Abstract

Human CYP2A13 is believed to be important in the metabolic activation of tobacco-specific nitrosamines in the respiratory tract; therefore, genetic polymorphisms of the CYP2A13 gene may be associated with interindividual differences in the risks of tobacco-related tumorigenesis. Our earlier studies identified a frequent single-nucleotide polymorphism in CYP2A13 exon 5, Arg257Cys, which led to an approximate 50% decrease in metabolic activities. In the present study, three additional coding region mutations (Arg25Gln, Arg101Stop, and Asp158Glu) and several mutations in the introns and flanking regions were identified in a Chinese patient population. Of particular interest is the Arg101Stop mutation, which was due to a C>T change in exon 2. Thus, individuals homozygous for this nonsense mutation would not have a functional CYP2A13 protein and, therefore, might have reduced sensitivity to xenobiotic toxicity resulting from CYP2A13-mediated metabolic activation in the respiratory tract. The frequencies of the coding region mutations were further examined using random samples of white, black, Hispanic, and Asian newborns from New York. The frequency of the Arg25Gln mutation in Asian newborns (9.6%) was very similar to that found in the Chinese population (10.9%). On the other hand, the Arg101Stop mutation was not detected in 136 newborn samples examined (23 white, 21 black, 19 Hispanic, and 73 Asian), suggesting that this mutation may be unique for the Chinese patient population. Haplotype analysis indicated that the Arg25Gln and Arg257Cys mutations are parts of a common haplotype. However, an additional haplotype that consists of the 25Gln but not the 257Cys allele was also identified.
Materials and Methods

DNA Samples Used for SNP Analysis. Human genomic DNA was isolated from 200 μl of whole blood from each subject using a QIAamp DNA Blood Mini kit (QIAGEN, Valencia, CA). All of the 32 subjects examined in this study were Chinese donors with various head and neck diseases: 16 individuals with laryngeal or laryngopharyngeal cancer, two with nasal tumors, five with nasal polyps, and nine with other nontumor head-and-neck diseases. The study was approved by the Institutional Review Boards at the participating institutions.

Amplification of CYP2A13 Gene by Long-Distance PCR. Primers shown in Table 1 were designed according to the CYP2A13 gene sequence from the completed human genome data base (GenBank accession no. AC008962). Primers 5'-F1 and 3'-R2 were used to amplify the full-length CYP2A13 gene. PCR amplification was carried out in a PerkinElmer thermal cycler 9600 instrument (Applied Biosystems, Foster City, CA). The reaction mixtures, in a total volume of 50 μl, contained about 50 to 200 ng of genomic DNA or 50 to 200 pg of PCR product as a template. 1.8 mM MgSO₄, 60 mM Tris-SO₄ (pH 9.1), 18 mM (NH₄)₂SO₄, 200 μM each dNTP, 200 nM each primer, and 2 μl of ELONGase Enzyme Mix (Invitrogen, Carlsbad, CA). After an initial denaturation at 94°C for 2 min, 35 cycles of amplification, each consisting of a denaturation at 94°C for 45 s, annealing at 65.4°C for 45 s, and an extension at 68°C for 30 min, were carried out, followed by a final extension at 68°C for 10 min. An H₂O blank (no template) control was routinely used for detecting potential contamination of reagents.

DNA Sequencing. All long-PCR products were gel-purified using a QIAquick Gel Extraction kit (QIAGEN) and then subjected to direct sequencing using an automated DNA sequencer from Applied Biosystems (model 3100) at the Molecular Genetics Core of the Wadsworth Center. Each sample was initially sequenced with a total of 10 different primers, as follows: 5'-F1, 5'-R1, Exon1F2, Exon2F2, Exon3F2, Exon5F, Exon6R, Exon7R, Exon8R, and Exon9R2 (see Table 1 for sequences). When PCR product from a single amplification was insufficient for all sequencing reactions, combined PCR products from two or more amplifications using the same genomic DNA sample or the first PCR product as a template were used. Mutations identified were confirmed by a second, independent PCR and by sequencing experiments using primers shown in Table 1, for detection of potential PCR and sequencing errors.

Determination of Allele Frequencies. The frequencies of exon 1, 2, and 3 variant alleles were determined using anonymous human genomic DNA samples isolated from newborn blood spots from the New York State Newborn Screening Program. Use of these specimens was approved by the Institutional Review Board at the New York State Department of Health. The methods of DNA preparation and of PCR amplification of DNA fragments covering exon 1, 2, and 3 have been described previously (Zhang et al., 2002). The lower and upper limits of the 95% confidence interval for the allele frequency were calculated, with a correction for continuity, using a program (http://faculty.vassar.edu/lowry/prop1.html) that is based on methods described by Newcombe (1998).

