Duplications and Defects in the CYP2A6 Gene: Identification, Genotyping, and In Vivo Effects on Smoking

YUSHU RAO, EWA HOFFMANN, MOHAMMAD ZIA, LAURENT BODIN, MARILYN ZEMAN, EDWARD M. SELLERS, and RACHEL F. TYNDALE

Centre for Addictions and Mental Health (Y.R., E.H., M.Zi., L.B., E.M.S., R.F.T.), Centre for Research in Women’s Health (M.Ze., E.M.S., R.F.T.), and Departments of Pharmacology (Y.R., E.H., M.Zi., L.B., E.M.S., R.F.T.), Psychiatry (E.M.S.), and Medicine (E.M.S.), University of Toronto, Toronto, Canada

Received November 8, 1999; accepted June 30, 2000 This paper is available online at http://www.molpharm.org

ABSTRACT

In humans, 80% of nicotine is metabolized to the inactive metabolite cotinine by the enzyme CYP2A6, which can also activate tobacco smoke procarcinogens (e.g., 4-(methylnitrosamo)-1-(3-pyridyl)-1-butanone). Previously, we demonstrated that individuals who are nicotine-dependent and have defective CYP2A6 alleles (*2, *3) smoked fewer cigarettes; however, we recognize that the genotyping method used for the CYP2A6*3 allele gave a high false-positive rate. In the current study we used improved genotyping methods to examine the effects of the defective CYP2A6*2 and CYP2A6*4 alleles on smoking behavior. We found that those with the defective alleles (N = 14) smoked fewer cigarettes per day than those homozygous (N = 277) for wild-type alleles (19 versus 28 cigarettes per day, P < .001). In addition, we identified a duplicated form of the CYP2A6 gene, corresponding to the gene deletion CYP2A6*4 allele, developed a genotyping assay, assessed the gene copy number, and examined its prevalence in Caucasian smokers (N = 296). We observed an ascending rank order for plasma cotinine and breath carbon monoxide levels (an index of smoke inhalation) in individuals with null (CYP2A6*2 and CYP2A6*4) alleles (N = 14), those homozygous for wild-type (CYP2A6*1/*1) alleles (N = 277), and those with our newly identified CYP2A6 gene duplication (N = 5). The phenotype, as determined by plasma nicotine/cotinine ratios, had a descending rank order for these three genotype groups that did not reach significance. Although further characterization is required for the duplication gene variant, these results extend our previous findings and suggest a substantial influence of CYP2A6 genotype and phenotype on smoking behavior.

Genetic variation of CYP2A6 alters coumarin and nicotine (NIC) metabolism (Yamano et al., 1990; Iscan et al., 1994; Messina et al., 1997). Initially, a wild-type (CYP2A6*1) and two defective alleles (CYP2A6*2 and CYP2A6*3) were identified. CYP2A6*2 is a null allele with no activity toward probe substrates, although the methodology for detection, function, and allele frequency of the CYP2A6*3 allele are controversial (Yamano et al., 1990; Fernandez-Salguero et al., 1995; Oscarson et al., 1998; Benowitz et al., 2000). Recently, a CYP2A6 gene deletion (CYP2A6*4) was characterized (Yokoi and Kamataki, 1998; Nunoya et al., 1999; Oscarson et al., 1999b); the mechanism proposed for the creation of the deleted allele is similar to that found for the deleted (CYP2D6*5) and duplicated (CYP2D6*2X2) alleles of CYP2D6, involving unequal crossover between CYP2D6 and adjacent CYP2D genes (Gaedigk et al., 1991). The existence of a CYP2A6 gene deletion variant infers the existence of a CYP2A6 gene duplication (Fig. 1A).

In humans, 80% of NIC is inactivated by metabolism to cotinine (COT; Benowitz et al., 1994). Determining the variation in NIC inactivation is important because of NIC's role in producing tobacco dependence and regulating smoking behavior. We, and others, have demonstrated that CYP2A6 is responsible for the majority of the metabolic inactivation of NIC to COT (Nakajima et al., 1996b, 2000; Messina et al., 1997; Benowitz et al., 2000) and for the metabolism of COT to trans-3-hydroxyCOT, 5'-hydroxyCOT and possibly norCOT (Nakajima et al., 1996a; Murphy et al., 1999).

Dependent smokers adjust their smoking behavior to maintain constant blood and brain NIC levels (McMorrow and Foxx, 1983; Russel, 1987). Consistent with this, we previously found that heterozygotes for defective (CYP2A6*2 or CYP2A6*3) alleles smoked fewer cigarettes (CIGs) per week than smokers homozygous for wild-type CYP2A6*1 alleles (129 versus 159 CIGs per week) and were less likely to become NIC dependent (Pianezza et al., 1998). Repeating the

ABBREVIATIONS: NIC, nicotine; CIG, cigarette; COT, cotinine; CO, carbon monoxide; PCR, polymerase chain reaction; dNTP, deoxyribonucleotide triphosphate; bp, base pair(s); ppm, parts per million.
CYP2A6*2 genotyping on these samples (Pianezza et al., 1998) with new techniques, an allele-specific assay (Oscarson et al., 1998) and a restriction digestion assay (Chen et al., 1999), demonstrated a conversion of nine individuals previously genotyped as CYP2A6*2/*2 to CYP2A6*1/*2 but no change in the individuals previously genotyped as CYP2A6*1/*2. The revised CYP2A6*2 allele frequencies were 2.7% in the never tobacco-dependent group (N = 184) and 2.1% in the tobacco-dependent group (N = 164), consistent with recent studies of Caucasians (2.3% (Chen et al., 1999); 1.1, 1.4, and 3.0% (Oscarson et al., 1998)).

