7.9

to –8.6 kb; the primers CYP2B6/XREMF2/R2 were used to amplify and sequence the region

from –8.3 to -8.95 kb.

Analysis of *CYP2B6* Sequences The Phred-Phrap-Consed package (University of Washington, Seattle, http://droog.mbt.washington.edu/PolyPhred.html) was used to assemble the sequences. This program automatically detects the presence of heterozygous single nucleotide substitutions using fluorescence-based sequencing of PCR products. To represent a true variant, the SNP had to be present in sequences generated by the forward and reverse primers, and the presence of a SNP in a sample was confirmed by repeating the PCR and sequencing. Sequences were also analyzed with the Lasergene software (DNASTAR, Inc., Madison, WI, USA). The wild-type and variant sequences were analyzed by a splice prediction program (http://www.fruitfly.org/seq_tools/splice.html) to determine whether any identified SNPs might have contributed to aberrant splicing. The promoter sequences were searched by the program TESS (Transcription Element Search System; http://www.cbil.upenn.edu/tess/) to detect the disruption of any putative transcription factor– binding sites. Nuclear hormone receptor binding motifs were detected using the computer program NUBIscan (http://www.nubiscan.unibas.ch) (Podvinec et al., 2002).

Statistical Analysis to Evaluate Possible Genotype-Phenotype Relationships.
Phenotypes were compared between genders and among racial groups with the exact Wilcoxon and exact Krusal-Wallis tests, respectively. Phenotypes were compared between two genotypes and among three genotypes with the exact Wilcoxon and exact Krusal-Wallis tests, respectively. Proc-StatXact for SAS users: statistical software for exact nonparametric inference (1997) (CYTEL Software Corporation, Cambridge, MA) was used for statistical

analysis. When the overall test among three genotypes (homozygous major, heterozygous, homozygous minor) was significant at an alpha level of 0.10 or less, pairwise comparisons were conducted. *P* values were not adjusted for multiple comparisons because we did not want such strict control on the type 1 error rate since this would be expected to miss a potential significant factor or association.

Linkage Disequilibrium. The significance of linkage disequilibrium between pairs of polymorphic sites was assessed by using Fisher's exact test as described (Weir, 1996) to analyze genotypic data.

Determination of Haplotype Frequencies in Different Ethnic Groups. To calculate the haplotype frequencies, we used the six loci for which the maximum amount of genotyping data was available. The haplotype frequencies were estimated on the basis of the expectation–maximization algorithm (Excoffier and Slatkin, 1995). Haplotype frequencies were determined separately for each ethnic group by counting the estimated haplotypes in each ethnic group. The association between haplotypes and the different CYP2B6 phenotypes were evaluated on the basis of a score test from the haplo.score package(http://www.mayo.edu/statgen), which

and which is based on generalized linear models applicable to a number of traits (Schaid et al., 2002) The haplotype frequencies for females and males were also estimated separately and together. Because the current version of the haplo.score package could not cope with a covariate, i.e.sex, we reconstructed individual haplotypes by

PHASE (Stephens et al., 2001) and used Wilcoxon's rank sum test to compare the

phenotypes between group 1 (with at least one haplotype) and group 2 (without the

haplotype).

The content of CYP2B6 mRNA, protein and activity were quantitated in human livers.

Real-time quantitative PCR revealed a large degree of variation in the relative levels of

CYP2B6 mRNA (Table 2, Fig 1a). The mean relative value for 22 females (1016 $\pm$ 2557) was

3.9 times greater than that for males (259 $\pm$ 676). Among females, Hispanics had the highest

mean relative level of *CYP2B6* mRNA. Among males, Caucasians and Hispanics had similar

mean relative values but was greater than the mean relative value for African American males.

Quantitation of CYP2B6 Protein. CYP2B6 protein concentration was determined in

human liver microsomes (Table 2, Fig 1b). Fifty-nine of the 60 human liver samples contained

detectable amounts of CYP2B6 protein (not shown). Females had a mean CYP2B6 protein

level that was 1.6-times higher than the mean CYP2B6 protein level in male livers. Among

females the highest mean CYP2B6 protein levels were found in Hispanics. In contrast,

Hispanic males had a mean level of CYP2B6 protein which was similar to that of Caucasians,

but higher than the mean value for African American males.

Quantitation of the *N*-Demethylase Activity of CYP2B6. It has been conclusively

demonstrated, using specific monoclonal antibodies and inhibitors, that the extent of N-

demethylation of *S*-mephenytoin (i.e., the biotransformation of *S*-mephenytoin to nirvanol) is

a specific indicator of the level of CYP2B6 activity (Ekins et al., 1998; Heyn et al., 1996; Ko

et al., 1998). Therefore, we measured CYP2B6 activity in the 60 samples described in the

preceding section and in samples from 13 African Americans by measuring the amount of

nirvanol formed by the N-demethylation of *S*-mephenytoin (Table 2, Fig 1c). Our HPLC

method did not detect N-demethylation of mephenytoin in one (4.5%) of 22 liver samples

from females and in nine (20%) of 45 samples from males. CYP2B6 enzyme activity in liver

tissue of females was 1.7 times greater than that in liver tissue of males. CYP2B6 activity

was 3.6 and 4.9 times higher in Hispanic females than in Caucasian females and African

American females (P<0.0215 and P<0.0381, respectively). Among the males, Hispanics had a

mean activity level that was lower than the activity in Caucasians and African Americans.

Correlation of Hepatic CYP2B6 mRNA Level, Protein Quantity, and Enzymatic

Activity. A very good positive correlation was seen between CYP2B6 activity level and

CYP2B6 protein quantity in females (r^2=0.88) and in males ($r^2 = 0.73$). Similarly, a good

positive correlation was found between the relative level of *CYP2B6* mRNA and the amount

of CYP2B6 protein in liver samples from females (r^2=0.80), but the positive correlation

between the two in liver samples from males was not as great (r^2= 0.60). Relative values were

also available for CYP1A2, CYP2D6, CYP4A11 (not shown), and CAR mRNAs (all

normalized to the level of GAPDH mRNA, which served as a control for mRNA quality).

CYP2B6 mRNA expression was not correlated with expression of CYP1A2, CYP2D6, or

CYP4A11 mRNA. Because all of the relative quantities of mRNAs were normalized to that

of GAPDH, there was minimal chance that differences in RNA quality among the samples led

to spurious associations. If data for all persons were included in the analysis, the relative level

of *CYP2B6* mRNA was not correlated with that of CAR mRNA (r^2=0.19). However, CAR

and *CYP2B6* mRNAs were highly correlated (r^2=0.76) if we excluded data for two subjects

who were receiving CYP2B6 inducers (dilantin and fluconazole) at the time of tissue donation and for one subject whose mRNA was highly degraded. The relative level of CAR mRNA was higher in females than in males (Fig. 1d).

Alternative Splicing of *CYP2B6* mRNAs. Although alternatively spliced transcripts of *CYP2B6* have been reported, it is unclear whether alternative splicing is an important determinant of polymorphic expression of CYP2B6 in the liver. Therefore, *CYP2B6* exons 1 through 9 were amplified from mRNA isolated from the same 60 human liver samples phenotyped for mRNA and protein expression and the PCR products were analyzed by agarose gel electrophoresis (Fig. 2a). When we conducted one round of amplification of full-length *CYP2B6* cDNA generated from *CYP2B6* mRNA present in high, intermediate, and low levels in representative liver samples, the *CYP2B6* cDNA of the expected size (~1500 bp) was amplified primarily in those samples with the highest level of *CYP2B6* mRNA (Fig. 2b). *CYP2B6* mRNA was more readily detected in the liver samples from females than in those from males. Notably, although only five samples from Hispanics were among the 60 samples analyzed, samples from two of the three female Hispanics and from one of the two male Hispanics had the highest level of *CYP2B6* mRNA (Fig. 2a).

To detect *CYP2B6* mRNA in samples that contained the lowest amounts of *CYP2B6* mRNA (as indicated by real-time PCR), the amount of template cDNA was increased for first-round amplification, and a second round of amplification with nested primers was performed. Under these robust conditions, the expected *CYP2B6* cDNA was found only in those samples with the highest level of *CYP2B6* mRNA, whereas several smaller cDNAs (the

chief variant was 1000 bp long) were seen in many of the samples (Fig. 2b). The level of variant cDNAs increased as the level of *CY2B6* cDNA representing properly spliced *CYP2B6* mRNA decreased.

To obtain shorter *CYP2B6* transcripts that could be more readily characterized and to avoid amplifying previously reported *CYP2B6* splice variants in which portions of exon 8 were deleted (Miles et al., 1989), we used primers identical to sequences in exons 2 and 7 to amplify the *CYP2B6* cDNA. Apart from the cDNA of the expected size, a variant cDNA of approximately 500 bp (splice variant 1, SV1) was observed (Fig. 3a). This variant lacked sequences corresponding to *CYP2B6* exons 4, 5, and 6. Sequencing of intron 3 and exon 4 in genomic DNA from the liver samples that expressed SV1 revealed several SNPs. The exon-4 genotype 15631G>T was most frequently associated with SV1 (Fig. 3a). The SNP in intron 3 was not independently related to SV1 production, but the ratio of SV1 to full-length cDNA was greater in persons carrying the combination of the exon-4 (15631G>T) and intron-3 (15582C>T) genotypes than in those with only one of the two SNPs (Fig. 3a).