Haplotype Analysis. DNA samples containing heterozygous mutations at both exon 1 and exon 5 SNP sites were used as a template to amplify the full-length CYP2A13 gene using primers 5'-F1 and 3'-R2, as described above. The 9456-bp PCR product was used as a template in a nested PCR using primers 5'-F2 and 3'-R1, which amplify an 8723-bp fragment. The nested PCR mixtures contained, in a total volume of 50 μl, about 50 to 200 pg of template, 1.5 mM MgSO₄, 60 mM Tris-SO₄ (pH 9.1), 18 mM (NH₄)₂SO₄, 200 μM each dNTP, 200 nM each primer, and 1 μl of ELONGase Enzyme Mix. After an initial denaturation at 94°C for 2 min, 35 cycles of amplification, each consisting of a denaturation at 94°C for 20 s, annealing at 66°C for 1 min, and an extension at 68°C for 8 min, were carried out, followed by a final extension at 72°C for 30 min. The nested PCR products were purified by agarose gel electrophoresis, with use of crystal violet for staining DNA (Invitrogen), and cloned into the pcR-XL-TOPO plasmid vector (Invitrogen). Sequence analysis of individual clones, which represent single alleles, was carried out using primers 5'-R1, Exon1R, Exon5F, and Exon9R2.

Results

A genomic fragment containing the human CYP2A13 gene was amplified from each of 32 genomic DNA samples. The amplified fragment (Fig. 1) spans 9456 bp, starting at 1117 bp before the ATG start codon and ending at 869 bp after the TGA stop codon. All PCR products were then sequenced using primers as shown in Table 1. The specificity of the long-PCR primers was confirmed by the absence of any corresponding CYP2A6 or CYP2A7 sequences (data not shown). The sequenced regions covered about 1 kilobase before the ATG start codon, all exons and exon-intron junctions, and about 100 bp after the TGA stop codon. All mutations detected in the long-PCR product were confirmed by the sequencing of a second PCR product obtained in a separate experiment with the original genomic DNA as a template; this step indicated that the mutations did not result from PCR or sequencing errors.

A total of 14 SNPs were detected, as shown in Table 2; all variant alleles were detected as heterozygotes. Of the 14 SNPs, four are located in the coding region, whereas the other ten are located in noncoding regions: two in the 5'-flanking region, two in the 3'-flanking region, and one in the 5'-H11002 flanking region, two in the 3'-H11002 flanking region, and one in the 5'-H11002 flanking region.
untranslated region, and six in the introns. Five of the 14 SNPs had apparent allele frequencies of greater than 10%, whereas the other nine had apparent allele frequencies of 1.8 to 4.8% in this small sample.

Among the four SNPs detected in the coding region, one (Arg257Cys, caused by a 3375C > T missense mutation, with an apparent allele frequency of 10.9%) was the same as had been previously identified (Zhang et al., 2002), whereas the other three (one each in exons 1, 2, and 3) had not been previously reported. The SNP in exon 1 was a 74G > A missense mutation, leading to a predicted amino acid change from Arg to Gln at position 25. This SNP was detected in seven subjects, with an apparent allele frequency of 10.9%. The SNP detected in exon 2 was a 578C > T missense mutation, resulting in an amino acid change from an Arg to a stop codon at position 101. This SNP was detected in two subjects, with an apparent allele frequency of 3.2%. The SNP in exon 3, which was detected in only one subject (1.8%), was a missense mutation (1706C > G), which is expected to cause a conserved Asp158Glu substitution.

None of the 10 SNPs found in the noncoding regions had been detected in the previous study (Zhang et al., 2002). The two SNPs in the 5’-flanking region, −729C > T (at 729 bp before ATG) and −411G > A, had apparent allele frequencies of 4.7 and 15.6%, respectively, whereas the two SNPs in the 3’-untranslated region, 7520C > G and 7571G > C, had apparent allele frequencies of 4.7 and 10.9%, respectively. The other six SNPs were detected in introns 1, 2, 3, 5, 6, and 8 (Table 2), with apparent allele frequencies between 1.8 and 15.6%.

A rare mutation in exon 3, 1662G > C (Gly144Arg), previously detected in a Hispanic newborn (Zhang et al., 2002), and a silent mutation in exon 2, 523C > T, previously detected in an Asian newborn (Zhang et al., 2002), were not detected in these Chinese samples. The occurrence of the intron-5 and intron-7 mutations reported earlier (Zhang et al., 2002) was not examined in the present study.

