Sequence diversity in *CYP3A* promoters and characterization of the genetic basis of polymorphic CYP3A5 expression

Peter Kuehl[1,2]*, Jiong Zhang[2,3]*, Yvonne Lin[4], Jatinder Lamba[5], Mahfoud Assem[5], John Schuetz[5], Paul B. Watkins[6], Ann Daly[7], Steven A. Wrighton[8], Stephen D. Hall[9], Patrick Maurel[10], Mary Relling[5], Cynthia Brimer[5], Kazuto Yasuda[5], Raman Venkataramanan[11], Stephen Strom[12], Kenneth Thummel[4], Mark S. Boguski[2] & Erin Schuetz[5]

**These authors contributed equally to this work.*

Variation in the CYP3A enzymes, which act in drug metabolism, influences circulating steroid levels and responses to half of all oxidatively metabolized drugs. CYP3A activity is the sum activity of the family of CYP3A genes, including *CYP3A5*, which is polymorphically expressed at high levels in a minority of Americans of European descent and Europeans (hereafter collectively referred to as 'Caucasians'). Only people with at least one *CYP3A5*1* allele express large amounts of CYP3A5. Our findings show that single-nucleotide polymorphisms (SNPs) in *CYP3A5*3* and *CYP3A5*6* that cause alternative splicing and protein truncation result in the absence of CYP3A5 from tissues of some people. CYP3A5 was more frequently expressed in livers of African Americans (60%) than in those of Caucasians (33%). Because CYP3A5 represents at least 50% of the total hepatic CYP3A content in people polymorphically expressing CYP3A5, *CYP3A5* may be the most important genetic contributor to interindividual and interracial differences in CYP3A-dependent drug clearance and in responses to many medicines.

Introduction

A major goal of the Human Genome Project has been completed with publication of the draft sequence of the human genome[1,2]. Understanding the biological and medical implications of variation in the human genome sequence is another major aim of the Human Genome Project[3,4]. Cytochromes P450 (CYP) are particularly amenable for rapid identification of functionally important sequence variations, because CYP catalytic activities can be determined by various *in vivo* and *in vitro* drug bioassays. Moreover, the applications of such information are particularly relevant for pharmacogenomics, in which knowledge of SNPs in CYP genes may lead to individualized drug dosing and improved therapeutics[5]. Here we elucidate whether sequence variations in the promoters of genes encoding enzymes of the CYP3A family contribute to variability in CYP3A metabolism.

The cytochromes P450 evolved to catalyze the metabolism of numerous structurally diverse exogenous and endogenous molecules. Approximately 55 different CYP genes are present in the human genome and are classified into different families and subfamilies on the basis of sequence homology. The CYP families have arisen through a process of gene duplication and gene conversion. Members of the CYP3A subfamily catalyze the oxidative, peroxidative and reductive metabolism of structurally diverse endobiotics, drugs, and protoxic or procarcinogenic molecules[6]. The CYP3A members are the most abundant CYPs in human liver and small intestine[7,8]. Substantial interindividual differences in CYP3A expression, exceeding 30-fold in some populations[9], contribute greatly to variation in oral bioavailability and systemic clearance of CYP3A substrates, including HIV protease inhibitors, several calcium channel blockers and some cholesterol-lowering drugs. Variation in CYP3A expression is particularly important for substrates with narrow therapeutic indices, such as cancer chemotherapeutics[10] and the immunosuppressants cyclosporin A and tacrolimus (FK506). Such variation in CYP3A can result in clinically significant differences in drug toxicities (for example, nephrotoxicity) and response (for example, graft survival). Moreover, because CYP3A metabolizes estrogens to 2-hydroxyesterone, 4-hydroxyestrone and 16α-hydroxylated estrogens, all of which have been implicated in estrogen-mediated carcinogenicity[11], variation in CYP3A may influence the circulating levels of these estrogens and the risk of breast cancer.

Human CYP3A activities reflect the heterogeneous expression of at least three CYP3A family members: CYP3A4, CYP3A5 and

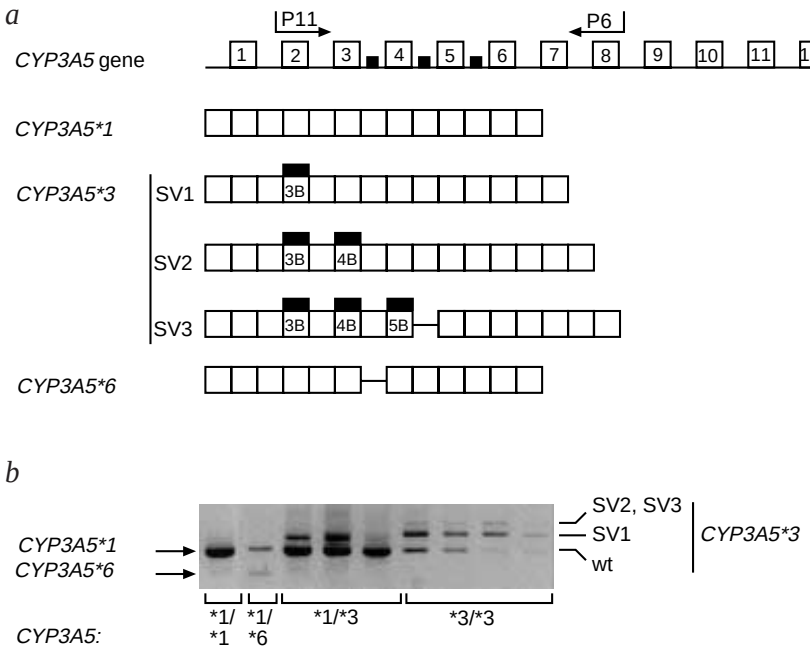

Fig. 1 Splice variants of *CYP3A5*. ***a***, Schematic drawing of the cDNA structures of the various *CYP3A5* alleles. Exons are numbered and boxed, and primers P11 (exon 2/3) and P6 (exon 8) are indicated. The location of intronic sequences coding for exons 3B, 4B and 5B derived from intron sequences are indicated as filled boxes. ***b***, RT–PCR analysis of mRNAs encoded for by the various *CYP3A5* alleles. Total mRNA from human livers was reverse transcribed and amplified using PCR primers P11 (exon 2/3) and P6 (exon 8) and the products analyzed on agarose gels. *CYP3A5* cDNA was amplified by 30 and 35 cycles of RT–PCR from hepatic RNA of people expressing high and low amounts of CYP3A5 protein, respectively. The PCR product of *CYP3A5*1* is the expected size of 539 bp; the *CYP3A5*6* mRNA lacking exon 7 is 391 bp; the normally spliced *CYP3A5*3* mRNA (539 bp) (wt) and the *CYP3A5*3* SV1 mRNA containing exon 3B (670 bp) and SV2 mRNA containing exon 3B and 4B (775 bp) and the SV3 mRNA containing exons 3B, 4B and 5B and missing exon 6 (769 bp). The *CYP3A5*1/*1*, *CYP3A5*1/*3* and *CYP3A5*3/*3* genotypes (AA, AG and GG, respectively, at nt 22,893 in *CYP3A5* (AC005020)).