Amplification was done to check for the presence of a previously reported SV that includes 44 nucleotides from intron 3 (exon 3B) but lacks the first 29 nucleotides of exon 4 (Yamano et al., 1989; Miles et al., 1990). We found four SVs (SV2-5) (Fig. 3b), several of which had not been reported previously. These cDNAs represented variant mRNAs that included either 44 bp (exon 3A) (SV4, SV5) or 130 bp (exon 3B) (SV2, SV3) of sequence in intron 3. Some cDNAs representing alternatively spliced transcripts contained exon 4 (SV2, SV4), and others contained all of exon 4 except the first 29 bp (SV3, SV5) (Fig. 3c). Because of the presence of premature termination codons, all four variants encoded truncated proteins.

CYP2B6 exons 7 through 9 were also amplified from the same 60 liver cDNAs. Two previously reported SVs (Yamano et al., 1989; Miles et al., 1989) were observed in approximately half of the samples. One of these variants lacked sequences corresponding to part of exon 8 , whereas the other variant lacked sequences for exon 8 but contained a 58-bp insertion (exon 8A) from intron 8 (Fig. 3d). Both variants contained premature termination codons.

Analysis of Splice Sites and Skipping of Exons 4 Through 6. We used an information theory–based model to determine the information content of the splice donor and acceptor site sequences and whether the strength of the splice site (i.e., the R_i value) was related to alternative splicing (Table 3). Unlike other measures of binding strength, the splice site information we used is related to the thermodynamic entropy and therefore the free energy of binding (Schneider, 1997; Rogan et al., 1998; Thompson et al., 2002). One basis for alternative splicing of exons 4 through 6 appears to be that the skipped upstream acceptor site in intron 3 had a lower information content (R_i=7.0 bits) than the ultimate downstream acceptor site intron 6 (R_i=10.6 bits) (Table 3). For the cryptic exons, all of the splice sites (except the exon 8A acceptor site) had nominal R_i values (R_i values greater than the minimum information content required for splicing, i.e., 1.6 bits). Nevertheless, the acceptor site for exon 3A is stronger than the exon 3 acceptor site; this feature is similar to that of the cryptic exons in *CYP3A5* (Rogan et al., 2003). Thus, the higher information content of some of the cryptic splice sites is related to their enhanced use in splicing and to the inclusion of some cryptic exons (e.g., exon 3A) in *CYP2B6* mRNA.

Because the most common SV lacked exons 4 through 6, we further analyzed genomic

DNA for SNPs surrounding exon 4 to see whether they might be predicted to affect splicing.

Our analysis indicated that the exon 4 SNPs led to several differences in splicing. First, the

15631G>T SNP disrupts a cryptic AG splice acceptor site that is used in creating SV3 and

SV5, which lacks the first 29 bp of exon 4. Second, the exon-4 SNP 15631G>T appears to lie

within an exonic splicing enhancer (ESE). ESEs enhance pre-mRNA splicing, are present in

most exons within approximately 30 bp of the exon border, and serve as binding sites for the

serine/arginine–rich (SR) proteins, which are thought to promote exon definition by either

directly recruiting the splicing apparatus or antagonizing the action of nearby silencer

elements (Cartegni et al., 2002). The 15631G>T substitution in exon 4, the most frequently

observed exonic SNP, was located approximately 30 bp from the 5´ end of exon 4. Computer

analysis of the sequence CCA[G/T]TCC, which surrounds this exon-4 SNP indicated that the

sequence is a potential binding site for the SR protein srp40. The threshold value for srp40

binding is 2.67. The wild-type sequence of exon 4 has a score of 2.85, and replacement of G

with T drops the score to 2.19. Thus, the ESE in exon 4 promotes appropriate splicing of

exon 4, but the SNP in this ESE leads to the skipping of exon 4 (Fig. 3a).

Analysis of *CYP2B6* Exons And Their Flanking Introns. A total of seven exonic SNPs

were identified in *CYP2B6* (Tables 4, 5), and five were missense mutations. The most

frequent SNP was the 15631G>T substitution in exon 4 and resulted in the Gln172His

change. Although analysis of DNA from German Caucasians indicated that the Gln172His

substitution was always associated with the non-synonymous SNP (Lys262Arg) in

CYP2B6*6 and *CYP2B6*7* (Lang et al., 2001), we found that the Gln172His replacement was infrequently associated with other non-synonymous SNPs and represented a newly identified allele, *CYP2B6*9*. The previously reported 64C>T and 12740G>C SNPs in exons 1 and 2, respectively, were found to be linked. Both SNPs occurred at frequencies of 9% and 14% in Caucasians and Hispanics, respectively, but were absent in African Americans. A newly identified missense SNP, 13072A>G in exon 3, results in a Lys139Glu change and represented the previously unknown allele *CYP2B6*8*. This SNP was found only in a single Caucasian. In exon 5, two SNPs were identified: a silent 18000C>T and a 18053A>G that resulted in a Lys262Arg replacement. The Lys262Arg allele (*CYP2B6*4*) was found in 4% of Caucasians, 17% of African Americans, and 14% of Hispanics. This finding was in contrast to an earlier report of a frequency of 32.6% in German Caucasians. The last exonic SNP that we observed was the previously reported 25505C>T substitution, which leads to a Arg487Cys change and was found in 13% of Caucasians and 9% of African Americans, but was absent in Hispanics and the single Asian subject.

Analysis of SNPs in *CYP2B6* 5´Flanking Sequences. Because SNPs in DNA regulatory sequences can impact gene expression we resequenced a total of 3.8 kb (the proximal -2.4 kb and a distal 1.2 kb (from –7.9 to –8.95 kb) of the 5´-flanking region. The region resequenced included the proximal CAR-binding site (the phenobarbital-responsive enhance module, PBREM) and the distal one (xenobiotic-responsive enhancer module, XREM) (Wang et al., 2003). We identified 10 SNPs, nine of which were present in at least 1% of subjects in an ethnic group (Fig. 4). No SNPs were discovered in either the PBREM or the XREM. The

most frequent of these SNPs was a -750T>C substitution, which occurred in 48%, 83%, and

79% of Caucasians, African Americans, and Hispanics, respectively (Table 4). The next most

common alleles were those that possessed the –1456T>C and the –2320T>C changes. The

SNP at -2320 lies within a putative HNF4-binding site, whereas the SNPs at -750 and -757

lie within putative binding sites for HNF1 and Sp1, respectively (Fig. 4). The SNP at -8427

(T>C) resides in a glucocorticoid receptor–binding site (T/Cgtgtc). The SNP at -8207 is

immediately downstream of a glucocorticoid receptor–binding site and within a putative

nuclear hormone receptor binding motif (DR2) as predicted by Nubiscan (Podvinec et al.,

2002).

Several SNPs appear to be specific to particular ethnic groups. For example, the SNP

at –1224 and the SNPs in exons 3 and 5 were found only in Caucasians. African Americans

and Hispanics had several SNPs that were not seen in Caucasians (e.g., SNPs at -8427, -1578,

-757). The exon-9 SNP was not present in any of the Hispanics studied.

Association Between *CYP2B6* Genotype and Phenotype. Inclusive recruitment from

racially diverse populations without stratification can confound association studies

because of the variance in allele frequency or the presence or absence of variants in one

population but not in another. Therefore, we determined whether genotypes were associated

with phenotypes in the entire study population and in stratified populations. *CYP2B6*

genotype-phenotype associations were not observed when data from the entire study

population were analyzed. However, when the phenotypes were first stratified for sex alone

or for sex and ethnicity, associations between phenotypes and genotypes were observed

(Table 5). Three SNPs were significantly associated with CYP2B6 expression (Table 6). Among all persons the Arg487Cys non-synonymous change resulted in significantly lower CYP2B6 activity in Caucasian females who were heterozygous or homozygous for the Arg^{487} allele. The SNP in the intron-3 branch site was also associated with lower levels of CYP2B6 activity in female subjects. For example, compared with those homozygous for the reference allele (15582CC) those subjects who were heterozygous (CT) (n=32) and homozygous (TT) (n=5) for the variant had 1.46-fold and 1.85-fold lower levels of CYP2B6 protein, respectively.

Additional SNPs were correlated with decreased CYP2B6 protein levels, but their association did not reach statistical significance. The –750T>C SNP in the putative HNF1-binding site showed a modest effect on CYP2B6 protein levels. Compared with persons homozygous for the –750T>C change, those persons with one or no allele containing this SNP had 1.42-fold and 1.81-fold, respectively, less CYP2B6 protein (Fig. 5). The SNP in exon 1, which resulted in the Arg22Cys substitution, was associated with higher CYP2B6 activity in all females but not in males. A potential confounding factor was that some samples were from persons who may have been undergoing treatment with CYP2B6 inducers at the time that the liver tissue was donated. However, the dose and duration of exposure were not known, and no consistent effect on CYP2B6 expression could be seen among those with potential exposure to CYP2B6 inducers. Finally, we made no adjustments for multiple statistical comparisons because we did not want such strict control on type I error rate because this would increase the chance of missing a potential significant association. An additional confounding factor inherent in our study is that African Americans and Hispanic

Americans are heterogeneous populations because of the genetic admixture with European Americans. Thus, it will be important to verify these results with larger studies designed with sufficient power to confirm these associations.