The frequencies of the Arg25Gln, Arg101Stop, and Asp158Glu alleles were further examined using random samples of white, black, Hispanic, and Asian newborns in the New York State Newborn Screening Program. For Arg25Gln, nine heterozygotes and two homozygotes for the 74G > A mutation were detected among the 102 samples analyzed, with apparent allele frequencies of 2.1% (95% confidence interval, 0.1–12.5%) in white (n = 24), 11.5% (4.8–24.1%) in black (n = 26), 1.9% (0.1–11.6%) in Hispanic (n = 26), and 9.6% (3.6–21.8%) in Asian (n = 26) newborns. The frequency in the Asian newborns (9.6%) was very similar to that found in the Chinese samples (10.9%). On the other hand, the Arg101Stop mutation was not detected in 136 newborn samples examined (23 white, 21 black, 19 Hispanic, and 73 Asian), suggesting that this mutation may be unique for the Chinese patient population. For Asp158Glu, one heterozygote (Hispanic) was detected among 86 newborns examined (17 white, 21 black, 25 Hispanic, and 23 Asian), indicating that this allele, although infrequent, is not exclusively found in Chinese individuals.

The seven Chinese samples heterozygous for the Arg25Gln allele were also heterozygous for −411G > A, 3375C > T (exon 5, Arg257Cys), 7233T > G (intron 8), and 7571G > C (3’-untranslated region), suggesting that these mutations may be parts of a common haplotype. The potential linkage between the Arg25Gln and the Arg257Cys alleles was also apparent when the distribution of the Arg25Gln allele in the 102 newborn DNA samples described above was compared with that of the Arg257Cys allele (Table 3), which had been determined previously in the same samples by PCR-SSCP (Zhang et al., 2002). The data in Table 3 also indicate the occurrence of a haplotype (found in one individual) that consists of the 25Gln, but not the 275Cys allele. The linkage between the Arg25Gln and the Arg257Cys alleles was subsequently confirmed by sequencing subcloned long-PCR products that correspond to single chromosomes. Results from four Chinese DNA samples containing heterozygous mutations at both exon 1 and exon 5 SNP sites consistently showed that the 74G > A and 3375C > T alleles were on the same chromosome, as were the −411G > A, 7233T > G, and 7571G > C alleles.

**Discussion**

We have identified 13 additional SNPs in the CYP2A13 gene in the present study, which increases the total number of known variants in this gene to 20, including the seven variants detected in our previous study using SSCP (Zhang et al., 2002). Haplotype analysis indicated that the Arg25Gln and the Arg257Cys mutations are parts of a common haplotype. However, the Arg25Gln mutation was not detected in the earlier study, in which the Arg257Cys mutation was detected by SSCP at 7.7% in Asian, 1.9% in white, 14.4% in black, and 5.8% in Hispanic newborns (Zhang et al., 2002). It is likely that the 74G > A mutation in exon 1 could not be detected by SSCP under the conditions used in that study. Notably, six of the newly identified