CYP3A7. The CYP3A genes are adjacent to each other on chromosome band 7q21, but the genes are differentially regulated[12]. Two CYP3A pseudogenes, *CYP3AP1* and *CYP3AP2*, were recently identified[12]. Functional CYP3A4 is found in most adults, with 10- to 40-fold variation in its expression. CYP3A7 is predominantly expressed in fetal life, and its expression seems to be silenced shortly after birth; however, some people express *CYP3A7* mRNA into adulthood[13]. CYP3A4 activity is correlated with *CYP3A4* mRNA concentrations, indicating that transcriptional control is the primary mechanism for regulating expression of *CYP3A4*. CYP3A5 was previously detected in livers and small intestines of some adult Caucasians, but the basis for this 'polymorphic' expression is unknown[14–16]. Recently, two mutations in what was believed to be the *CYP3A5* promoter were found to be associated with polymorphic CYP3A5 expression[17]. Sequencing of the CYP3A locus, however, revealed that these SNPs are actually in the promoter of the pseudogene *CYP3AP1* (ref. 12) and, thus, cannot be the basis for polymorphic CYP3A5 expression. The *CYP3A4*, *CYP3A5* and *CYP3A7* promoters contain multiple putative transcription-factor–binding sites. In particular, a recently discovered nuclear receptor, PXR/SXR (for pregnane X-receptor/steroid and xenbiotic receptor), that binds to a PXR response element in the *CYP3A4* and *CYP3A7* promoters is important for regulation of the CYP3A proteins[18,19]. Therefore, we hypothesized that polymorphisms affecting CYP3A activity are present in DNA regulatory sequences. Through a systematic analysis of genomic DNA from a large number of patients with various drug metabolism phenotypes, and through analysis of DNA from the Coriell DNA Polymorphism Discovery Resource (DPDR; ref. 3), we identified all common variants in the promoter regions of *CYP3A4*, *CYP3A5*, *CYP3AP1* and *CYP3A7*, and demonstrated ethnic differences in allele frequencies. By performing quantitative immunoblotting and *in vitro* kinetic assays of drug metabolism, we evaluated the relationship between these common genetic variants and ethnic- and tissue-specific phenotypic variability. In addition, analysis of human liver *CYP3A5* cDNA revealed that only those people with a *CYP3A5*1* allele produce high levels of full-length *CYP3A5* mRNA and express CYP3A5. Those with the *CYP3A5*3* allele have sequence variability in intron 3 that creates a cryptic splice site and results in the generation of *CYP3A5* exon 3B; this *CYP3A5*3* allele encodes an aberrantly spliced mRNA with a premature stop codon. This finding explains the molecular defect responsible for one of the most common polymorphisms in drug-metabolizing enzymes.

Results

We initially performed SNP analysis of the promoter regions of the CYP3A genes on DNA from the smallest DPDR subset (Table 1). The minimum DPDR panel contained five variant alleles with a frequency greater than or equal to 1% in the combined study population. These variant alleles included the *CYP3A4*1B* allele, in which an A→G variant at nt –288 is present in the nifedipine-specific element of the promoter[20]. We detected the *CYP3AP1*1* reference allele (G at nt –44 in the promoter) in 43 people (28 heterozygotes and 15 homozygotes). In an unusual haplotype, hereafter referred to as *CYP3A7*1C* (Table 1), an approximately 60-bp stretch (nt –129 to –188) of the *CYP3A7* promoter seems to have been replaced with sequence identical to the same region in the *CYP3A4* promoter. This replacement indicates the occurrence of a gene conversion event[21].

The Coriell DPDR panel was designed to represent the genetic diversity of United States residents who have ancestors from Europe, Africa, the Americas and Asia. We found no variants in the DPDR panel that were not also seen in the combined study population, but additional *CYP3A* alleles found in our larger combined population were not present in the DPDR. Most of these alleles were rare (that is, they were seen in only one or two people). To assess the usefulness of the DPDR panel in predicting SNPs, we calculated the probability that common variants would be missed in a random sampling of our combined population. The chance of missing the polymorphic alleles *CYP3A4*1B* and *CYP3AP1*1* was less than 5%, even if the sample consisted of as few as 10 study subjects from the combined population group. If we were to randomly select a sample of 25 subjects, the probability

Table 1 • Frequencies of SNPs in *CYP3A* promoters

Gene	Location	Name of variant allele[e]	Reference allele	Variant allele	DPDR[a] reference allele	DPDR[a] variant allele[f]	Combined population[b] reference allele	Combined population[b] variant allele[f]	Caucasians[c] reference allele	Caucasians[c] variant allele[f]	African Americans[d] reference allele	African Americans[d] variant allele[f]	Region	Effect of variant allele
CYP3A4[g]	–341	*CYP3A4*1C*	T**g**gga	T**G**gga	1.00	0.00	0.99	0.00	0.99	0.01	1.00	0.00		
	–288	*CYP3A4*1B*[h]	ggca**A**gaga	ggca**G**gaga	0.81	0.19	0.87	0.13	0.98	0.02	0.65	0.35	NFSE	
	+43	*CYP3A4*1D*	ccag**C**aaag	ccag**A**aaag	1.00	0.00	0.99	0.00	0.99	0.01	1.00	0.00		
CYP3AP1[g,h]	–117	*CYP3AP1*1B*	ttcc**C**cata	ttcc**T**cata	1.00	0.00	0.99	0.01	1.00	0.00	0.97	0.03		
	–109	*CYP3AP1*1C*	agaa**T**atga	agaa**C**atga	1.00	0.00	0.99	0.01	1.00	0.00	0.97	0.03		
	–44	*CYP3AP1*3*	cccc**G**cctc	cccc**A**cctc	0.38	0.62	0.22	0.78	0.08	0.92	0.50	0.50	Sp1	associated with decreased hepatic and intestinal CYP3A5 protein
CYP3A7[g]	–211	*CYP3A7*1B*	actc**C**ccag	actc**T**ccag	1.00	0.00	0.99	0.00	0.99	0.01	1.00	0.00		
	–188	*CYP3A7*1C*[i]	ttgt**G**tatg	ttgt**T**tatg	0.87	0.13	0.96	0.04	0.97	0.03	0.94	0.06	HNF-5	enhanced 3A7 mRNA after birth
	–181	*CYP3A7*1C*[i]	tgat**T**ctac	tgat**A**cctc	0.87	0.13	0.96	0.04	0.97	0.03	0.94	0.06		
	–179	*CYP3A7*1C*[i]	attc**T**acat	attc**C**acat	0.87	0.13	0.96	0.04	0.97	0.03	0.94	0.06		
	–178	*CYP3A7*1C*[i]	ttct**A**cata	ttct**T**cata	0.87	0.13	0.96	0.04	0.97	0.03	0.94	0.06		
	–167	*CYP3A7*1C*[i]	atat**T**aact	atat**G**aact	0.87	0.13	0.96	0.04	0.97	0.03	0.94	0.06	PXR-RE	
	–159	*CYP3A7*1C*[i]	tcaa**T**ggag	tcaa**A**ggag	0.87	0.13	0.96	0.04	0.97	0.03	0.94	0.06	PXR-RE	
	–129	*CYP3A7*1C*[i]	gatt**A**tttg	gatt**C**tttg	0.87	0.13	0.96	0.04	0.97	0.03	0.94	0.06	octamer motif	
	+13	*CYP3A7*1D*	gcag**G**gcag	gcag**A**gcag	0.94	0.06	0.99	0.01	0.99	0.01	1.00	0.00		
	+55	*CYP3A7*1E*	gcac**G**ctgc	gcac**A**ctgc	0.94	0.06	0.97	0.03	1.00	0.00	0.92	0.08		
CYP3A5	+18	*CYP3A5*1B*	gcag**G**gaag	gcag**A**gaag	NA	NA	0.98	0.02	0.97	0.03	1.00	0.00		
	+30	*CYP3A5*1C*	cagg**C**aaac	cagg**T**aaac	NA	NA	0.96	0.04	0.97	0.03	0.93	0.07		

[a]Analyzed people=8. [b]Analyzed people analyzed people=119 (*CYP3A4/7*), 123 (*CYP3A5P1*) or 74 (*CYP3A5*). [c]Analyzed people=79 (*CYP3A4/7*), 83 (*CYP3A5P1*) or 59 (*CYP3A5*). [d]Analyzed people=40 (*CYP3A4/7/P1*) or 15 (*CYP3A5*). [e]The *CYP* alleles are named according to recommended nomenclature guidelines[44] (http://www.imm.ki.se/CYPalleles). [f]The frequency of the variant alleles was calculated by using the formula: frequency = [2*(number of people homozygous for the variant allele)+(the number of heterozygous people)]/[2*(total number of people)]. The frequency of the reference allele was equal to the following: 1–(variant allele frequency). For DNA from families, we included only data gathered from the parents' samples in the frequency calculations. [g]The *CYP3A4*1*, *CYP3A7*1* and *CYP3A5*1* reference alleles are D11131, AF280107 and AC005020, respectively. Apparent errors in the original sequence of the *CYP3AP1* promoter are corrected and deposited as the consensus sequence for *CYP3AP1*1*. The variants are numbered relative to the initiation site of transcription, which is defined as +1. To renumber alleles according to the translation start site (A in initiation codon ATG is +1), subtract 103 bases from the SNP location (for example, the location of *CYP3AP1*3* changes from –44 to –147). [h]*CYP3A4*1B* and *CYP3AP1*1* are partially linked in African Americans. [i]These seven *CYP3A7* variants (hereafter referred to as *CYP3A7*1C*) are located between nt –129 and –188 and are in complete linkage.

of missing polymorphic *CYP3A7*1C* alleles would be 15%, that of missing *CYP3A7*1E* would be 25%, and that of missing *CYP3A7*1D* would be greater than 60%. The probability of missing the rare alleles *CYP3A4*1C*, *CYP3A4*1D*, *CYP3AP1*1B* and *CYP3AP1*1C* would be greater than 60%. Thus the DPDR 8-sample subset was more useful in accurately predicting the identity of *CYP3A* SNPs than a random sampling of 25 people from the combined population.