Linkage Disequilibrium (LD) and Haplotype Structure of *CYP2B6*. In our study, LD referred to the preferential association of two SNPs within a population. Fisher's exact test was used to detect significant pairwise LD between many of the SNPs in *CYP2B6* (Fig. 6). LD was highly significant between more sites in Caucasians than in African Americans or Hispanics (Fig. 6). This result may be due to the lower number of African American subjects or may reflect the fact that a greater number of haplotypes are typically seen in African Americans. A high degree of linkage was observed between the SNP in the HNF4-binding site in the promoter (-2320T>C) and the SNP in the intron-3 branch point (15582C>T).

Using the six most common SNPs and multisite haplotype inference (Table 7), we identified 11 total and 8 main haplotypes. Haplotype refers to a series of SNPs that are linked on a single chromosome. Haplotypes I, III, IV, and VII accounted for 94% of the observed haplotypes in Caucasians. Haplotypes I through VII accounted for 95% of the haplotypes in African Americans. Haplotype III (-2320T/int3 branch point C) was associated with high levels of CYP2B6 protein in Caucasian females (P<0.049), but haplotype IV (-2320C/int 3 brance point T) was associated with low quantities of CYP2B6 protein in Caucasian females (P<0.035). Haplotype VII (exon 9, 1459C, Cys[487]) was significantly associated with low CYP2B6 activity in Caucasian females (P<0.004).

Effect(s) of Non-synonymous SNPs in *CYP2B6* on CYP2B6 Protein Function. One new

way to prioritize those SNPs that might have a functional consequence is to use computer

programs such as SIFT (sorting intolerant from tolerant) (NG and Henikoff, 2002), a

sequence homology–based tool that evaluates whether amino acid changes are in

evolutionarily conserved or unconserved regions and thus predicts whether any of the non-

synonymous SNPs would affect protein function (NG and Henikoff, 2002). Therefore, we

used this program to identify those SNPs that are most likely to affect CYP2B6 protein

function. Our analysis of the amino acid sequences of CYP2B6 and those of 137 other CYP2

proteins predicted that only the exon-3 SNP that results in Lys139Glu would affect the

function of CYP2B6 (Fig. 7). Interestingly, this substitution occurs at an amino acid (Lys)

that is conserved in CYP2B in all species studied so far and is also conserved in members of

the CYP2C family, e.g., human CYP2C9 and CYP2C19, and rabbit CYP2C5.

Many base changes resulted in replacement of a highly conserved amino acid by a

different (but biochemically similar) amino acid present in the same position in another

species (e.g., exon 4 CYP2B6 Q^{172} was changed to H^{172} which is found in monkey, pig, dog

and rabbit CYP2B orthologs at this position). Although none of the amino acid changes were

found in the six putative substrate recognition sites of CYP2B6 (Gotoh, 1992), some

investigators have suggested that mutations outside of the substrate recognition sites can

affect substrate activity (Domanski et al., 1999). The Arg22Cys substitution occurs in the

membrane anchor region, whereas the Gln172His replacement occurs in the amino acid hinge

between the D' and E helices. There are only four amino acids in CYP2B6 that are in regions

that are highly evolutionarily unconserved across species. Amazingly, the Arg487Cys occurs

in one of these evolutionarily unconserved amino acids and was not predicted by the SIFT program to affect protein function. Nevertheless, the SNP responsible for this amino acid change is associated with low levels of CYP2B6 protein and activity (Table 6) (Ariyoshi et al., 2001; Auboeuf et al., 2002). Although recent reports have reported SIFT as the best predictor of SNPs likely to affect function (Leabman et al., 2003) it is important to consider that SIFT analysis provides only one context in which substitutions that are likely to affect CYP2B6 expression can be identified.

We conducted a study to discover SNPs in *CYP2B6* because this enzyme metabolizes a growing number of clinically important medications and because there is considerable variation in its expression in humans; this variation is believed to be due to genetic factors such as SNPs. Although a previous study identified polymorphisms in German Caucasians (Lang et al., 2001) and an Arg487Cys variant has been shown to be associated with low CYP2B6 expression in some persons, the frequency of this variant's allele in the population could not explain the extent of variation in human CYP2B6 expression. In our study the Arg487Cys variant was also associated with low expression in liver tissue, but only in that of females. In addition, differences in CYP2B6 expression in the liver tissues of females was associated with a SNP in the intron-3 branch site (15582C>T) of *CYP2B6* and a SNP in a putative HNF4-binding site (-2320T>C) in the promoter.

Our strategy of SNP discovery included the resequencing of mRNAs from liver samples of subjects from different ethnic groups as a means to discover sequence variation and to simultaneously characterize alternatively spliced mRNAs. Indeed, the completion of the draft sequence of the human genome has led to the idea that alternative splicing is an important mechanism of generating the complexity of proteins seen in humans. Analysis of alternatively spliced mRNAs can also reveal transcripts that are associated with polymorphic expression of genes (e.g., *CYP3A5* (Kuehl et al., 2001)). We found *CYP2B6* cDNAs that represented numerous aberrantly spliced transcripts, several of which had been previously identified (Yamano et al., 1989; Miles et al., 1988; Miles et al., 1989). The main SV, SV1, lacked sequence corresponding to exons 4, 5, and 6 of the *CYP2B6* gene and was found in

many of the samples analyzed. The presence of a SNP in exon 4 and a SNP in intron 3 were correlated with the appearance of SV1, and the intron-3 SNP was highly associated with CYP2B6 phenotype.

The intron 3 SNP appeared to enhance the proportion of mRNA transcripts skipping exons 4-6 in those persons with the exon 4 ESE SNP (Fig. 3a).   RNA splicing depends not only on the donor and acceptor sequences at the exon/intron boundaries, but also on intronic sequences (branch site) located within 40 nt of the AG splice acceptor. An invariant Adenine of the branch site consensus region undergoes nucleolytic attack by the terminal G nucleotide of the splice donor site. The intron 3 15582C>T alters the consensus C that is immediately adjacent 3' of the invariant A in the putative branch site sequence (Fig 8). Previous studies have shown that mutations in many of the branch site nucleotides can abolish splicing or lead to a reduction in the proper splicing and to switching of the pattern of 3' splice site selection even if the conserved A residue is not affected.

It is difficult to know the exact proportion of any of the *CYP2B6* SVs in the pool of CPY2B6 transcripts because most of the SVs contained premature termination codons and therefore would probably undergo nonsense-mediated decay. This possibility could explain the relatively low abundance of these SVs . Several factors may have contributed to the skipping of exons 4 through 6. Splice site selection appeared to be related to both the SNP in the exon-4 ESE and the strength of the interaction between the splice site and the spliceosome, because the skipped intron-3 acceptor site was weaker (i.e., it had a lower R_i value) than the intron-6 acceptor site used in the alternative transcript. Other reports have

shown that there is a preference for stronger acceptor sites in splicing (Thompson et al., 2002; Rogan et al., 1998).

The ability to relate the exon-4 SNP to CYP2B6 phenotype is complicated by the possibility that the exon-4 SNP has pleiotropic effects (Fig. 8). The disruption of the ESE by the SNP in exon 4 leads to the skipping of exon 4 and the generation of an alternatively spliced *CYP2B6* mRNA. This SNP would be predicted to be associated with a smaller pool of appropriately spliced *CYP2B6* mRNA and decreased CYP2B6 activity. However, the exon-4 SNP also disrupts a cryptic acceptor 3' splice site in the reference sequence (*AG*/T). The cryptic 3'-splice acceptor *AG* is used to create some of the SVs that lack the first 29 bp of exon 4 (Fig. 3b). If the exon-4 SNP did not exert pleiotrophic effects, then it would be predicted to decrease the production of these SVs and this decrease would be expected to result in more of the appropriately spliced *CYP2B6* mRNA. The exon-4 SNP, which results in the Gln172His substitution, was recently introduced into *CYP2B6* cDNA, and the protein encoded by this variant allele had higher 7-ethoxycoumarin *O*-deethylase activity than the protein encoded by the reference allele (Ariyoshi et al., 2001; Jinno et al., 2003). These types of cDNA-based assays were unable to determine whether the exon-4 SNP affected splicing of the pre-mRNA. The results of the phenotype analysis in our study population showed that there was no relationship between this exon-4 SNP alone and CYP2B6 activity.

Our ability to identify the functionally important *CYP2B6* genotypes is hampered by the fact that many of the SNPs are in LD and in multiple haplotypes. For example, a high degree of linkage was observed between two SNPs, (e.g., the -2320T>C SNP in the putative HNF4-binding site and the intron-3 branch point SNP). Because multiple haplotypes exist,

the SNPs in LD in one haplotype may or may not be associated with SNPs in another haplotype. It must also be considered that some phenotypes are the sum of regulatory and splicing genotypes, as there is growing evidence that transcription and splicing are coincident. Because transcriptional regulation can be important for splicing choices (Maniatis and Reed, 2002), SNPs in the HNF4-binding site may influence splice-site selection.