It is clear that the original genotyping assay for CYP2A6*3 was inaccurate; it has been proposed that a gene conversion in the 3′-flanking region (in the position of the original reverse primer R4; Fernandez-Salgueiro et al., 1995), occurring in 30 to 40% of the CYP2A6*1 alleles, results in the CYP2A6*3 genotype misclassification (Oscarson et al., 1999a). Newer assays suggest that the frequency of the CYP2A6*3 is extremely low (0–0.7%, Chen et al., 1999; Oscarson et al., 1999b). Our kinetics data indicated that liver samples previously genotyped as having the CYP2A6*3 allele demonstrated slower NIC metabolism (Messina et al., 1997; R. F. Tyndale and E. M. Sellers, unpublished data). In addition, genotyping of samples in which CYP2A6*3 had been previously identified (Pianezza et al., 1998) for other variant alleles indicated that some of these samples contained the CYP2A6*4 allele; the identification of a CYP2A6*4 allele in an individual originally classified as having a CYP2A6*3 allele has also been observed by Oscarson et al. (1999a). In addition there appear to be a number of other nucleotide changes, and resultant amino acid changes, in the CYP2A6-coding region from these subjects which are currently being investigated (R. F. Tyndale and E. M. Sellers, unpublished data). These data suggest: 1) that some of the individuals who we previously genotyped as having the CYP2A6*3 allele may have alternative null allelic variants that may, or may not, account for our previous observations and 2) that we need to reassess the role of genetically variable CYP2A6 in the risk for tobacco dependence with larger numbers of subjects (because of the lower estimates of the allelic variants).

In this study, we focused on retesting the second observation from our previous study (Pianezza et al., 1998), which suggested that NIC-dependent (DSM-IV, American Psychiatric Association, 1994) individuals with CYP2A6*2 or CYP2A6*3 null alleles smoked fewer CIGs per day. Specifically, we demonstrated decreased CIGs per day and lower plasma COT and breath carbon monoxide (CO) levels in individuals with CYP2A6*2 or CYP2A6*4 null alleles. We also identified a putative CYP2A6 gene duplication variant, established a genotyping method for this variant, determined the allele frequencies, and then examined the impact of this novel variant on in vivo indices of smoking in Caucasians (N = 296).

Materials and Methods

**Primer and Sequencing.** Oligonucleotide primers for polymerase chain reaction (PCR) assays and DNA sequencing (Table 1) were synthesized by the Hospital for Sick Children Biotechnology Service Center (Toronto, Canada). Cosmid DNA from clones 19296, 19019, 17943, and 27292 (gratefully received from Dr. Linda Ashworth, Human Genome Center, Liverpool, CA) containing CYP2A6, CYP2A7, CYP2A7P, and CYP2A13, respectively (Hoffman et al., 1995) were used to test the CYP2A gene specificity of the primers and to obtain intronic DNA sequence. CYP2A gene-specific and nonspecific forward primers (exon and intron 1, 3, 6, 7, 8, and 9) and 3′-flanking reverse primers were used to amplify genomic DNA containing wild-type, CYP2A7/6 (CYP2A6*4 gene deletion), and CYP2A6/7 (CYP2A6 gene duplication) DNA. Sequencing was performed by the Core Molecular Biology Facility, York University (Toronto, Canada). Sequence alignments were performed using DNASIS for windows (Hitachi Software, Genetic Systems, San Francisco, CA).