As has been previously reported[22], *CYP3A4*1B* occurred at a much higher frequency in African Americans (q=0.35) than in Caucasians (q=0.02; χ^2=48.9, P<0.001). We found no SNPs in the regions containing the proximal (nt –169 to –152) and distal (nt –7836 to –7208; ref. 23) PXREs of *CYP3A4*. Sequence analysis of the *CYP3A5* promoter (nt +90 to –280) revealed two SNPs in the 5´ UTR (Table 1). *CYP3A7*1C* was three times more common in an African American population (q=0.06) than in a Caucasian population (q=0.03). Among the 10 unrelated French Caucasians in the study, however, 3 people were heterozygous for *CYP3A7*1C* (q=0.15), whereas only 2 of the 96 unrelated non-French Caucasians carried 1 *CYP3A7*1C* allele (q=0.01). *CYP3AP1*1* was the most common minor allele in Caucasians and African Americans. This allele is also divergently distributed among African Americans (q=0.50) and Caucasians (q=0.08, χ^2=51.1, P<0.001). By examining other panels of DNA, we found that the frequency of *CYP3AP1*1* alleles (q values) in Japanese was 0.15; in Chinese, 0.3; in Mexicans, 0.13; in Southeast Asians (excluding Japanese and Chinese), 0.5; in Pacific Islanders, 0.3; and in Southwestern American Indians, 0.5.

We determined the association of *CYP3A* alleles with CYP3A phenotypes. None of the SNPs in the *CYP3A5* or *CYP3A4* promoters, including the *CYP3A4*1B* allele, were associated with altered CYP3A activities of people whose phenotypes had been determined *in vivo* by the erythromycin breath test or nifedipine clearance or *in vitro* by assays for total midazolam hydroxylase or for basal or rifampin-inducible testosterone 6β-hydroxylase activity. Previous studies have noted the polymorphic expression of CYP3A5 in the liver and intestine of approximately 25% of people studied[14,16,24]. Although there is good concordance of *CYP3AP1*1* (G at nt –44) with high levels of CYP3A5 in the liver[17], *CYP3AP1*1* cannot be responsible for polymorphic CYP3A5 expression. Because we did not find a *CYP3A5* promoter variant that correlated with CYP3A5 expression, we determined whether *CYP3AP1*1* was in linkage disequilibrium with another variant in *CYP3A5* that might be causal. Because the *CYP3AP1* genotype is correlated with CYP3A5 expression[17], we used mRNAs from Caucasian and African American livers representing the *CYP3AP1*1/*1*, *CYP3AP1*1/*3* and *CYP3AP1*3/*3* genotypes to generate and sequence the entire *CYP3A5* cDNA. The *CYP3A5* mRNA content was greater in people with a *CYP3AP1*1* allele than in those people homozygous for *CYP3AP1*3*. All livers contained the full-length *CYP3A5* mRNA transcript encoded by *CYP3A5*1* (Fig. 1). Many people carrying *CYP3AP1*3* also carried *CYP3A5*3*, which encodes only a small amount of normally spliced *CYP3A5* mRNA and a splice variant-1 (SV1) mRNA. SV1 results from a SNP at nt 22,893 (in AC005020), which creates a cryptic consensus splice site and exon 3B, an exon derived from intron-3 sequences of *CYP3A5* (Fig. 1 and Table 2). The mutated sequence in intron 3 of *CYP3A5*3* has a splice site prediction score of 0.9. Because of the premature termination codon in exon 3B, the SV1 mRNA encodes a protein whose sequence is truncated after amino acid 102. We found a second

Table 2 • *CYP3A5* alleles

CYP3A5	exon	Starts in AC005020*	Ends in AC005020	5′ splice site	3′ splice site	Functional consequence
	1	15826	15983		ctatctgtgagtaa	
	2	19601	19694	ctttgtagatatgg	cgtcaggtgagttg	
	3	21224	21276	tctcccagggtctc	gtgggggtgagtat	
3B, insertion from intron 3		22894	23025	tctttca[A/G]tatctc	ctaggggtatggat	*CYP3A5*3*- 22893G>A; insertion of exon 3B into SV1, SV2, SV3; premature stop after aa 102
	4	23130	23229	ccacacagaacgta	cgaagggtaagcat	
4B, insertion from intron 4		23931	24036	attcgtagatttgt	g[A/G]caggttct	*CYP3A5*3*-insertions of exon 4B into SV2 and SV3
	5	28744	28857	aactctagtcttt	aaggaggtatgaaa	
5B, insertion from intron 5		28999	29096	atgtacagaaaaga	ctacaggtactgat	*CYP3A5*3*-insertion of exon 5B into SV3
	6	29120	29208	gtgcttagatgttc	gaaagagtaagtag	*CYP3A5*3*-deletion of exon 6 from SV3
	7	30495	30643	ccactcagcatctt	caataagtatgtgg	*CYP3A5*6*-30597G>A; deletion of exon 7 from splice variant; frame shift and premature stop at 184 aa
	8	31714	31841	tcccacagtactct	caaaaggtaaaatc	
	9	32927	32993	gcttctagcaccga	acaaaggtaaccaa	
	10	35150	35310	cctttcagctctgt	aataaggtgagggg	
	11	43030	43256	ccttccaggcacca	tgaaaggtacaagt	
	12	45577	45736	ggaactaggttcag	acacaggtcagtac	
	13	47409	47615	tattgtagatcccc		

*The accession numbers for the reference *CYP3A5*1* cDNA and *CYP3A5*3* are J04813 and AC005020, respectively. SV1, splice variant-1; SV2, splice variant-2; SV3, splice variant-3. *CYP3A5*3* has a G at nt 22,893 and a consensus splice site of [tttcA**G**tatc] and aberrant splicing of *CYP3A5* mRNA. *CYP3A5*1* contains an A at nt 22,893 generating a sequence of [tttcA**A**tatc] and no aberrant splicing of *CYP3A5* mRNA. To renumber the mutation sites according to the translation start site (nt 15913 is A in the ATG initiation codon), subtract 15912 bases from the SNP location (for example, the location of *CYP3A5*6* is 14685G>A).

splice variant (SV2) mRNA that encoded exons 3B and 4B in many people carrying *CYP3A5*3*. Although sequence analysis around the predicted splice site for exon 4B revealed a SNP at nt 24,035 (Table 2), this SNP and the generation of SV2 were not related (Table 2). A third splice variant (SV3) mRNA comigrated with SV2 in an agarose gel (Fig. 1) and contained exons 3B, 4B and 5B, but not exon 6. No SNPs were found around exon 5B or exon 6 that correlated with the SV3 transcript. The fact that mRNAs with premature stop codons are more unstable and rapidly degraded[25] may explain the lower amount of *CYP3A5* mRNA in people homozygous for *CYP3A5*3* compared with people carrying at least one *CYP3A5*1* allele. We examined the frequency of the *CYP3A5*1* allele and found it to be distributed divergently among African Americans (q=0.45) and Caucasians (q=0.15). By examining other panels of DNA, we found that the frequency of *CYP3A5*1* alleles (q values) in the DPDR subset of 8 people was 0.31; in Japanese, 0.15; in Chinese, 0.35; in Mexicans, 0.25; in Southeast Asians (excluding Japanese and Chinese), 0.33; in Pacific Islanders, 0.35; and in Southwestern American Indians, 0.6. *CYP3A5*6* was identified in only 3 of 20 African Americans and encoded a normal *CYP3A5* mRNA and a splice variant mRNA in which exon 7 was deleted. This deletion resulted in a frameshift, causing the encoded protein to be truncated at amino acid 184. Sequence analysis of genomic DNA and cDNA synthesized from the spliced variant mRNA from those individuals skipping exon 7 identified a single sequence variation: a 30597G>A synonymous mutation in exon 7 (nt 711 G>A in the *CYP3A5* cDNA). This genotype was concordant with the loss of exon 7. This silent mutation may cause exon 7 skipping by disrupting an exonic splicing enhancer[26].