It is possible that *CYP2B6* transcription and splicing are coordinately regulated by nuclear hormone receptors and their coactivators (Auboeuf et al., 2002). This possibility is particularly intriguing in light of the recent identification of PGC-1 (peroxisome proliferator–activated receptor 1) as a coactivator of the gene coding for the CAR receptor (a master regulator of CYP2B) (Shiraki et al., 2003)). PGC-1 contains an SR domain (found in the family of splicing factors that bind to ESEs) and an RNA-binding domain, and thus may function to help couple splicing to transcription. PGC-1 colocalizes with splicing factors in the nucleus and interacts with SR proteins. *CYP2B6* promoter–dependent recruitment of regulatory splicing factors, such as PGC-1 , to CAR could influence *CYP2B6* transcription and the selection of splice sites.

Significant differences in CYP2B6 expression in liver were found between sexes. Indeed, CYP2B6 activity was below quantifiable limits in 4.5% of female and in 20% of male liver microsomes. This difference appears to translate to humans in vivo because a difference between sexes in metabolism of the anticancer drug ifosfamide, a CYP2B6 substrate, has been reported (Schmidt et al., 2001). The relative level of CAR mRNA, a master regulator of CYP2B (Wei et al., 2000), is higher in females than in males (Fig. 1), and could directly affect

differences between the sexes in the rate of *CYP2B6* transcription. In rodents, sexually dimorphic expression of hepatic CYP2B is regulated by the pattern of growth hormone secretion (Teglund et al., 1998). In humans the patterns of growth hormone secretion are sexually dimorphic (Jaffe et al., 1998), and growth hormone secretion is a regulator of CYP expression (Jaffe et al., 2002). However, whether the differences between the sexes in growth hormone secretion in humans regulate either CAR or CYP2B expression remains to be determined. CYP2B6 is the main enzyme involved in the hydroxylation of efavirenz (an inhibitor of HIV-1 reverse transcriptase) (Ward et al., 2003) and in the bioactivation of ifosfamide (Schmidt et al., 2001). Therefore, the effectiveness of efavirenz and ifosfamide may be lower and higher, respectively, in females (especially Hispanic females) than in males.

It was important to look for relationships between *CYP2B6* genotype and phenotype not just in the combined study population but also in groups stratified on the basis of sex and of ethnicity, as other have done (Stengard et al., 2002). This stratification according to sex and ethnicity was crucial because our results clearly demonstrated that liver tissue of females express significantly higher amounts of CYP2B6 than do liver tissue of males, that CYP2B6 expression varied among the different ethnic groups, and that the frequencies of the SNPs and haplotypes differed among ethnic groups. The reason that that we detected a genotype-phenotype association in females but not in males may be due to the high level of CYP2B6 expression in females and the low level in males. However, the CYP2B6 variability may depend on a combination of SNPs that differ in each ethnic group, and that a particular SNP may depend on another gene product (e.g., CAR) or environment (e.g., sex hormones) to exert its effect. Our results also suggest that no single SNPs are associated with CYP2B6 activity

in both sexes. Therefore, we should not necessarily be asking which SNPs are absolute

markers of CYP2B6 phenotype, but rather what are the genetic variants associated with

CYP2B6 expression in each ethnic population stratified by sex.

References

Ariyoshi N, Miyazaki M, Toide K, Sawamura Y and Kamataki T (2001) A single nucleotide polymorphism in CYP2B6 found in Japanese enhances catalytic activity by autoactivation. *Biochem Biophys Res Commun* **281**: 1256-1260.

Auboeuf D, Honig A, Berget S M and O'Malley B W (2002) Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science* **298**: 416-419.

Cartegni L, Chew S L and Krainer A R (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**: 285-298.

Chang T K H, Bandiera S M and Chen J (2003) Constitutive androstane receptor and pregnane X receptor gene expression in human liver: interinidvidual variability and correlation with CYP2B6 mRNA levels. *Drug Metab Dispos* **31**: 7-10.

Domanski T L, Schultz K M, Roussel F, Stevens J C and Halpert J R (1999) Structure-function analysis of human cytochrome p-450 2B6 using a novel substrate, side-directed mutagenesis, and molecular modeling. *J Pharmacol Exp Ther* **290**: 1141-1147.

Ekins S, Vandenbranden M, Ring B J and Wrighton S A (1997) Examination of purported probes of human CYP2B6. *Pharmacogenetics* **7**: 165-179.

Ekins S, Vandenbranden M, Ring BJ, Gillespie JS, Yang TJ, Gelboin H V and Wrighton SA (1998) Further characterization of the expression in liver and catalytic activity of CYP2B6. *J Pharmacol Exp Ther* **286**: 1253-1259.

Ekins S, Bravi G, Ring BJ, Gillespie TA, Gillespie JS, Vandenbranden M, Wrighton SA and Wikel JH (1999) Three-dimensional quantitative structure activity relationship analyses of substrates for CYP2B6. *J Pharmacol Exp Ther* **288**: 21-29.

Excoffier L and Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* **12**: 921-927.

Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **267**: 83-90.

Heyn H, White RB and Stevens JC (1996) Catalytic role of cytochrome P4502B6 in the N-demethylation of S-mephenytoin. *Drug Metab Dispos* **24**: 948-954.

Jaffe CA, Ocampo-Lim B, Guo W, Krueger K, Sugahara I, Demott-Fribert R, Bermann M and Barkan AL (1998) Regulatory mechanisms of growth hormone secretion are sexually dimorphic. *J Clin Invest* **102**: 153-164.

Jaffe CA, Turgeon DK, Lown K, Demott-Fribert R and Watkins PB (2002) Growth hormone secretion pattern is an independent regulator of growth hormone actions in humans. *Am J Physiol Endocrinol Metab* **283**: E1008-E1015.

Jinno H, Tanaka-Kagawa T, Ohno A, Makino T, Matsushima E, Hanioka N and Ando M (2003) Functional characterization of cytochrome P450 2B6 allelic variants. *Drug Metab Dispos* **31**: 398-403.

Ko J W, Desta Z and Flockhart DA (1998) Human N-demethylation of (S)-mephenytoin by cytochrome P-450s 2C9 and 2B6. *Drug Metab Dispos* **26**: 775-778.

Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, Watkins PB, Daly A, Wrighton SA, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Venkataramanan R, Strom S, Thummel K, Boguski M S and Schuetz E (2001) Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nature Genet* **27**: 383-391.

Lang T, Klein K, Fischer J, Nussler AK, Neuhaus P, Hofmann U, Eichelbaum M, Schwab M and Zanger UM (2001) Extensive genetic polymorphism in the human CYP2B6 gene with impact on expression and function in human liver. *Pharmacogenetics* **11**: 399-415.

Leabman MK, Huang c C, DeYoung J, Carlson EJ, Taylor TR, de la Cruz M, Johns SJ, Stryke D, Kawamoto M, Urban TJ, Droetz DL, Ferrin TE, Clark AG, Risch N, Herskowitz I and Giacomini K M (2003) Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc Natl Acad Sci U S A* **100**: 5896-5901.

Maniatis T and Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* **416**: 499-506.

Miles JS, Spurr NK, Gough AC, Jowett T, McLaren AW, Brook JD and Wolf CR (1988) Novel human cytochrome P450 gene (P450IIB): chromosomal localization and evidence for alternative splicing. *Nucleic Acids Res* **16**: 5783-5795.

Miles JS, McLaren AW and Wolf CR (1989) Alternative splicing in the human cytochrome P450IIB6 gene generates a high level of aberrant messages. *Nucleic Acids Res* **17**: 8241-8255.

Miles JS, McLaren AW, Gonzalez FJ and Wolf CR (1990) Alternative splicing in the human cytochrome P450IIB6 gene: use of a cryptic exon within intron 3 and splice acceptor site within exon 4. *Nucleic Acids Res* **18**: 189.

NG PC and Henikoff S (2002) Accounting for human polymorphisms predicted to affect protein function. *Genome Research* **12**: 436-446.

Podvinec M, Kaufmann MR, Handschin C and Meyer UA (2002) NUBIScan, an in silico approach for prediction of nuclear receptor response elements. *Mol Endocrinol* **16**: 1269-1279.

Rogan PK, Faux BM and Schneider TD (1998) Information analysis of human splice site mutations. *Human Mutation* **12**: 153-171.

Rogan P K, Svojanovsky S and Leeder J S (2003) Information theory-based analysis of CYP2C19, CYP2D6 and CYP3A5 splicing mutations. *Pharmacogenetics* **13**: 207-218.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM and Poland GA (2002) Score tests for association of traits with haplotypes when linkage phase is ambiguous. *Am J Hum Genet* **70**: 425-434.

Schmidt R, Baumann F, Hanschmann H, Geissler F and Preiss R (2001) Gender difference in ifosfamide metabolism by human liver microsomes. *Eur J Drug Metab Pharmacokinet* **26**: 193-200.

Schneider TD (1997) Information content of inidividual genetic sequences. *J Theor Biol* **189**: 427-441.