**Subjects and Sampling.** All study protocols were approved by the Ethics Review Committee of the Sunnybrook and Women’s College Health Science Center. A structured questionnaire was used to obtain information concerning demographics, as well as history and pattern of psychoactive drug use [drug dependence assessment with DSM-IV (American Psychiatric Association, 1994)]. All subjects (N = 400) met the following criteria: 1) healthy male or female, 2) 16 to 70 years of age, 3) current smoker (50% light smokers, currently smoking <15 CIGs/day, and 50% heavy smokers, currently smoking >15 CIGs/day for each gender), and 4) willingness to sign the consent form. For this study we restricted the analysis to Caucasians (N = 296 of the 400 with three or more Caucasian grandparents) consisting of 155 female smokers (66 light and 89 heavy) and 141 male smokers (61 light and 80 heavy). Between 4 and 8 PM, subjects were assessed for breath CO with a Micro II Smokelyzer (Bedford Scientific Ltd., Upchurch, England). A single venous blood sample was acquired, and plasma NIC and COT were assessed by high-performance liquid chromatography (Pacifici et al., 1993). Genomic DNA was extracted from venous blood samples using the QIAamp Blood Kit (Qiagen Inc., Santa Clarita, CA). CYP2A6*2 and 4 assays were performed as previously described (Oscarson et al., 1998, 1999b).
CYP2A6 Gene Duplication Assay. A two-step genotyping assay was developed for the duplicated allele (Fig. 1B) based on inverting the gene specificity of the assay used for detecting the CYP2A6*4 allele (Oscarson et al., 1999b). Specifically, the first step used a forward primer with sequence common to both CYP2A6 and CYP2A7 in exon 7 (2Aex7F) with a 3'-flanking reverse primer that is CYP2A7 specific (2A7R1, analogous to 2A6R1, which is used for the CYP2A6*4 allele, Fig. 1B). The PCR reaction mixtures (25 μl) contained 0.25 μM each primer, 200 μM dNTPs, 1.8 mM MgCl₂, 1 U of Taq DNA polymerase (Gibco BRL, Life Technologies, Burlington, Ontario, Canada), and 50 ng of DNA. The reaction conditions were as follows: initial denaturation at 95°C for 1 min, followed by 35 cycles of denaturing at 95°C for 15 s, annealing at 60°C for 20 s, and extension at 72°C for 3 min, with a final extension of 7 min at 72°C. The second PCR step used nested gene-specific CYP2A7 (2A7ex8F) or CYP2A6 (2A6ex8F) forward primers with a nested CYP2A7-specific reverse primer (2A7R2) to identify the CYP2A7 wild type and CYP2A6 duplicated variants, respectively. The PCR reaction mixtures (25 μl) contained 0.25 μM each primer, 200 μM dNTPs, 1.8 mM MgCl₂, 1 U of Taq DNA polymerase (Gibco BRL, Life Technologies), and 1 μl of first step PCR-generated DNA. The reaction conditions were as follows: initial denaturation at 95°C for 1 min, followed by 15 cycles of denaturing at 95°C for 15 s, annealing at 44°C for 20 s, and extension at 72°C for 4.5 min, with a final extension of 10 min at 72°C.

A modified one-step assay for the detection of the duplication variant was also developed (Fig. 1C) that used a CYP2A7-specific (2A7ex8F) or CYP2A6-specific (2A6ex8F) forward primer with the CYP2A7 (2A7R1) reverse primer for detection of the wild-type CYP2A7 and duplicated CYP2A6 variants, respectively. The PCR reaction mixtures (25 μl) contained 0.25 μM each primer, 200 μM dNTPs, 1.8 mM MgCl₂, 1 U of Taq DNA polymerase (Gibco BRL, Life Technologies), and 50 ng of DNA. The reaction conditions were as follows: initial denaturation at 95°C for 1 min, followed by 35 cycles of denaturing at 95°C for 15 s, annealing at 44°C for 20 s, and extension at 72°C for 4.5 min, with a final extension of 7 min at 72°C. PCR products (8 μl, 1258 bp) were separated by gel electrophoresis on a 1.2% agarose gel stained with ethidium bromide. This assay can also be performed using the CYP2A7-specific (2A6ex8F) forward primer with sequence common to both CYP2A6 and CYP2A7 gene, we quantified the gene-specific (2A7R1, analogous to 2A6R1, which is used for the CYP2A6*4 gene, also CYP2A6*4, although not tested here) but not from the duplicated CYP2A6/7 gene, which has CYP2A7 3'-flanking sequence (the PCR product is 1883 bp). To assess the amount of 3'-flanking CYP2A6 DNA that was present in the samples, we amplified genomic DNA using a common forward primer (2Aex7F) with CYP2A6-specific (2A6R2) reverse primer. This primer pair amplifies DNA from the wild-type CYP2A6 gene (and also CYP2A6*4, although not tested here) but not from the duplicated CYP2A6/7 gene, which has CYP2A7 3'-flanking sequence (the PCR product is 1883 bp). To assess the amount of 3'-flanking CYP2A6 DNA that was present in the samples, we amplified genomic DNA using a common forward primer (2Aex7F) with CYP2A7-specific (2A7R2) reverse primer. This primer pair amplifies DNA from the wild-type CYP2A7 gene and also the CYP2A6/7 duplication variant. The PCR reaction mixtures (25 μl) contained 0.25 μM each primer, 200 μM dNTPs, 1.2 mM MgCl₂, 0.6 U of Taq DNA polymerase (Gibco BRL, Life Technologies), and 50 ng of DNA. The reaction conditions were as follows: initial denaturation at 94°C for 1 min, followed by 35 cycles of denaturing at 94°C for 15 s, annealing at 55°C for 20 s, and extension at 72°C for 3 min, with a final extension of 4 min at 72°C.