To determine the association of the *CYP3A5* and *CYP3AP* alleles with CYP3A5 expression, we measured CYP3A5 protein concentration in liver specimens. We found a greater frequency of CYP3A5 (≥21 pmol/mg protein) in African Americans (11/20; 50%) compared with Caucasians (9/27; 33%). All Caucasians and most African Americans with a CYP3A5 content less than 21 pmol/mg protein were homozygous for the *CYP3A5*3* allele with G at nt 22,893 leading to cryptic *CYP3A5* mRNA splicing, whereas those with CYP3A5 content ranging from 21 to 202 pmol/mg had at least one *CYP3A5*1* allele (A at nt 22,893 with no aberrant *CYP3A5* mRNA splicing; Fig. 2*a*,*b*). Among the livers of Caucasians, the *CYP3A5* allele frequencies conformed to Hardy–Weinberg equilibrium. There was also complete concordance between the *CYP3A5*1* and *CYP3AP1*1* genotypes and between *CYP3A5*3* and *CYP3AP1*3* genotypes in Caucasians. Among African Americans, however, there were several outliers when the *CYP3AP1* genotype was used to predict African American CYP3A5 phenotype. Specifically, African American liver 788 had a relatively high level of CYP3A5 (54 pmol/mg protein) that was not predicted by a *CYP3AP1*3/*3* genotype, but was explained by the *CYP3A5*1/*3* genotype. African American livers 624 and 958.29 had relatively low levels of CY3A5 protein that were not predicted by a *CYP3AP1*1/*1* genotype, but were better explained by *CYP3A5*1/*6* genotypes because the *CYP3A5*6* allele generates an mRNA lacking exon 7. The final outlier (African American liver 50) had only a trace amount of CYP3A5 protein that was not predicted by the *CYP3AP1*1/*3* genotype or the *CYP3A5*1/*3* genotype. This discrepancy may be due to secondary degradation of CYP3A5 and other proteins following poor tissue preservation, or to a second mutation in the coding region of *CYP3A5* that confers instability of the protein (for example, *CYP3A5*2*; ref. 27).

We examined the rates of midazolam hydroxylation by the same 47 livers. CYP3A4 and CYP3A5 each produce two metabolites: 1′-hydroxymidazolam (1′-OH MDZ) and 4-hydroxymidazolam (4-OH MDZ). The mean reaction velocities were 2.5-fold (P=0.03) higher for the livers of Caucasians and 2.2-fold (P=0.19) higher for the livers of African Americans with at least one *CYP3A5*1* allele, compared with those livers of people who were homozygous for *CYP3A5*3* (Fig. 3). The ratio of 1′-OH MDZ to 4-OH MDZ is dependent on substrate concentration and the type of CYP3A enzyme[28]. At a midazolam concentration of 8 μM, the ratio was 5.5 when we used recombinant CYP3A4 and 16.1 when we used recombinant CYP3A5. Thus, with liver microsomes, the observed product ratio depends on the relative amounts of CYP3A4 and CYP3A5 protein. The average ratio of 1′-OH MDZ to 4-OH MDZ was higher in Caucasian livers with at least one *CYP3A5*1* allele (8.1±3.0 for **1/*3* genotype) than in those with two *CYP3A5*3* alleles (5.9±1.7 for **3/*3*; P=0.09). The

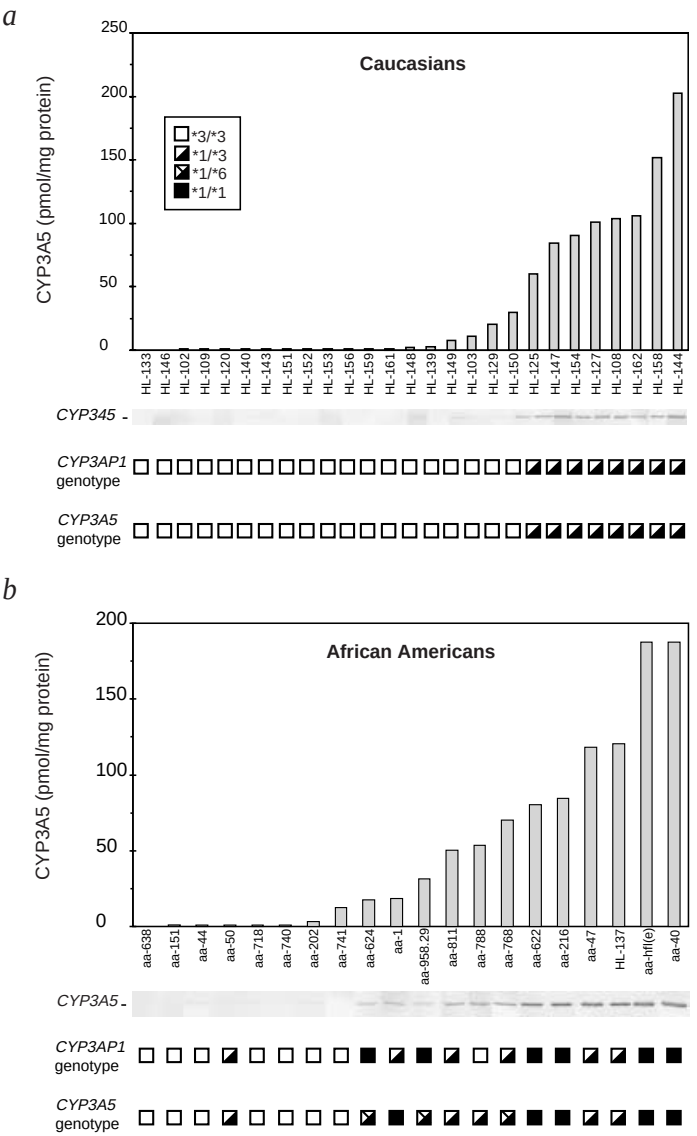

Fig. 2 Relationship between CYP3A5 phenotype and *CYP3A5* genotype in livers of Caucasians (***a***) and African Americans (***b***). Microsomal content of CYP3A5 protein in human livers (HL) was determined by immunoblot analysis. Identification of the *CYP3AP1* and *CYP3A5* genotypes were made by direct sequencing of genomic DNA: the *CYP3AP1*1/*1*, *CYP3AP1*1/*3* and *CYP3AP1*3/*3* genotypes (GG, AG and AA, respectively at nt –44 of the *CYP3AP1* promoter). The *CYP3A5* genotypes are defined in Fig. 1.

average ratio of 1´-OH MDZ to 4-OH MDZ was higher in African American livers with at least one *CYP3A5*1* allele (9.8±3.8 for **1/*1,*3,*6* genotypes) than in those with two *CYP3A5*3* alleles (8.3±1.6 for **3/*3*; P=0.41).

Measurement of CYP3A quantity in the same 47 livers showed that CYP3A4 content was not significantly different between Caucasian and African American livers. Within each ethnicity there was no difference in mean CYP3A4 content between livers with or without at least one *CYP3A5*1* allele. The amount of total CYP3A in Caucasian and African Americans was approximately threefold higher in those with the *CYP3A5*1* allele than in people not carrying this allele (P=0.001 and P=0.01, respectively). Moreover, CYP3A5 protein represented more than 50% of the total CYP3A content in one-third of Caucasian livers and over one-half of African American livers. Thus, CYP3A5 represented a considerably greater proportion of total CYP3A than previously estimated[14] and contributed significantly to CYP3A content and catalytic activity.

Because CYP3A5 is variably expressed in human intestine[16], we took advantage of a unique cohort of paired samples of small intestine and liver to determine whether the *CYP3AP1*1* genotype-phenotype concordance is seen in extrahepatic tissues. All people with large quantities of CYP3A5 in the liver also had relatively high intestinal levels of CYP3A5 and were heterozygous or homozygous for the *CYP3A5*1* allele (Fig. 4). Further, the intestinal midazolam product ratio for people with at least one *CYP3A5*1* allele (9.6±2.5 for the AG genotype) differed from that for people with two *CYP3A5*3* alleles (6.0±0.3; P=0.06).