Shiraki T, Sakai N, Kanaya E and Jingami H (2003) Activation of orphan nuclear constitutive androstane receptor requires subnuclear targeting by peroxisome proliferator-activated receptor coactivator-1 . *J Biol Chem* **278**: 11344-11350.

Stengard JH, Clark AG, Weiss KM, Kardia S, Nickerson DA, Salomaa V, Ehnholm C, Boerwinkle E and Sing CF (2002) Contributions of 18 additional DNA sequence variations in the gene encoding apolipoprotein E to explaining variation in quantitative measures of lipid metabolism. *Am J Hum Genet* **71**: 501-517.

Stephens M, Smith N and Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978-989.

Sueyoshi T, Kawamoto T, Zelko I, Honkakoski P and Negishi M (1999) The repressed nuclear receptor CAR responds to phenobarbital in activating the Human CYP2B6 Gene. *J Biol Chem* **274**: 6043-6046.

Teglund S, McKay C, Schuetz E, van Deursen J M, Stravopodis D, Wang D, Brown M, Bodner S, Grosveld G and Ihle JN (1998) Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell* **93**: 841-850.

Thompson TE, Rogan PK, Risinger JI and Taylor JA (2002) Splice variants but not mutations of DNA polymerase are common in bladder cancer. *Cancer Res* **62**: 3251-3256.

Wang H, Faucette S, Sueyoshi T, Moore R, Ferguson s, Negishi M and LeCluyse EL (2003)
A novel distal enhancer module regulated by pregnane X receptor/constitutive androstane
receptor is essential for the maximal induction of CYP2B6 gene expression. *J Biol Chem* **278**:
14146-14152.

Ward BA, Gorski CJ, Jones DR, Hall SD, Flockhart DA and Desta Z (2003) The cytochrome
P450 2B6 is the main catalyst of efavirenz primary and secondary metabolism: implication
for HIV/AIDs therapy and utility of efavirenz as substraet marker of CYP2B6 catalytic
activity. *J Pharmacol Exp Ther* **306**: 287-300.

Wei P, Zhang J, Egan-Hafley M, Liang S and Moore DD (2000) The nuclear receptor CAR
mediates specific xenobiotic induction of drug metabolism. *Nature* **407**: 920-923.

Weir BS. Genetic data analysis II, Sinauer Associates, Sunderland, MA, 1996.

Yamano S, Nhamburo P T, Aoyama T, Meyer UA, Inaba T, Kalow W, Gelboin H V,
McBride O W and Gonzalez F (1989) cDNA cloning and sequence and cDNA-directed
expression of human P450IIB1: Identification of a normal and two variant cDNAs derived
from the CYP2B locus on chromosome 19 and differential expresion of the IIB mRNAs in
human liver. *Biochem* **28**: 7340-7348.

Unnumbered **Footnote to the Title**

Legends for Figures

Fig. 1. Differences between sexes and ethnic groups in CYP2B6 and CAR expression. *CYP2B6* mRNA, protein, and activity levels and *CAR* mRNA expression were determined in liver tissue of Caucasian, Hispanic, and African American female and male subjects. The results are presented as mean $\pm$ SD, with n indicating the numbers of subjects. The relative levels of *CYP2B6* and *CAR* mRNA have been normalized to the level of GAPDH mRNA in the same liver. The combined populations (solid bars) also included females and males of unknown ethnicity.

Fig. 2. Amplification of human liver *CYP2B6* cDNA. Total RNA isolated from human livers was used to amplify full-length *CYP2B6* cDNA, which was subsequently analyzed on agarose gels. (a) Female (F) and male (M) *CYP2B6* cDNAs were arranged from highest to lowest relative levels of *CYP2B6* mRNA (left to right). Samples from Hispanics are indicated (*). (b) The cDNA templates (1, 5, or 10 µl) generated from representative samples having high, intermediate, and low relative levels of *CYP2B6* mRNA underwent one or two rounds of nested PCR using the indicated primers (bottom and top right). The wild-type *CYP2B6* mRNA is approximately 1500 bp, and multiple lower molecular-weight splice variants are present. M=molecular weight standards. The number for each representative sample is indicated above the appropriate lane.

Fig. 3. *CYP2B6* splice variants. (a) Effect of the SNP in the intron-3 branch site and in exon 4 on the skipping of exons 4, 5, and 6 in females. Exons 2 through 7 of *CYP2B6* were amplified, and the products were analyzed on agarose gels. Splice variant 1 (SV1) lacks sequence corresponding to exons 4 through 6. The absence of the intron-3 branch site SNP or the exon-4 SNP (wild type) is indicated by the open boxes, the presence of one allele containing one of these SNPs (heterozygous) is indicated by the half-filled boxes, and the presence of two alleles containing the same type of SNP (homozygous) variant is indicated by the filled boxes. (b) Amplification of *CYP2B6* variants containing intron-3 insertions was done by using nested primers (arrows point to the bands representing the different SVs) and by using the product from the exon 2-7 amplification as template. SV2 and SV3 contained exon 3B (130 bp from intron 3). SV4 and SV5 contained exon 3A (44 bp from intron 3). SV2 and SV4 contained exon 4, whereas SV3 and SV5 lacked the first 29 bp of exon 4. (c) Diagrammatic illustration of SV2 through SV5. (d) Diagrammatic illustration of SVs in which either exon 8A was inserted and exon 8 was skipped or exons 8A and 8 were absent.

Fig 4. Schematic representation of the location of SNPs in the 5'-flanking region of *CYP2B6*. Indicated are the locations of SNPs in putative transcription factor–binding sites in relation to the positions of PBREM, XREM, and additional putative nuclear hormone receptor–binding sites. The boldface arrow indicates the direction of transcription and its initiation site (+1).

Fig. 5. CYP2B6 protein levels in the liver as a function of HNF1 genotype in females and males. Results are shown as the mean $\pm$ SD. Open and filled boxes represent persons

homozygous for the reference and variant alleles, respectively, and partially shaded boxes represent heterozygotes.

Fig. 6. Distribution of linkage disequilibrium of SNPs in the *CYP2B6* gene of Caucasians, African Americans, and Hispanics. The position of SNPs occurring in the *CYP2B6* gene indicated in Exons and Introns (int) are as described numerically in Table 4. The *p* values less than 0.001 are denoted by the black box; those between 0.001 and 0.01, the dark gray box; those between 0.01 and 0.05, the intermediate gray box; and those greater than 0.05, by the lightest box.

Fig. 7. Alignment of a partial amino acid sequence of human CYP2B6 with those of CYP2B homologs from other species. The CLUSTAL method was used for alignment. Non-synonymous amino acid changes are indicated above the aligned sequences. Amino acids with similar chemical properties are grouped by color, and dashes denote gaps.

Fig. 8. Schematic illustration of the location of the *CYP2B6* intron-3 and exon-4 SNPs relative to the spliceosome. The three possible functional consequences of the exon-4 SNP are indicated. The splice site donor GT, the acceptor AG and *CYP2B6* branch site sequence (including the invariant Adenine (*A*) and the intron 3 branch site 15582C>T SNP) are indicated. The three possible functional consequences of the exon 4 15631 G>T SNP (ESE disruption by SNP; non-synonymous $Q^{172}H$ change; disruption of the cryptic AG splice acceptor) are indicated.

a.

b.

c.

d.

Caucasians
Hispanics
African Americans
Combined

Females

Males

a.

F

M

M

* = hispanic

b.