Conditions of linearity were established for each primer pair using serial dilutions of the cosmid clone containing CYP2A6 (for 2Aex7F with 2A6exR and 2AexF with 2A6R2 PCR reactions) and CYP2A7 (for 2Aex7F with 2A7R2 PCR reactions). Genomic CYP2A6, CYP2A7, and CYP2A13 DNA from the cosmid clones were used to confirm isozyme specificity of the PCR reactions and primer pairs. The reaction mixtures for each of the three sets of CYP2A primer pairs (assayed separately) were the same as used for the one-step genotyping assay. In a separate experiment we controlled for the amount and quality of the genomic DNA (50 ng) from the samples by amplifying the housekeeping gene β-actin (conditions from Tyndale et al., 1994); the assay linearity was established using a serial dilution of liver DNA (20 cycles of PCR used).

### TABLE 1

<table>
<thead>
<tr>
<th>Primer* Name</th>
<th>Sequence</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A6ex1F&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-GCT GAA CAC AGA GCA GCT GTA CA-3'</td>
<td>Exon 1</td>
</tr>
<tr>
<td>2A6ex3F&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-GGG TGG TAT TCA GCA ACG GG-3'</td>
<td>Exon 3</td>
</tr>
<tr>
<td>2A6ex5F&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-CTC CAC GAC GCC CT-3'</td>
<td>Exon 3</td>
</tr>
<tr>
<td>2A6ex8F&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5'-GAC GAA CAT GGC CTA GAT G-3'</td>
<td>Exon 7</td>
</tr>
<tr>
<td>2Aex7F&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5'-GGG CAA CAT GCC CTA CAT G-3'</td>
<td>Exon 7</td>
</tr>
<tr>
<td>2A6ex9F&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-CAC CTA AGG AGA TTC AGG TCC-3'</td>
<td>Intron 8</td>
</tr>
<tr>
<td>2A6ex9F&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-CAA AGG AGA TGA CGG CAC AGC-3'</td>
<td>Intron 8</td>
</tr>
<tr>
<td>2A6ex5R&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-GCA GCC TGG TTA TTT TTT TCC CA-3'</td>
<td>Intron 8</td>
</tr>
<tr>
<td>2A6ex4R&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-GTA GGT TGG TAT CCA GGA CTA GG-3'</td>
<td>Intron 8</td>
</tr>
<tr>
<td>2A6ex4R&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5'-TCC TGT TGG TTA TTT TTT TCC CA-3'</td>
<td>Intron 8</td>
</tr>
</tbody>
</table>

* F, forward upstream primer; R, reverse downstream primer.

<sup>a</sup> Primers from Oscarson et al. (1998).
<sup>b</sup> Primers from Fernandez-Salgueiro et al. (1995).
<sup>c</sup> Primers from Oscarson et al. (1999b).
After quantitative PCR conditions were established each sample was assessed with each primer pair two to three times on different days. In addition, using our duplication variant genotyping assay, we identified an additional 11 individuals in our database with the duplication variant (N = 16 total genotyped as homozygous CYP2A6*1/*1 plus the duplication variant); these individuals were included in the assessment of the amount of genomic CYP2A DNA, but their smoking demographics were not available.

Statistics. The null hypothesis was tested (i.e., increased number of CYP2A6 gene copies results in increased indices of smoking) by one-tailed t tests based on the pooled error term from a one-way ANOVA. Significance was set at P ≤ .05.

Results

Identification of a Duplication Variant. To test whether unequal crossover events between the CYP2A6 and CYP2A7 genes had occurred (Fig. 1A), resulting in deleted and duplicated CYP2A6 alleles, we amplified DNA from individuals with low and high NIC oxidase activity using CYP2A7 forward and CYP2A6 3' flanking reverse primers for deletion variants and CYP2A6 forward and CYP2A7 3' flanking reverse primers for duplication variants. Amplification products as well as the CYP2A6 and CYP2A7 cosmid clones were sequenced from exon 8 to 350 bp downstream of the stop codon (Fig. 2). The duplication crossover junction extends 219 bp upstream of the stop codon to 49 bp downstream of the stop codon (268 bp), in contrast to the crossover junction for the CYP2A6*4A deleted allele that occurs more than 106 bp downstream of the stop codon but consistent with the crossover position of the recently identified CYP2A6*4D allele (Oscarson et al., 1999a,b). The duplication crossover junction is defined by 15 positions of upstream sequence, which are identical with CYP2A6, and 35 downstream positions, which are identical with CYP2A7. It includes 4-bp positions (810, 819, 836, 892), which are uninformative because of reported CYP2A6 and CYP2A7 sequence polymorphisms (GenBank accession numbers: U22028, M33317, M33318; Nunoya et al., 1999; Oscarson et al., 1999a,b). Of note, after DNA sequencing, it was observed that two of the five DNA samples with the duplication variant contained a T at nucleotide 819 (Fig. 2) in contrast to the wild-type G, which would result in an amino acid change from glycine (GGC) 479 to valine (GTC). This is the same...
nucleotide change identified by Oscarson et al., (1999a), and when found in the CYP2A6 gene was referred to as the CYP2A6*5 variant; their paper suggests that this nucleotide alteration changes glycine 479 to a leucine amino acid, resulting in a null allele.