The *CYP3A4*1B* allele was originally proposed to be associated with altered CYP3A4 hepatic activity[20,29,30], but this proposal has been controversial[31,32]. Because the *CYP3A4*1B* and *CYP3AP1*1* or *CYP3A4*1B* and *CYP3A5*1* alleles can all be present in the same person, we hypothesized that it is ultimately the *CYP3A5* genotype that influences the overall functional activity of CYP3A. African Americans frequently carried both *CYP3AP1*1* and *CYP3A4*1B* alleles (Table 3), and using data from those people homozygous for *CYP3A4*1B* or *CYP3AP1*1*, we determined the linkage of the two alleles (data not shown; χ^2=12.8, P<0.001). The association of the *CYP3A5*1* and *CYP3A4*1B* alleles was not significant in African Americans, but approached statistical significance in Caucasians.

The *CYP3A7* promoter contains a set of 7 tightly linked variants that replaced 60 bp of the *CYP3A7* promoter with the identical region from the *CYP3A4* promoter; thus this replacement altered 3 transcription-factor–binding sites[33]: hepatic nuclear factor-5 (HNF-5), an octamer motif and the PXRE. Although CYP3A7 is generally expressed only during fetal development, *CYP3A7* mRNA is detected in a small fraction of adults[13]. We speculated that these nucleotides are important for the increased expression of hepatic CYP3A4 and loss of CYP3A7 expression in most people after birth. We evaluated hepatic *CYP3A7* mRNA expression in nine adult livers (five with the variant *CYP3A7*1C* allele) to determine whether this replacement of *CYP3A7* promoter elements by *CYP3A4* sequences is related to expression of CYP3A7. The two people with the highest levels of *CYP3A7* mRNA carried *CYP3A7*1C*, and four of five people with *CYP3A7*1C* alleles expressed *CYP3A7* mRNA (data not shown). One person with the *CYP3A7*1C* allele had no *CYP3A7* mRNA, however, and two people lacking *CYP3A7*1C* had very low but detectable levels of *CYP3A7* mRNA. Thus, *CYP3A7*1C* is associated with *CYP3A7* mRNA expression, but is not the sole explanation for *CYP3A7* expression in adults.

Discussion

Our findings show that SNPs in *CYP3A5* that cause alternative splicing and truncation of CYP3A5 protein provide a molecular explanation for the absence of CYP3A5 protein from some people. The most common cause of the loss of hepatic CYP3A5

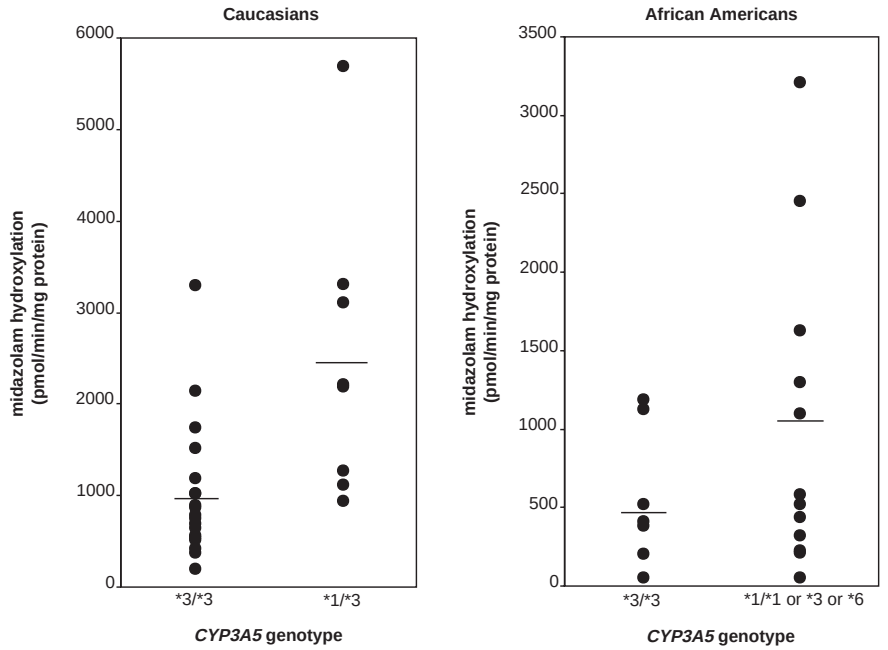

Fig. 3 Rates of midazolam hydroxylation in livers with different *CYP3A5* genotypes. Total rate of midazolam hydroxylation in liver microsomes was measured. The *CYP3A5* genotypes are defined in Figs. 1 and 2. Bars indicate mean values for the subpopulations.

expression is a SNP at nt 22,893 (AC005020) in intron 3 of *CYP3A5*3*; this SNP generates a cryptic splice site and exon 3B. Splicing of this exon into the *CYP3A5*3* transcript introduces a stop codon. Thus, translation of the abnormal transcript generates a protein that is prematurely terminated at amino acid 109. The new *CYP3A5*6* allele, which was identified in African Americans, contained a 30597G>A mutation in exon 7. This SNP correlates with the deletion of exon 7 from the *CYP3A5* mRNA and lower CYP3A5 catalytic activity in three African Americans. CYP3A5 expression was also closely associated with the *CYP3AP1*1* genotype[12,17]; however, the CYP3A5 phenotype was better predicted by *CYP3A5* genotype. Comparison of *CYP3AP1*1* and *CYP3A5*1* genotypes among the DPDR subset revealed only 37.5% linkage; in Mexicans, only 65% linkage; and in Japanese, Southeast Asians, Chinese and Pacific Islanders, 90–95% linkage. Although we detected small amounts of normally spliced *CYP3A5* mRNA and resulting protein even in those people homozygous for *CYP3A5*3*, cryptic splices sites within intron 3 were used preferentially. Thus, only those people with at least one *CYP3A5*1* allele polymorphically express high amounts of CYP3A5. These results add to the growing list of splicing alterations that affect the expression of clinically important genes[34].

In addition to determining the basis for polymorphic CYP3A5 expression, we demonstrated that in people with at least one *CYP3A5*1* allele, the peak hepatic content of CYP3A5 is much higher than previously thought[24,35] and is approximately the same as the reported peak level of CYP3A4 (ref. 8). For most Caucasians and African Americans who carry the *CYP3A5*1* allele, we also found that CYP3A5 accounts for at least 50% of the total CYP3A content. Thus, CYP3A5 should contribute substantially to the total metabolic clearance of the many CYP3A substrates. Indeed, we predict that people with the highest clearance and lowest oral bioavailability of CYP3A substrates will be heterozygous or homozygous for *CYP3A5*1*. These people might be more likely to encounter a lack of efficacy from a standard dose of active parent drug. Moreover, because CYP3A4 and CYP3A5 have varying degrees of catalytic capability and regioselectivity toward some substrates and because CYP3A4 and CYP3A5 are differentially inhibited by some compounds, polymorphic CYP3A5 expression will contribute to differences in metabolite profiles and in susceptibility to inhibitory drug interactions. One can also envision an increased risk of adverse effects from CYP3A5-generated toxic metabolites in people that predominantly express CYP3A5. For instance, the rate of 3α-hydroxylation of aflatoxin B_1 (the inactivation step) by CYP3A5 is approximately 1% of that of CYP3A4, whereas CYP3A5 efficiently catalyzes the 8,9-epoxidation (the activation step) of aflatoxin B_1 (ref. 36). Thus, in those for whom CYP3A5 is the dominant hepatic CYP3A, the relative lack of aflatoxin B_1 detoxification compared to activation of this agent might enhance the genotoxicity of aflatoxin[36]. Because polymorphic CYP3A5 is one factor contributing to individual variation in CYP3A-mediated metabolism of drugs[14,35], simple DNA-based tests can now be used to determine how individual differences in CYP3A5 contribute to the overall metabolic fate of these CYP3A substrates, to their pharmacodynamic variability and to disease risk.