CYP2B6 Full Length Amplification

a. exon 2-7

intron 3 br. pt. C>T
exon 4 G>T

CYP2B6
SV1

C>T
G>T

2 3 4 5 6 7

2 3 7   SV1

b. exon 3-4

SV2
SV3
SV4
SV5

c. SV2-SV5

3   130bp 3B   29bp 4

+29
−29

3 3B 4   SV2
3 3B 4   SV3

3   44bp 3A   29bp 4

+29
−29

3 3A 4   SV4
3 3A 4   SV5

d. exon 7-9

7   A>G   8   C>T   58 bp 8A   A>G   9

166   52   105   100

7 8A 9

6 7 9

-1578C/T

-1186C/G

-8427T/C

-1224A/G

-757C/T

-8207G/C

-2320T/C

-1778A/G

-750T/C

-1456T/C

-10Kb   XREM   ?   -7.5Kb -2.5Kb   PBREM   +1

GRE

HNF4 site

Sp1 site

HNF1 site

-8597

ACAAC**ACAGCA**GGAGAC**AGGTCC**TG**GGGTCAGG**AA**AGTCCA**TGGTGGCAT

NR5   NR4   NR3

NR8

CTTTCCT**AGGTCA**GGAAAG**TGTGGG**T**TGTAC**TTTGT**TGGTCC**TCTCCTCGTG

NR6   NR7

-8495

-8207G/C

ER6

DR2

AGGGAC**TGACCC**CTGCCA**GGGGCA**GGG**AGGGGA**CA**GGCTCA**CAGCAGT

-8242   -8195

DR3

DR5

Caucasians African Americans Hispanics

- ■ <0.001
- ▨ 0.001~0.01
- ▨ 0.01~0.05
- ▢ >0.05

	Ex 1 R22C	Ex 3 K139E	Ex 4 Q172H	Ex 5 K262R	Ex 9 R487C
hum2b6	LVQ-RHPNT	FCMGKRSVE	FLFQSITA	DPSAPDLID	TYQIRFLPR
macaque2B17	LVQ-RHPNA	FCMGKRSVE	FLFHSITA	DPSAPQDLID	TYQIRFLPR
pig2b22	WG----HPSS	FCMGKRSVE	VYFHSSTA	DPSLPRDLID	PYQIQFLPR
dog2b11	MAR-GHPKA	FCMGKRSVE	FFFHSNTA	DPSAPRDFID	VYQISFLSR
rabbit2b4	LFR-GHPKA	FCMGKRSVE	LLFHSITS	DPSMPRDFID	SYQIRFLAR
rabbit2b5	LFR-GHPKA	FCMGKRSVE	LLFHSVTA	DPSNPRDFID	SYQIRFLAR
mus2b9	LVR-GHAKI	FCMGKRSVE	FLFHSVTA	DPSAPRDVID	ANQIYFLAR
mus2b10	LAR-GHPKS	FCMGKRSVE	FLFQCITA	DPSVPRDFID	TYQICFLAR
mus2b10ike	LAR-GHPKS	FCMGKRSVE	FLFQCITA	DPSVPRDFID	TYQICFLAR
mus2B13	LVR-GHAKI	FCMGKRSVE	FLFQCITA	DPSBPRDFID	TYQICFLAR
mus2b19	LVSQSQPKT	FCMGKRSVE	FEFQCVTA	DPSWPRDFID	QYQIHFLSR
mus2h20	LAR-GHPKS	FCMGKRSVE	FYFQCIVA	DPSWPRDFID	EFE
rat2b1	LVR-GHPKS	FCMGKRSVE	FLFQCITA	DPSAPRDFID	TYQICFSAR
rat2b2	LVR-GHPKS	FCMGKRSVE	FLFQCITA	DPSAPRDFID	TYQICFSAR
rat2b3	LVR-GHAKV	FCIGHQSVE	FLFQCVTS	DPNAPRDFID	VYKICFVAR
rat2b12	LVSQSQPKT	FCMGNRSVD	FLFQSIAA	DPSAPRDFIN	PYQINFLSR
rat2b15	LASQNRPKT	FCMGKRSVE	SLFQSIAA	DPNTPRDFIN	PYQIHFLSR
rat2boes	LIR-GQGNG	FCMGKRSVE	FHLQCITA	DPSTPRDFID	TYQICFLSR

■ Identical □ Conservative ■ Block of Similar ■ Weakly Similar □ Non Similar

ESE SCORE

CCATTCC 2.19 MUTANT

2.67 Threshold for srp40

CCAGTCC 2.85 WT

Consensus
Branch site

EX 3 GT

TNCTAAC

CYP2B6: CTTGA (C/T)CT

AG EX 4

$Q^{172}H$

AG = Cryptic splice site
29 bp deletion, SV3,5

T = loss of cryptic splice acceptor

			A (°C)	C	E
			3	3	4
			8	8	0
			8	8	1
			7	8	1
			8	8	1
			5	8	1

			A (°C)	C	E
			6	5	1
			6	5	1
			5	5	1
			6	5	1
			6	5	1
			8	5	1
			6	5	1
			7	5	1
			6	5	1
			8	5	1
			5	5	3
			5	5	1
			6	5	1

Gender	Label	N[d]	Median	Minimum	Maximum	Mean	Std Error
Female	Activity[a]	28	4.50	0.00	42.52	7.05	1.60
	Protein[b]	22	28.66	0.00	129.67	35.00	5.90
	mRNA[c]	22	195.37	1.07	12161.22	1016.47	545.28
Male	Activity	45	3.08	0.00	17.75	4.22	0.67
	Protein	38	14.69	5.43	58.11	21.25	2.52
	mRNA	38	25.55	1.36	3590.58	259.98	109.69
ALL PERSONS	**Label**	**N**	**Median**	**Minimum**	**Maximum**	**Mean**	**Std Error**
Black	Activity	19	4.25	0.00	14.41	4.60	1.00
	Protein	6	12.24	0.00	19.13	10.71	2.78
	mRNA	6	21.42	3.78	424.61	87.25	67.64
Hispanic	Activity	7	6.89	0.00	42.52	10.70	5.60
	Protein	7	27.81	7.54	129.67	43.91	16.38
	mRNA	7	270.60	1.36	12161.22	2052.72	1690.53
Caucasian	Activity	43	3.32	0.00	23.42	4.94	0.77
	Protein	43	20.47	5.43	66.16	26.33	2.60
	mRNA	43	47.50	1.07	3590.58	384.64	113.35
MALES	**Label**	**N**	**Median**	**Minimum**	**Maximum**	**Mean**	**Std Error**
Black	Activity	12	3.55	0.00	14.41	4.76	1.40
	Protein	5	12.30	5.74	19.13	12.85	2.18
	mRNA	5	17.03	3.78	424.61	97.03	81.97
Hispanic	Activity	4	1.66	0.00	6.89	2.55	1.64
	Protein	4	18.71	7.54	40.69	21.41	7.88
	mRNA	4	138.51	1.36	770.69	262.27	180.76
Caucasian	Activity	28	3.15	0.00	17.75	4.35	0.87
	Protein	28	14.98	5.43	58.11	22.85	3.18
	mRNA	28	25.64	2.17	3590.58	296.22	146.29
FEMALES	**Label**	**N**	**Median**	**Minimum**	**Maximum**	**Mean**	**Std Error**
Black	Activity	7	4.70	0.00	9.55	4.34	1.40
	Protein	1	0.00	0.00	0.00	0.00	.
	mRNA	1	38.32	38.32	38.32	38.32	.
Hispanic	Activity	3	12.90	9.24	42.52	21.55	10.54
	Protein	3	69.86	22.18	129.67	73.90	31.10
	mRNA	3	962.07	196.72	12161.22	4440.00	3866.93
Caucasian	Activity	15	4.11	1.02	23.42	6.05	1.49
	Protein	15	28.86	8.13	66.16	32.84	4.13
	mRNA	15	194.01	1.07	2385.37	549.70	174.55

[a] Nirvanol Formation (nmol/min/mg protein)
[b] pmol CYP2B6 protein/mg microsomal protein
[c] relative CYP2B6 mRNA (arbitrary units)
[d] number of persons

Table 3. Relative strengths of exon acceptor and donor sites in appropriately and alternatively spliced exons

Acceptor R_i value (bits)	Exon #	Exon region (Acc# NG_000008)	Length (bp)	Donor R_i value (bits)
	1	214385 - 214555	171	7.1 @ 214556
-3.6 @ 227079	2	227080 - 227242	163	7.2 @ 227243
5.2 @ 227375	3	227376 - 227525	150	7.8 @ 227526
5.7 @ 229080	3A	229082 - 229125	44	4.5 @ 229128
5.7 @ 229080	3B	229082 - 229211	130	6.4 @ 229212
7.0 @ 229983	4	229984 - 230144	161	5.9 @ 230145
6.4 @ 230015	4 (alternate acceptor)	229984 - 230012	29	
10.3 @ 232297	5	232298 - 232474	177	3.5 @ 232475
3.1 @ 233072	6	233073 - 233214	142	6.0 @ 233215
10.6 @ 235376	7	235377 - 235564	188	9.6 @ 235565
8.5 @ 235752	8	235753 - 235894	142	8.0 @ 235895
-0.1 @236051	8A	236052 - 236109	58	4.4 @ 236110
5.2 @ 239724	9	239725 - 239906	182	

Table 4. Frequency of *CYP2B6* Alleles

Allele frequency (% individuals with at least one variant allele)[a]