**Genotyping Assay for the Duplication Variant.** With an approach based on the CYP2A6*4 deletion assay (Fig. 1B; Oscarson et al., 1999b), we designed and tested a two-step genotyping assay for the detection of the CYP2A6*7 duplicated and CYP2A7 wild-type genes. Assay specificity was tested using DNA from individuals of known genotypes (CYP2A6*-1, *2, and *4), as well as sequenced duplication variants (Fig. 3A). We also developed a rapid one-step assay (Fig. 1C) for the wild-type CYP2A7 and duplicated CYP2A6 variants using the gene-specific exon 8 CYP2A6 or CYP2A7 forward primer paired with the gene-specific CYP2A7 R1 3' reverse primer (Fig. 3B). This assay can also be performed using the CYP2A7 R2 reverse primer. Using either the two-step or one-step assays we detected the duplicated variants but had no false-positive results from the samples without the duplication. However, we were able to detect a wild-type CYP2A7 gene product in a sample genotyped as CYP2A6*4/*4 (Fig. 3), which is not predicted from the scheme illustrated in Fig. 1. We assayed two other homozygous CYP2A6*4/*4 samples from our database and detected a CYP2A7 exon 8–3’ PCR product but no CYP2A6-coding region product (e.g., Fig. 4B).

**Is the Duplication Variant Present with, or instead of, the Wild-Type CYP2A6 Gene?** To assess whether the novel hybrid variant CYP2A6*7 that we had identified existed (as predicted, Fig. 1A) with the wild-type CYP2A6 gene, as opposed to replacing it, we determined the amount of DNA from the coding region of CYP2A6, as well as from the 3'-flanking region of CYP2A6 and CYP2A7, in individuals with the duplication variant and compared the amount of DNA to those with a homozygous wild-type genotype. Fig. 4A illustrates the standard curve for the PCR primer pair spanning exon 1 to 4 using a serial dilution of DNA from the cosmid clone containing the CYP2A6 gene. A typical ethidium stained agarose gel of the PCR products from this amplification is shown in Fig. 4B. Using this assay we measured the amount of DNA PCR product in five samples with the wild-type (CYP2A6*1/*1) genotype and five samples containing the duplication variant (Dup) as illustrated in Fig. 4B. Amplification using these primers with a CYP2A6*4/*4 homozygous individual or CYP2A7 cosmid clone as template DNA indicates gene specificity of the assay. For each sample assayed we also assessed the amplification of genomic DNA using β-actin primers. The standard curve for β-actin was created using dilution curves of hepatic genomic DNA and is illustrated in Fig. 4C.

When five samples in each group were reassayed six times (as illustrated in Fig. 4B), and adjusted for individual levels of β-actin, we found that those samples with the duplication had higher levels of CYP2A6-coding DNA (exon 1–4) relative to those with a wild-type genotype (247 ± 40 versus 122 ± 6 optical density units, respectively, P < .004). This provided the first evidence for a gene duplication event resulting in more copies of the coding region of CYP2A6 rather than the

---

**Fig. 3.** A, genotyping assay results from the two-step duplication genotyping method. The first step uses a forward primer (2A-ex7F) common to CYP2A6 and CYP2A7 and a CYP2A7-specific reverse primer (2ATR1). The second step utilizes CYP2A7-specific (2A7ex8F) or CYP2A6-specific (2A6ex8F) forward primers with a nested CYP2A7-specific primer (2ATR2) for detection of the wild-type CYP2A7 gene (first lane of each pair) or the duplicated CYP2A6 gene (second lane of each pair), respectively. The arrow indicates the 1180-bp second step PCR product (outer lanes show the 3054-, 2036-, 1636-, and 1018-bp molecular markers) B, the one-step assay combines the CYP2A7-specific (2A7-ex8F) or the CYP2A6-specific (2A6ex8F) forward primer with the CYP2A7-specific reverse primer (2ATR1). The first lane of each pair contains the CYP2A7 positive control and the second lane indicates detection of the duplication variant (arrow indicates the 1258-bp product).

**Fig. 4.** A, standard curve demonstrating a linear relationship between amount of CYP2A6 as template and detection of PCR product using exon 1–4 PCR primer pairs (see text for details). B, an ethidium-stained gel illustrating the PCR product formation using CYP2A6 exon 1 and 4 primer pairs. DNA samples of different genotypes, as well as a representative dilution curve of CYP2A6 cosmid genomic DNA (0.003–0.20 ng DNA), are shown. Samples with the wild-type genotype (*1/*1) have lower amounts of PCR product (50-ng DNA template) than samples with the duplication variant (Dup). The primers do not amplify DNA from the CYP2A7 gene, as indicated by the lanes containing DNA from the CYP2A7 cosmid clone (2A7) and a sample that is homozygous for the deletion variant (*4/*4) in contrast to the lanes containing the CYP2A6 cosmid DNA (2A6). C, a standard curve indicating the relationship between the amount of DNA (hepatic genomic DNA from two different samples, repeated three times) used and the amount of β-actin PCR product produced (for conditions see text). β-actin was used as a control for the DNA in the assessment of CYP2A copy number (Fig. 5).
novel variant being the result of a gene-conversion event with no increase in CYP2A6 gene copy. The ratio of PCR product was 4.2 (247:121) rather than the expected 3:2 that would be predicted if one assumes that the PCR product is derived from two copies in the CYP2A6*1/*1 group and two copies plus the duplicated allele in the other group. This suggested that additional duplicated copies may be present; hence, we repeated these studies in larger numbers of samples.