CYP3A5 is the primary CYP3A family member expressed outside the liver and intestine (for example, kidney, lung and polymorphonuclear leukocytes), indicating that CYP3A5 has an important physiologic function in these tissues. Indeed, CYP3A5 can mediate the metabolism of cortisol to 6β-hydroxycortisol, a physiologic regulator of Na^+ transport in renal epithelia[37]. Variable and polymorphic renal expression of CYP3A5 (refs. 38,39) may contribute to individual differences in the localized generation of 6β-hydroxycortisol within the nephron and may have an etiologic role in, for example, salt-sensitive hypertension by increasing renal retention of Na^+. Thus, CYP3A5 is not just a catalyst of drug detoxification, but, in organs such as the kidney, it may serve an important function in regulating the pool of endogenous paracrine or endocrine factors. Similarly, polymorphic expression of CYP3A5 may contribute to variable metabolism of steroids in the prostate and breast, to differences in the concentrations of circulating steroids and, hence, to risk of disease in these tissues.

Among the Caucasians and African Americans studied here, the *CYP3A5*1* genotype was associated with significant levels of CYP3A5 protein in the liver and small intestine, the two main tissues in which CYP3A5 is likely to make the greatest contribution to drug elimination. The polymorphic distribution of the *CYP3A5*1* allele indicates that relatively high levels of metabolically active CYP3A5 are expressed by an estimated 30% of Caucasians, 30% of Japanese, 30% of Mexicans, 40% of Chinese and more than 50% of African Americans, Southeast Asians, Pacific Islanders and Southwestern American Indians. The higher prevalence of CYP3A5 expression indicates that these non-Caucasians are more likely to experience higher clearance of drugs principally inactivated by CYP3A; are less

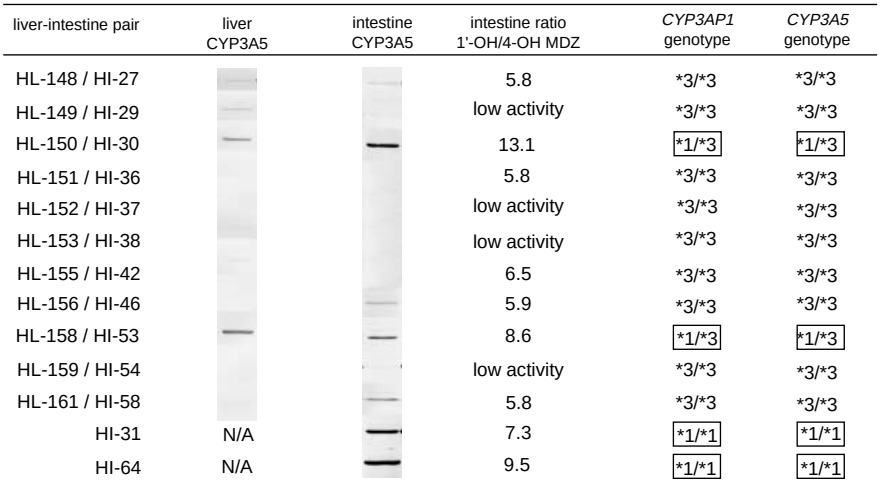

Fig. 4 Association between CYP3A5 expression and *CYP3A5* genotype in human liver and small intestine. Eleven pairs of tissue obtained from the same donors (all Caucasian) were analyzed for CYP3A5 protein content; the *CYP3AP1* and *CYP3A5* genotypes are defined in Figs. 1 and 2. CYP3A5 content ranged from undetectable levels to 31 pmol/mg of intestinal homogenate protein. NA, not available. Intestine ratio=1′-OH MDZ formation (pmol/min/mg)/4-OH MDZ formation (pmol/min/mg). Low activity, less than 25 pmol 1′-OH MDZ per min per mg homogenate protein (limit of quantitation).

likely to experience dose-limiting toxicities; and have different risks of diseases that are associated with the CYP3A5 expressor phenotype. Natural selection could drive the high frequency of *CYP3A5*1* alleles in African Americans and equatorial ethnic groups if, for example, there was a causal link between CYP3A5 expression and renal metabolism of endogenous molecules that affect Na^+ retention[37]. Such a link might confer a selective advantage in areas of water shortage.

Using the *CYP3A4*1B* allele as a marker, several investigators have postulated that *CYP3A4* is a candidate gene in the development of several disorders including prostate cancer and epipodophyllotoxin-induced secondary acute myelogenous leukemia[20,29,30]. Our finding of the simultaneous occurrence of *CYP3A4*1B*, *CYP3AP1*1* and *CYP3A5*1* alleles in some people indicates the need to re-evaluate this association and to determine whether the *CYP3A5* genotype and CYP3A5 expression are more closely associated with these disease risks.

Data on the frequencies of SNPs in human genes have been largely gathered from surveys of exons and introns[40,41], with little specific attention given to the frequency of SNPs in upstream regulatory regions, particularly those of differentially regulated gene family members. Previous studies have reported a nucleotide diversity of about 1 in 2,000 bp of noncoding DNA (consisting mostly of 5′ and 3′ UTRs and introns) and lower frequencies in coding sequences; the least frequent SNP, with a nucleotide diversity of approximately 1 in 20,000 bp, was a coding SNP that causes a nonconservative amino acid substitution[41]. In Caucasians, the observed nucleotide diversity in the *CYP3A4* promoter was 1 in 7,246 bp; in the *CYP3AP1* promoter, 1 in 2,577 bp; in the *CYP3A5* promoter, 1 in 3,175 bp; and in the *CYP3A7* promoter, 1 in 4,444 bp. In African Americans the observed nucleotide diversity in the *CYP3A4* promoter was 1 in 1,000 bp; in the *CYP3AP1* promoter, 1 in 649 bp; in the *CYP3A5* promoter, 1 in 2,841 bp; and in the *CYP3A7* promoter, 1 in 1,677 bp. It will be of interest to compare the level of nucleotide diversity within the *CYP3A* promoters with a much larger diverse survey of human gene promoters.

Traditionally, pharmacogenetics research used the strategy of identifying outliers in drug response, obtaining the DNA from outliers and identifying sequence variation in genes involved in that drug's metabolism. In the post-genomic era, reverse pharmacogenetics approaches are under way in which common polymorphisms are first identified in panels of DNA from anonymous people and then screening for these common sequence variations is carried out in people whose drug metabolism phenotype is known. For genes of pharmacogenetic interest, or for any gene, the fewer the number of DNA samples required to detect these common genetic variants, the more rapidly and cost effectively common variants can be identified. Moreover, these common variants are proposed to significantly contribute to risks of common disease[4]. Our results confirm that the smallest subset of eight DPDR samples was sufficient to identify all common polymorphisms in the *CYP3A* promoters and in the gene *CYP3A5*, including two polymorphisms that had a high probability of being missed in a random population sample of 25 people. This result indicates the robust power of this resource for rapid SNP discovery.

Methods

Population samples. We sequenced the *CYP3A* promoters of 159 people. Eight samples were from the DPDR (Coriell Cell Repositories). We also obtained samples from the following sources: 47 livers (27 Caucasian donors, 20 African American donors) whose midazolam 1′-hydroxylase and 4-hydroxylase activities and CYP3A5 and CYP3A4 protein were characterized (University of Washington, University of Pittsburgh, the Medical College of Virginia and St. Jude Children's Research Hospital); primary human hepatocytes from 8 people whose basal and drug-inducible testosterone 6β-hydroxylase activity had been characterized (University of Pittsburgh); kidneys from 12 people (Indiana University); 29 subjects (representatives of 7 families) whose nifedipine clearance had been determined (University of Newcastle upon Tyne, UK); 5 family members, one of whom was a poor nifedipine metabolizer (University of Michigan); 12 pediatric patients (St. Jude Children's Research Hospital); 14 subjects whose hepatic CYP3A phenotype had been determined by using the erythromycin breath test (University of Michigan); 10 subjects whose intestinal CYP3A4 content had been determined (University of Michigan); hepatocytes from 10 people for whom the drug induction of CYP3A protein had been characterized

Table 3 • Presence of *CYP3A4*1B* and *CYP3AP1*1* or *CYP3A4*1B* and *CYP3A5*1* in one person

Caucasians	*CYP3AP1*3/*3*	At least one *CYP3AP1*1* allele	*CYP3A5*3/*3*	At least one *CYP3A5*1* allele
*CYP3A4*1/*1*	63	10	14	3
*CYP3A4*1B*	1	2	0	2
African Americans				
*CYP3A4*1/*1*	11	3	3	3
*CYP3A4*1B*	3	15	2	9

CYP3AP1 and *CYP3A4* genotypes: Caucasians, χ^2=6.08; *P*=0.01; African Americans, χ^2=12.26; *P*=0.0004. *CYP3A4* and *CYP3A5* genotypes: Caucasians, Fishers exact test *P*=0.058; African Americans, Fishers exact test *P*=0.28.