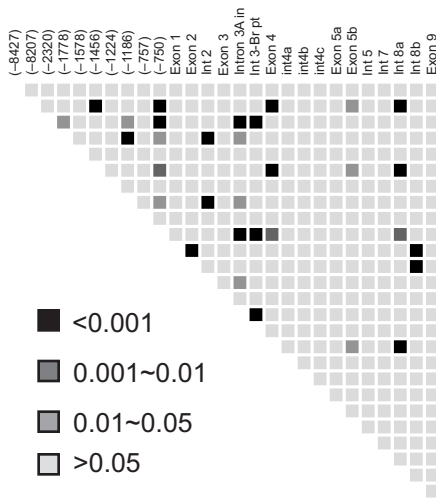| Position in Reference Sequence	Position in Gene	Position in mRNA	Sequence	Region	Caucasians	African Americans	Hispanics	Asians	Lang et al., 2001: Caucasians	Corresponding *CYP2B6* Allele Nomenclature
NG_000008	NG_000008[b]	NM_000767[b]			(n=86)		(n=14)	(n=2)		
205958	-8427		ttgcc(T/C)gtgtc	5'UTR	0 (0%)	[d] 0.22 (31.03%)	0.14 (28.57%)	0 (0%)		
206178	-8207		ggaca(G/C)gctca	5'UTR	0.21 (39.54%)	[d] 0.33 (51.72%)	0.43 (71.43%)	0 (0%)		
212065	-2320		cgata(T/C)ctggc	5'UTR(HNF4)	0.27 (51.2%)	[d] 0.17 (31.03%)	0.21 (28.57%)	1 (100%)		*CYP2B6*1B*
212607	-1778		agcac(A/G)gcaag	5'UTR	0.07 (13.95%)	[c] 0 (0%)	0.14 (28.57%)	1 (100%)		
212807	-1578		tcagg(C/T)gt agg	5'UTR	0 (0%)	[c] 0.25 (33.33%)	0.14 (28.57%)	0 (0%)		
212929	-1456		tccag(T/C) agaca	5'UTR	0.22 (41.9%)	[c] 0.5 (83.33%)	0.43 (71.43%)	0 (0%)		
213161	-1224		tgacc(A/G)cctct	5'UTR	0.01 (2.33%)	[c] 0 (0%)	0 (0%)	0 (0%)		*CYP2B6*1F*
213199	-1186		tcttg(C/G) tcctc	5'UTR	0.07 (13.95%)	[c] 0 (0%)	0.14 (28.57%)	1 (100%)		
213628	-757		catca(C/T)gcccg	5'UTR(Sp1)	0 (0%)	[c] 0.25 (33.33%)	0.07 (14.29%)	0 (0%)		
213635	-750		cccgg(T/C)taatt	5'UTR(HNF1)	0.48 (74.42%)	[c] 0.83 (100%)	0.79 (85.71%)	1 (100%)		*CYP2B6*1G*
214448	64	64	ttcag(C/T)gccac	Exon 1 (R22C)	0.09 (18.6%)	[c] 0 (0%)	0.14 (28.6%)	0 (0%)	5.3%	*CYP2B6*2A*
227124	12740	216	ggacc(G/C)aggcc	Exon 2	0.09 (18.6%)	[c] 0 (0%)	0.14 (28.6%)	0 (0%)	7.1%	
227301	12917		ggagt(A/T)tatgg	Intron 2	0.08 (16.3%)	[c] 0 (0%)	0.14 (28.6%)	1 (100%)		
227456	13072	415	tggga(A/G)agcgg	Exon 3 (K139E)	0.01 (2.3%)	[c] 0 (0%)	0 (0%)	0 (0%)		*CYP2B6*8*
228977	14593		agaga(C/G)agaca	Intron 3	0.27 (51.2%)	[c] 0.08 (16.7%)	0.21 (28.6%)	1 (100%)		
229966	15582		cttga(C/T)ctgct	Intron 3	0.27 (48.8%)	[d] 0.19 (34.5%)	0.21 (28.7%)	1 (100%)		
230015	15631	516	ttcca(G/T)tccat	Exon 4 (Q172H)	0.22 (41.9%)	[d] 0.28 (48.3%)	0.43 (71.4%)	0 (0%)	28.6% (38.6)[e]	*CYP2B6*9*
232054	17670		gaaag(G/A)agaga	Intron 4	0 (0%)	[c] 0.33 (33.3%)	0.14 (14.3%)	0 (0%)		
232139	17755		agata(G/C)atcaa	Intron 4	0 (0%)	[c] 0.33 (33.3%)	0.14 (14.3%)	0 (0%)		
232281	17897		cacca(C/T)ccctt	Intron 4	0 (0%)	[c] 0.33 (33.3%)	0.14 (14.3%)	0 (0%)		
232384	18000	732	gctta(C/T)attgg	Exon 5	0.01 (2.3%)	[c] 0 (0%)	0 (0%)	0 (0%)		
232437	18053	785	cccca(A/G)ggacc	Exon 5 (K262R)	0.04 (4.7%)	[c] 0.17 (16.7%)	0.14 (14.3%)	0 (0%)	32.6%	*CYP2B6*4*
233011	18627		tgggc(G/A)tatac	Intron 5	0 (0%)	[c] 0.17 (33.3%)	0.07 (14.3%)	0 (0%)		
235587	21203		cccat(A/G)gccct	Intron 7	0.05 (9.3%)	[c] 0 (0%)	0.07 (14.3%)	0 (0%)		
235947	21563		tatcc(C/T)caact	Intron 8	0.21 (39.5%)	[d] 0.50 (83.3%)	0.43 (71.4%)	0 (0%)		
239625	25241		tgcaa(A/G)tctgt	Intron 8	0.09 (18.6%)	[d] 0.02 (3.5%)	0.14 (28.6%)	0 (0%)		
239889	25505	1459	agatc(C/T)gcttc	Exon 9 (R487C)	0.13 (25.6%)	[d] 0.09 (17.2%)	0 (0%)	0 (0%)	14%	*CYP2B6*5*

[a] Alleles deposited: http://www.imm.ki.se/CYPalleles/
[b] Position is with respect to the translation start site of the *CYP2B6* gene; the A in ATG is +1 and the base immediately 5' is –1.
[c, d] Number of alleles screened : [c] 12, [d] 58.
[e] Ariyoshi et al., 2001: G516T had a frequency of 0.20 (38.6%) in a Japanese population.

HL#	Race	A	-8427	-8207	-2320	-1778	-1578	-1456	-1224	-1186	-757	-750	64	12740	12917	13072	14593	15582	15631	17670	17755	17897	18000	18053	18627	21203	21563	25241	25505	Act[a]	Pr[b]	mRNA[c]	CmRNA[d]
604[e]	H	6	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CT	CG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	AG	CC	AG	CC	42.52	130	12161.2	95.51
700	C	3	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CT	CG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AG	CC	23.42	66.16	393.44	15.22
636	H	5	TC	GC	TT	AA	CT	TC	AA	CC	CC	CC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	12.9	69.86	962.07	56.7
650[f]	C	9	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CT	CG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AG	CC	12.71	28.55	2385.37	164.35
836	AA	5	TT	GG	TC													CC	GG										CT	9.55			
746[g]	H	0	TT	CC	TT	AA	CC	CC	AA	CC	CC	CC	CC	GG	AA	AA	CC	CC	TT	GG	GG	CC	CC	GG	GG	GG	TT	AA	CC	9.24	22.18	196.72	75.36
864	C	8	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	9.24	46.37	101.83	61.34
629[g]	C	6	TT	GC	TC	AG	CC	TC	AA	CG	CC	CC	CC	GG	AT	AA	CG	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	8.23	38.85	803.41	77.02
35	AA	1	TC	GG	TT													CC	GG										CC	8.02			
667	C	2	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CT	CG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	GG	CC	AG	CC	5.87	31.62	46.53	2.63
634[g]	C	8	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	5.76	36.35	797.86	205.35
202	AA	5	TC	GG	TC													CC	GG										CC	5.46			
822		4	TT	GC	TC	AG	CC	TC	AA	CG	CC	CC	CC	GG	AT	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	5.38	17.21	74.54	32.25
850	AA	4	TT	CC	TT													CC	GT										CC	4.7			
776	C	6	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CC	GG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	4.3	58.63	1234.75	87.63
775	C	6	TT	GC	TC	AG	CC	TC	AA	CG	CC	CC	CC	GG	AT	AG	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	4.11	25.64	112.99	23.17
797	C	6	TT	GC	TC	AA	CC	TC	AA	CC	CC	CC	CC	GG	AA	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	3.6	20.47	5.06	1.15
794	C	8	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG			AA	CC	3.25	42.44	770.69	74.45
763	C	2	TT	GC	TC	AG	CC	TC	AA	CG	CC	CC	CC	GG	AT	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	3.2	8.13	1.07	10.73
677	A	0	TT	GG	CC	GG	CC	TT	AA	GG	CC	CC	CC	GG	TT	AA	GG	TT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	3.12	22.81	526.39	40.75
821		9	TT	GC	TC	AA	CC	TC	AA	CC	CC	CC	CC	GG	AA	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	2.98	15.6	157.59	17.97
858[g]	C	5	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CT	CG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	GG	CC	AG	CT	2.74	28.77	99.73	5.59
362	AA	4	TT	GG	TC													CT	GG										CC	2.67			
633	C	8	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CT	1.64	28.86	1269.46	101.79
819	C	9	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CC	GG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	AG	CC	AA	CT	1.64	20.45	194.01	43.72
747[f]	C	5	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CT	1.02	11.33	29.24	11.79
159	AA	7	TT	GG	CC													TT	GG										CC	BQL			
622[f]	AA	4	TC	GC	TT	AA	CT	TC	AA	CC	CT	CC	CC	GG	AA	AA	CC	CC	GT	AA	CC	TT	CC	AA	AG	GG	CT	AA	CC	BQL	BQL	38.32	26.63
3	AA	5	TT	GG	TC													CC	GG										CT				
253	AA	4	CC	GG	TT													CC	GG										CC				