Having established the linearity of the assays, we screened our database for additional DNA samples containing CYP2A6 duplication variants. We used 28 samples that were homozygous CYP2A6*1/*1 and 16 samples from individuals who we genotyped as having the CYP2A6/7 hybrid duplication variant (five from the current data set and 11 that were identified in our database). β-actin PCR product was also determined using DNA from each sample. As expected from the postulated mechanism (Fig. 1A), we found that significantly more DNA was amplified, using primers for CYP2A6 exon 1 to 4, in the samples with the duplicated variant compared to those without it (Fig. 5A).

To further extend these observations, we amplified DNA with a selective CYP2A6 reverse primer in the 3'-flanking region that would not amplify the duplication variant. We found similar amounts of PCR DNA product for the 3'-flanking region of CYP2A6 in samples from both groups, consistent with similar copy numbers of the wild-type CYP2A6 gene being present in samples from both groups (Fig. 5B). This indicated that the source of the increased amount of CYP2A6-coding region DNA observed in samples genotyped as having the duplication variant (Fig. 5A) was most likely due to the duplication variant itself rather than to multiple copies of the wild-type CYP2A6 gene. When we amplified the 3'-flanking region of CYP2A7, we expected, and found, increased amounts of this PCR DNA product in the samples genotyped with the duplication variant, consistent with the duplicated CYP2A6 allele having a 3'-flanking region of CYP2A7 sequence (Fig. 5C). Together these data provide initial evidence that the duplicated CYP2A6/7 gene variant exists with a wild-type CYP2A6 gene on chromosome 19, rather than replacing it as an allelic variant per se. As with the postulated mechanism (Fig. 1A), we found that significantly more DNA was amplified, using primers for CYP2A6 exon 1–4 (using 2A-ex1F and 2A6R1 primers) relative to β-actin DNA (Fig. 5D, 0.4), suggesting that there may be more than one copy of the duplicated variant in all or some of the individuals genotyped with the duplication variant.

CYP2A6 Allelic Frequencies in Caucasian Smokers. We examined the frequency of the variant alleles in Caucasian smokers with no increase in CYP2A6 gene copy. The ratio of PCR product was 4:2 (247:121) rather than the expected 3:2 that would be predicted if one assumes that the PCR product is derived from two copies in the CYP2A6*1/*1 group and two copies plus the duplicated allele in the other group. This suggested that additional duplicated copies may be present; hence, we repeated these studies in larger numbers of samples.

The genotype frequencies for either CYP2A6*1/*1 or CYP2A6*2 were not significantly different from the genotype frequencies predicted by Hardy-Weinberg equilibrium. These individuals (with CYP2A6*2 or *4 alleles) were combined to form a decreased activity group 1 (N = 14, one or fewer active CYP2A6 allele). The majority of the smokers were CYP2A6*1/*1 individuals and constituted group 2 (N = 277, two active alleles). The CYP2A6-duplicated gene was detected in five persons, indicating a gene duplication prevalence of 1.7% (5/296) in this population (group 3, N = 5; three or more active copies). As mentioned above, two of the five samples with the duplicated gene contained a nucleotide change G819T, consistent with the previously identified mutation in the wild-type gene (CYP2A6*5; Oscarson et al., 1999a).
We also assessed the impact of CYP2A6 genetic variation on in vivo smoking with both the current smoking levels and the number of CIGs smoked during the heaviest period of life-time smoking. The reported number of current CIGs per day was significantly lower for those in group 1 (13.5 ± 2.3) compared to those in group 2 (19.5 ± 0.7, P < .05). This was also true for the period of heaviest regular smoking (19 versus 28, P < .001, group 1 versus group 2, respectively). However, people smoke CIGs with different intensities, and self-reported CIG numbers smoked may not be an accurate or robust smoking index; therefore, CO levels were used as an additional index of smoking behavior and inhalation. An ascending rank order was observed for the impact of the CYP2A6 genetic variation on CO levels (parts per million (ppm), Table 2; Fig. 6A). Group 1 (one or fewer active alleles) had significantly lower CO levels than did individuals homozygous for wild-type alleles (group 2) and individuals with duplicated CYP2A6 (group 3). To further examine smoking and kinetic variables among the three groups, we compared COT plasma levels (Table 2; Fig. 6B). Significantly higher COT levels were observed in group 3 compared with both the group of homozygous wild-type individuals (group 2) or group 1, those with at least one null allele. Although not reaching significance, the ratio of plasma NIC to COT (a measure of phenotype) followed a descending rank order with the highest ratio found in group 1 (0.20 ± 0.11); group 2 had intermediate ratios (0.12 ± 0.03) and group 3, with the duplicated variants, had the lowest ratios (0.09 ± 0.01).