(INSERM and Medical College of Virginia); 11 pairs of small intestine and liver (some liver samples are the same as those mentioned above) and intestines of 3 Caucasians whose midazolam 1′-hydroxylase and 4-hydroxylase activities and CYP3A5 protein had been characterized (University of Washington). We sequenced the *CYP3AP1* promoter and *CYP3A5* introns 3, 4 and 6 in samples from 10 Mexicans, 9 Japanese, 10 Chinese, 10 Southeast Asians, 6 Pacific Islanders and 5 Southwestern American Indians (Human Variation Panels, Coriell Cell Repositories). Informed consent was obtained from all people and approval of this research was obtained according to the relevant institutional guidelines.

Primers and sequencing. We used primer3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) to design primers to amplify the promoter regions of *CYP3A4*, *CYP3A5*, *CYP3AP1* and *CYP3A7*; the amplicons were 400–600 bp. The sequences of the primers were as follows: *CYP3A4*, 5′–TGGGATGAATTTCAAGTATTTTG–3′ (f) and 5′–AGGTTTCCATGGCCAAGTCT–3′ (r); *CYP3A4* primers to sequence the distal PXRE sequences (nt –7,836 to –7,208), 5′–CCGATCAGAATAAGGCATTG–3′ (f) and 5′–GATTCACCTGGGGTCAACAC–3′ (r); *CYP3AP1* primers, 5′–GGGGATGGATTTCAAGTATTCTG–3′ (f) and 5′–GTCCATCGCCACTTGCCTTCT–3′ (r); *CYP3A7* primers, 5′–GTCTGGCTGGGTATGAAAGG–3′ (f) and 5′–GCCAAGTTTGGGATGAGAT–3′ (r); *CYP3A5*, 5′–GAGGATGGATTTCAATTATTCTA–3′ (f) and 5′–GTCCATCGCCACTTTCCTTC–3′ (r). Forward and reverse primers were tailed with universal sequencing primers (–40 M13 and –28 M13, respectively). Primer pairs were used for 35 cycles to amplify genomic DNA. The following conditions were used in each cycle: 95 °C for 15 s, 61 °C for 30 s, and 72 °C for 5 min. We removed unincorporated nucleotides and primers by incubating the PCR product with shrimp alkaline phosphatase and exonuclease for 30 min at 37 °C followed by 15 min at 80 °C (ref. 42). Primers for amplification of the full-length *CYP3A5* cDNA were P1, 5′–AACAGCCCAGCAAACAGCAGC–3′ (f), and P2, 5′–TAAGCCCATCTTTATTTCAAGGT–3′ (r). Primers for sequencing the *CYP3A5* cDNA were as follows: P3, 5′–GTTGCTATTAGACTTGAGAGGACT–3′; P4, 5′–TGTAAGGATCTATGCTGTCCTTC–3′; P5, 5′–CACAAATCGAAGGTCTTTAGGC–3′; P6, 5′–TCAAAAACTGGGGTAAGGAATG–3′; P7, 5′–GCCTAAAGACCTTCGATTTGTG–3′; P8, 5′–CATTCCTTACCCCAGTTTTTGA–3′; P9, 5′–AGTCCTCTCAAGTCTAATAGCAAC–3′; P10, 5′–GAAGGACAGCATAGATCCTTACA–3′; P11, 5′–CAGGGTCTCTGGAAATTTGACA–3′; P12, 5′–TCATTCTCCACTTAGGGTTCCA–3′; and P13, 5′–CAGCATGGATGTGATTACTGGC–3′. The primers used to amplify *CYP3A5* exon 3B, 4B and 5B insertions from genomic DNA were 5020_22719, 5′–CCTGCCTTCAATTTTTCACTG–3′ (f), and 5020_24161, 5′–GCAATGTAGGAAGGAGGGCT–3′ (r). The nested primers used to sequence the nt 22,893 site were 5020_22743, 5′–TAATATTCTTTTTGATAATG–3′ (f), and 5020_23205, 5′–CATTCTTTCACTAGCACTGTTC–3′ (r). The nested sequencing primers used to sequence the nt 24,035 site were 5020_23761, 5′–CAACAAAAACCGGCAAACTG–3′ (f), and 5020_24135, 5′–AGGATTTTCAGACTTAACAC–3′ (r). The primers used to amplify the exon 7 deletion in *CYP3A5*6* were 5020_28814, 5′–GGTCATTGCTGTCTCCAACC–3′ (f), and the P6 primer (r), and to sequence across exon 7 5020_30237, 5′–TATGACTGGGCTCCTTGACC–3′ (f), and 5020_30745, 5′–TGGAATTGTACCTTTTAAGTGGA–3′ (r).

We sequenced the proximal promoters by performing standard fluorescence-based sequencing with Amersham ET Dye Primers. The distal promoter region of *CYP3A4* and the *CYP3A5* promoter were sequenced by using BigDye Terminator sequencing, and products were resolved by polyacrylamide gel electrophoresis or by capillary gel electrophoresis. The resultant trace files were base-called by phred and assembled by phrap (http://www.genome.washington.edu). Polyphred[42] was used to detect potential heterozygosity. To be a true variant, the variant-containing sequence generated by the forward primer had to be identical to that generated by the reverse primer.

Western-blot analysis. We performed quantitative immunoblotting of CYP3A5 and CYP3A4 content of tissue preparations as described[16]; purified cDNA-expressed CYP3A5 (a gift from R. Peter) and CYP3A4 purified from human liver were the reference standards[11]. Liver microsomes (10–20 μg) and intestinal homogenates (50 μg), both of which had been prepared from organ donor tissue[16], were resolved by electrophoresis, and CYP3A5 was detected with anti-CYP3A5 antibody (Gentest) or anti-CYP3A4 antibody[16]. We determined the integrated optical density of each band by using a ChemiDoc (Biorad) and Quantity One program.

Midazolam kinetic protocol. Midazolam, $^{15}N_3$-midazolam, 1′-OH MDZ and 4-OH MDZ were provided by Roche Laboratories. All incubations were performed in duplicate in solutions containing potassium phosphate (0.1 M, pH 7.4) and human liver microsomes (20–100 μg) or intestinal homogenates (50–100 μg) or recombinant CYP3A4 or CYP3A5 (10 pmol; Gentest). Midazolam (final concentration, 8 μM) was added to the diluted tissue preparations, and the mixtures were preincubated at 37 °C for 5 min. We added NADPH (final concentration, 1 mM) to initiate the reaction. Incubation of human liver microsomes lasted for 2 min; incubation of recombinant CYP3A and intestinal samples lasted for 4 min. Reactions were terminated by the addition of 1 ml of ice-cold Na_2CO_3 (0.1 M, pH ~11). We used NCI GC-MS to measure the quantities of extracted 1′-OH MDZ and 4-OH MDZ (ref. 16). Reaction velocities and product ratios are presented as mean±s.d. Statistical comparisons of mean data for different genotypes were performed by using a 2-sided, unpaired *t*-test with unequal variances.

RT–PCR of *CYP3A7*. Total RNA (5–10 μg) from human liver was reverse-transcribed according to the manufacturer's instructions (Life Technologies). CYP3A7 cDNA was amplified from first-strand cDNA using oligonucleotides CYP3A7(S), 5′–ATTCCAAGCTATGTTCTTCATCAT–3′, and CYP3A7(AS), 5′–AATCTACTTCCCCAGCACTGA–3′, under described conditions[43], except the initial denaturation lasted 5 min, the annealing temperature was 58 °C and the reaction required 25 cycles. The PCR product was analyzed on agarose gels. Amplification of 28S rRNA served as a control for RNA integrity.

GenBank accession numbers. *CYP3AP* promoter, S74700; *CYP3AP1*1* promoter consensus sequence, AF35929. Sequences of *CYP3A5*3* alternatively spliced exons, AF355800, AF 355804, AF355801 and AF355802, and *CYP3A5*6*, AF355803.