Table 5B. Caucasian Males sorted by CYP2B6 Activity   CYP2B6

HL#	a	-8427	-8207	-2320	-1778	-1578	-1456	-1224	-1186	-757	-750	64	12740	12917	13072	14593	15582	15631	17670	17755	17897	18000	18053	18627	21203	21563	25241	25505	Act[a]	Pr[b]	mRNA[c]	CmRNA[d]
871	8	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CT	17.75	37.76	30.48	5.05
789[f]	8	TT	GG	TT	AA	CC	TT	AG	CC	CC	TT	CC	GG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	AG	CC	AA	CC	17.09	58.11	209.38	86.5
659[h]	3	TT	GC	TC	AG	CC	TC	AA	CG	CC	CC	CC	GG	AT	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	11.56	50.19	3590.58	388.8
806	3	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CT	9.57	56.22	1770.57	184.26
785	8	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CC	GG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AG	GG	GG	CC	AA	CC	7.24	41.78	243.88	36.16
780	4	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CT	6.95	28.47	310.83	23.79
867	0	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	AG	CC	AA	CC	4.47	33.44	6.96	3.03
784	9	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	4.08	7.246	21.6	6.32
781	6	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CC	GG	AT	AA	CG	TT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	4.06	16.93	56.49	9.62
777	9	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	3.78	12.46	12.82	4.77
824	2	TT	GG	TC	AG	CC	TT	AA	CG	CC	TC	CC	GG	AT	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	3.54	53.4	1541.37	1913.18
729	3	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CC	GG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CT	3.4	15.48	23.92	2.96
779	4	TT	GG	CC	AA	CC	TT	AA	CC	CC	CC	CC	GG	AA	AA	GG	TT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	3.32	19.33	87.43	20.87
774	3	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CC	GG	GG	GG	CC	CT	AA	GG	GG	CC	AA	CC	3.21	24.65	15.89	53.62
796	9	TT	CC	TT	AA	CC	CC	AA	CC	CC	CC	CC	GG	AA	AA	CC	CC	TT	GG	GG	CC	CC	GG	GG	GG	TT	AA	CC	3.08	14.49	8.57	24.41
771	8	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	2.8	12.43	12.73	5.6
758	7	TT	GG	TT	AA	CC	TC	AA	CC	CC	TT	CT	CG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AG	CC	2.71	13.08	19.16	5.09
817	6	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	2.56	7.467	98.36	39.68
773[g]	3	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CT	2.3	5.431	2.17	4.37
790	0	TT	GC	TC	AA	CC	TC	AA	CC	CC	CC	CC	GG	AA	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	1.76	11.78	8.46	0.46
766[g]	0	TT	GC	TC	AA	CC	TC	AA	CC	CC	CC	CC	GG	AA	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	1.75	6.851	32.45	115.65
754	2	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CT	CG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	GG	CC	AG	CC	1.72	18.25	25.99	12.55
861	2	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	1.65	42.42	56.1	0.2
825	3	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CC	GG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CT	1.38	11.38	47.5	8.28
656	3	TT	GG	TT	AA	CC	TT	AA	CC	CC	TT	CC	GG	AA	AA	CC	CC	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CT	BQL	13.54	25.28	0.19
761	7	TT	GG	TC	AA	CC	TT	AA	CC	CC	TC	CT	CG	AA	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AG	CC	BQL	11.18	20.68	0.08
762[g]	2	TT	GG	TC	AG	CC	TT	AA	CG	CC	TC	CC	GG	AT	AA	CG	CT	GG	GG	GG	CC	CC	AA	GG	AG	CC	AA	CC	BQL	5.509	3.46	0.98
765	3	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CT	CG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AG	CC	BQL	10.42	11.16	8.63

Table 5C. Hispanic Males sorted by CYP2B6 Activity   CYP2B6

HL#	a	-8427	-8207	-2320	-1778	-1578	-1456	-1224	-1186	-757	-750	64	12740	12917	13072	14593	15582	15631	17670	17755	17897	18000	18053	18627	21203	21563	25241	25505	Act[a]	Pr[b]	mRNA[c]	CmRNA[d]
783	3	TC	GC	TT	AA	CT	TC	AA	CC	CT	CC	CC	GG	AA	AA	CC	CC	GT	AA	CC	TT	CC	AA	AG	GG	CT	AA	CC	6.89	40.69	270.6	58.61
637[h]	4	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CT	CG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AG	CC	3.32	27.81	770.69	38.83
696	2	TT	GC	TC	AG	CC	TC	AA	CG	CC	CC	CC	GG	AT	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	BQL	9.603	1.36	0.24
743	6	TT	GG	CC	AG	CC	TT	AA	CG	CC	CC	CC	GG	AT	AA	GG	TT	GG	GG	GG	CC	CC	AA	GG	GG	CC	AA	CC	BQL	7.538	6.41	0.96

Table 5D. African American Males sorted by CYP2B6 Activity   CYP2B6

HL#	a	-8427	-8207	-2320	-1778	-1578	-1456	-1224	-1186	-757	-750	64	12740	12917	13072	14593	15582	15631	17670	17755	17897	18000	18053	18627	21203	21563	25241	25505	Act[a]	Pr[b]	mRNA[c]	CmRNA[d]
216	0	TC	GC	TT													CC	GT											CC	14.41			
1	2	TT	GC	TT													CC	GT											CC	11.85			
847	5	CC	GG	TT													CC	GG											CC	9.55			
844	8	TT	GC	TC													CT	GT											CC	6.02			
718	9	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CT	4.94	19.13	17.03	77.44
177	9	TT	GC	TC													CT	GT											CC	4.25			
359	6	TC	GG	TT													CC	GG											CC	2.84			
638	2	TT	GC	TC	AA	CC	TC	AA	CC	CC	CC	CC	GG	AA	AA	CG	CT	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CC	2.13	12.18	424.61	14.74
788	7	TT	CC	TT	AA	CC	CC	AA	CC	CC	CC	CC	GG	AA	AA	CC	CC	TT	GG	GG	CC	CC	GG	GG	GG	TT	AA	CC	1.07	5.743	3.78	31.26
151	3	TT	GG	TC													CT	GG											CT	BQL			
740	6	TT	GC	TT	AA	CC	TC	AA	CC	CC	TC	CC	GG	AA	AA	CC	CC	GT	GG	GG	CC	CC	AA	GG	GG	CT	AA	CT	BQL	12.3	25.81	11.55
811	0	CC	GG	TT	AA	TT	TT	AA	CC	TT	CC	CC	GG	AA	AA	CC	CC	GG	AA	CC	TT	CC	AA	AG	GG	CC	AA	CC	BQL	14.88	13.93	11.79
44	5	TT	GC	TT													CT	GT											CC				
237	3	CC	GG	TT													CC	GG											CC				
248	4	TT	GC	TT													CT	GG											CC				
303	4	TT	CC	TT													CC	TT											CC				
624	3	TT	CC	TT													CC	GT											CC				
937	1	TT	GG	TT													CT	GG											CC				

Abbreviations: AA, African American; C, Caucasian; H, Hispanic; A, Asian
[a] Nirvanol Formation (nmol/min/mg protein); BQL, below quantifiable limits
[b] pmol CYP2B6 protein/mg microsomal protein
[c,d] relative CYP2B6 and CAR mRNA (arbitrary units), respectively
[e-h] Subjects on known inducers of CYP2B6: [e]Fluconazole; [f] Phenobarbital, [g] Dexamethasone, [h] Dilantin,
Additional Patient Demographics: Human liver #

Table 6: Association between *CYP2B6* genotypes and phenotype

SNPs	Group	P	Phenotype
Exon 9 (R487C)	CF	0.0015	Activity
	All F	0.085	Activity
Intron3 branch site	CF	0.030	Protein
	All F	0.048[a]	Activity
	All F	0.030[d]	Protein
	All persons	0.069[f]	Protein
-8207(putative DR2)	All persons	0.070[e]	Activity
-2320 (HNF4)	All F	0.069[b]	Activity
-750 (HNF1)	All F	0.104[c]	Protein
Exon 1 (R22C)	All F	0.059	Activity

F=females; C = Caucasian

[a]CC vs. CT, P=0.083; CC vs. TT, P=0.119

[b] CC vs TT, P=0.034

[c]CC vs. TC, P=0.091; CC vs. TT, P=0.098

[d]CC vs. CT, P=0.021

[e]CC vs TT, P=0.030

[f]CC vs TT, P= 0.040

Table 7. Haplotype frequencies estimated separately for each population.

Haplotype	-8427	-8207	HNF4 -2320	15582C>T int3br pt	Q172H 516G>t exon 4	R487C 1459C>T exon9	African Americans (n=58)	Caucasians (n=86)	Hispanics (n=14)
I	T	C	T	C	T	C	0.28	0.20	0.43
II	C	G	T	C	G	C	0.22	0.00	0.14
III	T	G	T	C	G	C	0.16	0.38	0.21
IV	T	G	C	T	G	C	0.12	0.24	0.21
V	T	G	T	T	G	C	0.07	0.02	0.00
VI	T	C	T	C	G	C	0.05	0.00	0.00
VII	T	G	T	C	G	T	0.05	0.12	0.00
VIII	T	G	C	C	G	T	0.03	0.01	0.00
IX	T	G	C	C	G	C	0.02	0.00	0.00
X	T	C	C	C	T	C	0.00	0.01	0.00
XI	T	G	T	C	T	C	0.00	0.01	0.00

n=number of alleles

The six loci with greatest extent of nucleotide diversity have been used; The Hispanic values have lower confidence due to sample size.