---

**TABLE 2**

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Location</th>
<th>Predicted Amino Acid Change</th>
<th>Frequencya</th>
<th>αb</th>
</tr>
</thead>
<tbody>
<tr>
<td>−729C &gt; T</td>
<td>5’-Flanking region</td>
<td>No</td>
<td>4.7 (1.2–14.0)</td>
<td>32</td>
</tr>
<tr>
<td>−411G &gt; A</td>
<td>5’-Flanking region</td>
<td></td>
<td>15.6 (8.2–27.3)</td>
<td>32</td>
</tr>
<tr>
<td>74G &gt; A</td>
<td>Exon 1</td>
<td></td>
<td>10.9 (4.9–21.8)</td>
<td>32</td>
</tr>
<tr>
<td>260C &gt; A</td>
<td>Intron 1, 80 bp downstream of exon 1</td>
<td>No</td>
<td>4.7 (1.2–14.0)</td>
<td>32</td>
</tr>
<tr>
<td>578C &gt; T</td>
<td>Exon 2</td>
<td></td>
<td>3.2 (0.6–12.2)</td>
<td>31</td>
</tr>
<tr>
<td>672C &gt; A</td>
<td>Intron 2, 52 bp downstream of exon 2</td>
<td>No</td>
<td>4.8 (1.3–14.4)</td>
<td>31</td>
</tr>
<tr>
<td>1706C &gt; G</td>
<td>Exon 3</td>
<td></td>
<td>1.8 (0.1–10.8)</td>
<td>28</td>
</tr>
<tr>
<td>1757A &gt; G</td>
<td>Intron 3, 32 bp downstream of exon 3</td>
<td>No</td>
<td>1.8 (0.1–10.8)</td>
<td>28</td>
</tr>
<tr>
<td>3375C &gt; T</td>
<td>Exon 5</td>
<td></td>
<td>10.9 (4.9–21.8)</td>
<td>32</td>
</tr>
<tr>
<td>4947A &gt; G</td>
<td>Intron 5, 212 bp upstream of exon 6</td>
<td>No</td>
<td>4.7 (1.2–14.0)</td>
<td>32</td>
</tr>
<tr>
<td>5578A &gt; C</td>
<td>Intron 6, 196 bp upstream of exon 7</td>
<td>No</td>
<td>4.7 (1.2–14.0)</td>
<td>32</td>
</tr>
<tr>
<td>7233T &gt; G</td>
<td>Intron 8, 56 bp upstream of exon 9</td>
<td>No</td>
<td>15.6 (8.2–27.3)</td>
<td>32</td>
</tr>
<tr>
<td>7520C &gt; G</td>
<td>3’-Untranslated region</td>
<td>No</td>
<td>4.7 (1.2–14.0)</td>
<td>32</td>
</tr>
<tr>
<td>7571G &gt; C</td>
<td>3’-Untranslated region</td>
<td>No</td>
<td>10.9 (4.9–21.8)</td>
<td>32</td>
</tr>
</tbody>
</table>

a Apparent allele frequency. All mutations were detected as heterozygotes.
b Total number of samples analyzed.

95% confidence interval.
noncoding region SNPs, including −411G > A, 672C > A, 1757A > G, 7233T > G, 7520C > G, and 7571G > C, are also found in the dbSNP data base (www.ncbi.nlm.nih.gov).

The Arg257Cys mutation is known to cause a general reduction in CYP2A13 activity for all substrates tested, including NNK, 2′-methoxacetophenone, coumarin, hexamethylphosphoramide, N,N-dimethylalanine, and N-nitrosomethylphenylamine (Zhang et al., 2002), but it is not yet known whether the Arg25Gln mutation has any functional consequences. This mutation is unlikely to affect membrane retention or protein synthesis, since both Arg and Gln appear in orthologous CYP2A proteins (e.g., Arg in CYP2A13 and Gln in CYP2A6). Moreover, in CYP2C1, mutation of Lys21 (which appears to correspond to Arg25 of CYP2A13 according to a sequence alignment between CYP2A and CYP2C (Gotoh, 1992)) to Asn did not affect membrane retention of a green fluorescent protein reporter (Szczensa-Skorupa and Kemper, 2000). On the other hand, the effects of specific mutations in this linker region on catalytic activity have not been examined. Although this region is not part of the surface for interaction with NADPH-cytochrome P450 reductase or for substrate binding (Williams et al., 2000), and although the Arg25Gln change is unlikely to affect gross structure (since Gln25 occurs in functional CYP2As), this mutation does involve the loss of a positive charge. It remains to be determined whether such a change could have subtle effects on protein orientation and substrate access through the lipid bilayer. It will also be important to determine the impact of the Arg25Gln and Arg257Cys double mutations on CYP2A13 function since these mutations are linked in most cases. However, the Asp158Glu mutation in exon 3 is a conservative change and is less likely to have a significant impact on function.

The Arg101Stop mutation represents a null allele, since the prema-

ture stop codon will most likely lead to the synthesis of a truncated protein containing only the amino-terminal 100 residues, which could be unstable and would not be functional as a cytochrome P450 protein. The clear-cut functional consequence of this mutation will make it particularly interesting for future studies that corre-