Acknowledgments

We thank E. Green for advice and use of facilities. This work was supported in part by grant GM60346 and the NIH/NIGMS Pharmacogenetics Research Network and Database (U01GM61374, http://pharmgkb.org/) under grant U01GM (GM61393) and by grants GM32165, ES08658, P30 CA21765, and CA51001 from the National Institutes of Health and by the American Lebanese Syrian Associated Charities (ALSAC).

Received 13 October 2000; accepted 21 February 2001.

1. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Collins, F.S., Brooks, L.D. & Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231 (1998).
4. Collins, F.S., Guyer, M.S. & Chakravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
5. Evans, W.E. & Relling, M.V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286**, 487–491 (1999).
6. Rendic, S. & DiCarlo, F.J. Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab. Rev.* **29**, 413–580 (1997).
7. Cholerton, S., Daly, A.K. & Idle, J.R. The role of individual human cytochromes P450 in drug metabolism and clinical response. *Trends Pharmacol. Sci.* **13**, 434–439 (1992).
8. Shimada, T., Yamazake, H., Mimura, M., Inui, Y. & Guengerich, F.P. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: Studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.* **270**, 414–423 (1994).
9. Watkins, P.B. Cyclosporine and liver transplantation: will the midazolam test make blood level monitoring obsolete? *Hepatology* **22**, 994–996 (1995).
10. Kivisto, K.T., Koremer, H.K. & Eichelbaum, M. The role of human cytochrome P450 enzymes in the metabolism of anticancer agents: implications for drug interactions. *Br. J. Clin. Pharmacol.* **40**, 523–530 (1995).
11. Huang, Z., Guengerich, F.P. & Kaminsky, L.S. 16α-hydroxylation of estrone by human cytochrome P4503A4/5. *Carcinogenesis* **19**, 867–872 (1998).
12. Finta, C. & Zaphiropoulos, P.G. The human cytochrome P4503A locus. Gene evolution by capture of downstream exons. *Gene* **260**, 13–23 (2000).
13. Schuetz, J.D., Beach, D.L. & Guzelian, P.S. Selective expression of cytochrome P450 CYP3A mRNAs in embryonic and adult human liver. *Pharmacogenetics* **4**, 11–20 (1994).
14. Wrighton, S.A., Ring, B.J., Watkins, P.B. & Vandenbranden, M. Identification of a

polymorphically expressed member of the human cytochrome P-450III family. *Mol. Pharmacol.* **36**, 97–105 (1989).
15. Schuetz, J.D., Molowa, D.T. & Guzelian, P.S. Characterisation of a cDNA encoding a new member of the glucocorticoid-responsive cytochromes P450 in human liver. *Arch. Biochem. Biophys.* **274**, 355–365 (1989).
16. Paine, M.F. *et al.* Characterization of interintestinal and intrainestinal variations in human CYP3A-dependent metabolism. *J. Pharmacol. Exp. Ther.* **283**, 1552–1562 (1997).
17. Paulussen, A. *et al.* Two linked mutations in transcriptional regulatory elements of the CYP3A5 gene constitute the major genetic determinant of polymorphic activity in humans. *Pharmacogenetics* **10**, 415–424 (2000).
18. Lehmann, J.M. *et al.* The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J. Clin. Invest.* **102**, 1–8 (1998).
19. Blumberg, B. *et al.* SXR, a novel steroid and xenobiotic sensing nuclear receptor. *Genes Dev.* **12**, 3195–3205 (1998).
20. Rebbeck, T.R., Jaffe, J.A., Walker, A.H., Wein, A.J. & Malkowicz, S.B. Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. *J. Natl. Cancer Inst.* **90**, 1225–1229 (1998).
21. Gonzalez, F.J. The molecular biology of cytochrome P450s. *Pharmacol. Rev.* **40**, 243–288 (1989).
22. Sata, F. *et al. CYP3A4* allelic variants with amino acid substitutions in exons 7 and 12: evidence for an allelic variant with altered catalytic activity. *Clin. Pharmacol. Ther.* **67**, 48–56 (2000).
23. Goodwin, B.J., Hodgson, E. & Liddle, C. The orphan human pregnane X receptor mediates the transcriptional activation of *CYP3A4* by rifampicin through a distal enhancer module. *Mol. Pharmacol.* **56**, 1329–1339 (1999).
24. Aoyama, T. *et al.* Cytochrome P-450 hPCN3, a novel cytochrome P-450 IIIA gene product that is differentially expressed in adult human liver. *J. Biol. Chem.* **264**, 10388–10395 (1989).
25. Maquat, L.E. When cells stop making sense: effects of nonsense codons on mRNA metabolism in vertebrate cells. *RNA* **1**, 453–465 (1995).
26. Maquat, L.E. The power of point mutations. *Nature Genet.* **27**, 5 (2001).
27. Jounaidi, Y., Hyrailles, V., Gervot, L. & Maurel, P. Detection of a CYP3A5 allelic variant: a candidate for the polymorphic expression of the protein? *Biochem. Biophys. Res. Commun.* **221**, 466–470 (1996).
28. Gorski, J.C., Hall, S.D., Jones, D.R., BandenBranden, M. & Wrighton, S.A. Regioselective biotransformation of midazolam by members of the human cytochrome P4503A (CYP3A) subfamily. *Biochem. Pharm.* **47**, 1643–1653 (1994).
29. Felix, C.A. *et al*. Association of CYP3A4 genotype with treatment-related leukemia. *Proc. Natl. Acad. Sci. USA* **95**, 13176–13181 (1998).
30. Paris, P.L. *et al.* Association between a CYP3A4 genetic variant and clinical presentation in African-American prostate cancer patients. *Cancer Epidemiol. Biomarkers Prev.* **8**, 901–905 (1999).
31. Westlind, A., Lofberg, L., Tindberg, N., Andersson, T.B. & Ingelman-Sundberg, M. Interindividual differences in hepatic expression of CYP3A4: Relationship to genetic polymorphism in the 5′-upstream regulatory region. *Biochem. Biophys. Res. Commun.* **259**, 201–205 (1999).
32. Ball, S.E. *et al.* Population disribution and effects on drug metabolism of a genetic variant in the 5′ promoter region of CYP3A4. *Clin. Pharmacol. Ther.* **66**, 288–294 (1999).
33. Hashimoto, H. *et al.* Gene structure of CYP3A4, an adult-specific form of cytochrome P450 in human livers, and its transcriptional control. *Eur. J. Biochem.* **218**, 585–595 (1993).
34. Cooper, T.A. & Mattox, W. The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.* **61**, 259–266 (1997).
35. Wrighton, S.A. *et al.* Studies on the expression and metabolic capabilities of human liver cytochrome P450IIIA5 (HLp3). *Mol. Pharmacol.* **38**, 207–213 (1990).
36. Gillam, E.M.J. *et al.* Expression of cytochrome P450 3A5 in Escherichia coli: effects of 5′ modification, purification, spectral characterization, reconstitution conditions, and catalytic activities. *Arch. Biochem. Biophys.* **317**, 374–384 (1995).
37. Grogan, W.M., Phillips, V.M., Schuetz, E.G., Guzelian, P.S. & Watlington, C.O. Corticosterone 6b-hydroxylase in A6 epithelia: a steroid-inducible cytochrome P-450. *Am. J. Physiol. Cell Physiol.* **258**, C480–C488 (1990).
38. Schuetz, E.G. *et al.* Expression of cytochrome P4503A in amphibian, rat and human kidney. *Arch. Biochem. Biophys.* **294**, 206–214 (1992).
39. Haehner, B.D. *et al.* Bimodal distribution of renal cytochrome P450 3A activity in humans. *Mol. Pharmacol.* **50**, 52–59 (1996).
40. Halushka, M.K. *et al.* Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genet.* **22**, 239–247 (1999).
41. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet* **22**, 231–238 (1999).
42. Nickerson, D.A., Tobe, V.O. & Taylor, S.L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* **25**, 2745–2751 (1997).
43. Schuetz, E.G., Beck, W.T. & Schuetz, J.D. Modulators and substrates of P-glycoprotein and cytochrome P4503A coordinately up-regulate these proteins in human colon carcinoma cells. *Mol. Pharmacol.* **49**, 311–318 (1996).
44. Nebert, D.W. Suggestions for the nomenclature of human alleles: relevance to ecogenetics, pharmacogenetics and molecular epidemiology. *Pharmacogenetics* **10**, 279–290 (2000).