However, although the individuals with duplicated alleles (group 3) had higher breath CO and COT levels indicating greater smoking behavior than the other two groups, they reported smoking fewer CIGs per day, both at the current time (mean, 13.3 ± 3.3; range, 6 to 25) and at the time of heaviest smoking (23 CIGs/day; range, 15 to 35), than would be expected. However, they appeared to smoke the CIGs more intensely, as suggested by an almost double CO to CIG ratio (2.1 ± 0.7) compared with either group 2 (1.2 ± 0.1) or group 1 (1.2 ± 0.2, P ≤ .05) and by higher plasma NIC to CIG ratios (group 3, 3.0 ± 1.0) compared with group 2 (1.6 ± 0.4) or group 1 (1.4 ± 0.3, P < .02) (current CIGs per day used to match current COT and CO results). Thus, group 3 appears to compensate for more CYP2A6 gene copies by increasing the intensity of smoking, whereas those individuals with null alleles and slower NIC metabolism (group 1) compensated by decreasing the number of CIGs per day, but smoking them with the same intensity as those individuals with normal CYP2A6 levels (group 2). There were no statistically significant differences in smoking demographics between those individuals with the duplication variant with, or without, the T819G mutation. For example, CO levels were 23 ppm in the individuals with the mutation (N = 2) and 22 ppm in individuals without the mutation (N = 3, P = .9).

### Table 2

Impact of CYP2A6 genotype on indices of smoking

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2A6*2/*2, N = 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*1/*1, N = 277</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*2/*2, N = 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*1/*1, N = 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*1/*1, N = 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>8/M = 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO (ppm)</td>
<td>14 ± 2</td>
<td>143/M = 144</td>
<td>4/M = 1</td>
</tr>
<tr>
<td>Cotinine plasma levels (ng/ml)</td>
<td>217 ± 76a</td>
<td>265 ± 9a</td>
<td>378 ± 76</td>
</tr>
</tbody>
</table>

F, female; M, male.

* Significantly different from group 1 (P < .05).

Discussion

CYP2A6-mediated coumarin 7-hydroxylase and NIC oxidase activities are highly variable (Yamano et al., 1990; Iscan et al., 1994; Messina et al., 1997), suggesting many CYP2A6 gene variants of increased and decreased activity may exist. To determine the molecular mechanisms involved in the very low and high CYP2A6 activity individuals, we searched for deleted and duplicated copies of the CYP2A6 gene, derived from unequal crossover of the CYP2A7 and CYP2A6 genes (Fig. 1A), analogous to the mechanism proposed for CYP2D6 (Gaedigk et al., 1991). We identified both deletion and duplication variants using a PCR approach and developed a genotyping assay for the duplicated variant to assess the impact of these variants on NIC metabolism and smoking.

The location of the crossover point between CYP2A7 and CYP2A6 in the deletion variant (CYP2A6*4A) is 106 to 201 bp downstream from the stop codon (Fig. 2; Nunoya et al., 1999; Oscarson et al., 1999a,b). In contrast, the duplication variant (CYP2A6/7) that we have identified is derived from an upstream unequal crossover that spans the stop codon and is consistent with the reported crossover point of the CYP2A6*4D variant (Oscarson et al., 1999a). In addition, this region contains the CYP2A6*1B and CYP2A6*5 variants, which are also thought to have arisen by unequal crossover between the CYP2A6 and CYP2A7 genes (Oscarson et al., 1999a). From exon 8 to the 3′-flanking region is highly conserved among the CYP2A6 genes (Fig. 2), making them good candidates for unequal crossover events; it is very possible that there are additional uncharacterized duplication and deletion variants with crossover positions in this same region. Therefore, we adapted the genotyping method of Oscarson et al. (1999b) to identify CYP2A6 gene duplications occurring from unequal crossover anywhere within this region. We propose that screening of large populations for the duplicated allele could be done with a one-step assay (Figs. 1C and 3B), followed by confirmation with the two-step assay (Figs. 1B and 3A). The current CYP2A6*2, *3, *4, *5 and duplicated alleles do not account for all of the metabolic outliers identified in our studies or those of other investigators (Benowitz et al., 2000), indicating that other CYP2A6 variant alleles exist.