late monooxygenase. The clear-cut functional consequence of this muta-

tion will make it particularly interesting for future studies that corre-

late CYP2A13 activity for all substrates tested, including NNK, 2′-methoxacetophenone, coumarin, hexamethylphosphoramide, N,N-dimethylalanine, and N-nitrosomethylphenylamine (Zhang et al., 2002), but it is not yet known whether the Arg25Gln mutation has any functional consequences. This mutation is unlikely to affect membrane retention or protein synthesis, since both Arg and Gln appear in orthologous CYP2A proteins (e.g., Arg in CYP2A13 and Gln in CYP2A6). Moreover, in CYP2C1, mutation of Lys21 (which appears to correspond to Arg25 of CYP2A13 according to a sequence alignment between CYP2A and CYP2C (Gotoh, 1992)) to Asn did not affect membrane retention of a green fluorescent protein reporter (Szczensa-Skorupa and Kemper, 2000). On the other hand, the effects of specific mutations in this linker region on catalytic activity have not been examined. Although this region is not part of the surface for interaction with NADPH-cytochrome P450 reductase or for substrate binding (Williams et al., 2000), and although the Arg25Gln change is unlikely to affect gross structure (since Gln25 occurs in functional CYP2As), this mutation does involve the loss of a positive charge. It remains to be determined whether such a change could have subtle effects on protein orientation and substrate access through the lipid bilayer. It will also be important to determine the impact of the Arg25Gln and Arg257Cys double mutations on CYP2A13 function since these mutations are linked in most cases. However, the Asp158Glu mutation in exon 3 is a conservative change and is less likely to have a significant impact on function.

The Arg101Stop mutation represents a null allele, since the prema-

ture stop codon will most likely lead to the synthesis of a truncated protein containing only the amino-terminal 100 residues, which could be unstable and would not be functional as a cytochrome P450 protein. The clear-cut functional consequence of this mutation will make it particularly interesting for future studies that corre-

late CYP2A13 activity for all substrates tested, including NNK, 2′-methoxacetophenone, coumarin, hexamethylphosphoramide, N,N-dimethylalanine, and N-nitrosomethylphenylamine (Zhang et al., 2002), but it is not yet known whether the Arg25Gln mutation has any functional consequences. This mutation is unlikely to affect membrane retention or protein synthesis, since both Arg and Gln appear in orthologous CYP2A proteins (e.g., Arg in CYP2A13 and Gln in CYP2A6). Moreover, in CYP2C1, mutation of Lys21 (which appears to correspond to Arg25 of CYP2A13 according to a sequence alignment between CYP2A and CYP2C (Gotoh, 1992)) to Asn did not affect membrane retention of a green fluorescent protein reporter (Szczensa-Skorupa and Kemper, 2000). On the other hand, the effects of specific mutations in this linker region on catalytic activity have not been examined. Although this region is not part of the surface for interaction with NADPH-cytochrome P450 reductase or for substrate binding (Williams et al., 2000), and although the Arg25Gln change is unlikely to affect gross structure (since Gln25 occurs in functional CYP2As), this mutation does involve the loss of a positive charge. It remains to be determined whether such a change could have subtle effects on protein orientation and substrate access through the lipid bilayer. It will also be important to determine the impact of the Arg25Gln and Arg257Cys double mutations on CYP2A13 function since these mutations are linked in most cases. However, the Asp158Glu mutation in exon 3 is a conservative change and is less likely to have a significant impact on function.

The Arg101Stop mutation represents a null allele, since the prema-

ture stop codon will most likely lead to the synthesis of a truncated protein containing only the amino-terminal 100 residues, which could be unstable and would not be functional as a cytochrome P450 monoxygenase. The clear-cut functional consequence of this mutation will make it particularly interesting for future studies that corre-

late CYP2A13 genotype with respiratory tract diseases. Individuals with the null allele may be at a reduced risk of chemical toxicity from compounds metabolically activated by CYP2A13 in the respiratory tract. Nonfunctional alleles have also been found in the CYP2A6 gene (e.g., Nunoya et al., 1998; Oscarson et al., 1999), which has been shown to play a major role in the metabolic disposition of nicotine and may thus influence the levels of exposure to tobacco-related carcinogens, such as NNK, in active smokers (e.g., Oscarson, 2001; Xu et al., 2002; Yoshida et al., 2002). In fact, male smokers homozygous for the CYP2A6*1 allele seem to have an elevated risk for tobacco-induced lung cancers (Ariyoshi et al., 2002). It is likely that individuals defective in both CYP2A6 and CYP2A13 genes may be further protected against tobacco-related toxicity in the lung.

Functional consequences of the noncoding region SNPs are more difficult to predict. Of the two SNPs detected in the 5′-flanking region, neither is located in a binding site for any known transcription factor, as determined by a search of the TRANSFAC data base (Heinemeyer et al., 1998) using the TFSEARCH program (Y.

### Acknowledgments

We gratefully acknowledge the use of the Molecular Genetics Core of the Wadsworth Center. We thank Drs. Laurence Kaminsky and Adriana Verschoor for reading the manuscipt.

### References


activation of a tobacco-specific carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone.  
*Cancer Res* 60:5074–5079.