Although the frequencies of the variant alleles in Caucasian smokers were low, we were able to detect an effect of the different genotypes on smoking indices. Our current data retest one portion of our previous findings (Pianezza et al., 1998). In the present study we demonstrate that CIG smok-
ers with CYP2A6 null alleles (+2 or +4) smoke fewer CIGs per day than do homozygous wild-type smokers both currently (13.5 ± 2.3 versus 19.5 ± 0.7, P < .03) and at the time of heaviest smoking (19 versus 29, P < .001). In addition, we have shown that they have lower breath CO levels (Fig. 6A), a measure that does not rely on self-report. They also have lower COT levels (Fig. 6B), which indicates decreased smoking and metabolism of NIC to COT. These data indicate a role for CYP2A6 gene variants in affecting smoking behavior, with slower metabolizers smoking less than faster metabolizers. We have confirmed these data independently using inhibition of CYP2A6 in smokers in vivo, observing a significant decrease in smoking (e.g., decreased CO levels, increased latency between CIGs) in the presence of a CYP2A6 inhibitor (Sellers et al., 2000a).

The individuals with duplicated CYP2A6 (group 3) smoked more, as evidenced by higher CO levels (Fig. 6A) and plasma COT levels (Fig. 6B); however, they reported fewer CIGs per day than expected, leading us to hypothesize that they may smoke more intensely rather than more frequently. This is supported by higher NIC/CIG and CO/CIG ratios than found in the other two groups. Despite the lack of controlled phenotyping conditions (i.e., dose and timing not controlled by investigator, inhalation route used avoiding the first pass effect that often contributes substantially to the metabolic ratio, single time-point plasma collection), the ratios of NIC/COT demonstrated a rank order (not significant), with the lowest ratio in the duplicated group 3, followed by the intermediate wild-type group 2 and the group with null allele carriers (group 1). This suggests that, with some refinement of the methodology, plasma NIC/COT ratios in smokers may be useful for finding those individuals with variant CYP2A6 alleles in the absence of investigator-administered NIC or coumarin. Our data suggest that the individuals in this study have between one and three copies of the duplicated variant as well as the wild-type CYP2A6 (Fig. 5). It is clear that further analysis of the variant duplicated allele is required, including formal in vivo NIC and coumarin kinetic studies and smoking demographic analysis of larger numbers of individuals, as well as Southern blotting and expression studies to clarify copy number and impact of the nucleotide changes.

Tobacco smoke contains a number of tobacco-specific procarcinogen nitrosamines, e.g., N-nitrosodiyethylamine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butane, and N'-nitrosonornicotine, that CYP2A6 can activate via α-hydroxylation (Crespi et al., 1990; Patten et al., 1997). Therefore, individuals who have CYP2A6 null alleles may also be less efficient at bioactivating tobacco smoke procarcinogens to carcinogens, whereas those with duplications may be more efficient. This is of particular interest because ethnic variations in frequencies of CYP2A6 variant alleles exist (Oscarson et al., 1998; Yokoi and Kamataki, 1998, 1999b) and may be related to the ethnic differences in lung cancer incidence and histology (Groeger et al., 1997). The role of CYP2A6 in levels of smoking and procarcinogen activation is supported by the recent study of Miyamoto et al. (1999), who found that having the CYP2A6*4 allele resulted in a significant reduction in risk for lung cancer. The decreased risk observed could be due to the gene's impact on amount smoked (decreasing exposure to procarcinogens) and/or on the decreased activation of procarcinogens. To examine the in vivo role of CYP2A6 in the activation of procarcinogens, we have blocked CYP2A6 activity in smokers using methoxsalen, a CYP2A6 inhibitor. Our preliminary data suggest a significant rerouting of the N-nitrosodiyethylamine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butane nitrosamines from the mutagenic α-hydroxylation pathways to the nonmutagenic 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol glucuronidation pathway (Sellers et al., 2000b).

In summary, we have demonstrated reduced smoking behavior (CO levels, CIGs per day, and COT levels) for those with fewer copies of the active CYP2A6 gene compared with individuals homozygous for the wild-type allele. We have also identified a putative CYP2A6 gene duplication and established a genotyping assay for its detection. Individuals with the duplication variant had higher breath CO and COT levels, suggesting higher levels of smoking, although they reported fewer CIGs per day, suggesting that they smoke each CIG with greater intensity (higher CO/CIG and NIC/CIG ratios). These data demonstrate that CYP2A6 gene variants exist that have an impact on smoking behavior, suggesting a significant role for CYP2A6 in smoke exposure and potentially in the etiology of tobacco-related cancers. Our data suggest also that mimicking the decreased activity variants by inhibiting the activity of CYP2A6 may produce the same benefits that are imparted by the null alleles, providing novel therapeutic approaches to prevention and treatment of tobacco smoking.

![Fig. 6. Impact of CYP2A6 genetic variation on breath CO levels (A) and plasma COT levels (B). Group 1 (one or fewer active gene copies) contains individuals with null alleles, group 2 (two active gene copies) contains CYP2A6*1/*1 individuals, and group 3 (three or more active gene copies) contains individuals with duplicated CYP2A6 gene variants. Data are given as mean ± S.E.M. Arrows indicate pairs that are significantly different (P ≤ .05).](attachment:image_url)
Acknowledgments

We thank Dr. Sharon Miksys for careful review of the paper and Dr. Howard Kaplan for help with data analysis. We are also grateful for the constructive comments made by the reviewers.

References


