Short Communication

SINGLE NUCLEOTIDE POLYMORPHISMS OF THE HUMAN CYP2A13 GENE: EVIDENCE FOR A NULL ALLELE

(Received February 25, 2003; accepted May 15, 2003)

This article is available online at http://dmd.aspetjournals.org

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.

The human cytochrome P450 2A13 (CYP2A13) gene is located in the CYP2 gene cluster on chromosome 19 (Fernandez-Salgueiro et al., 1995; Hoffman et al., 2001). CYP2A13 is selectively expressed in the human respiratory tract (Koskela et al., 1999; Su et al., 2000) and is active in the metabolism of many xenobiotic compounds, such as 2′-methoxyacetophenone, 2,6-dichlorobenzonitrile, hexamethylphosphoramide, N,N-dimethylaniline, methyl tert-butyl ether, N-nitrosodimethylnitrosamine, and N-nitrosomethylphenylamine (Su et al., 2000). Of particular significance, CYP2A13 is the most efficient P450 enzyme known in the metabolic activation of a major tobacco-specific procarcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butane (NNK). Therefore, CYP2A13 may play important roles in xenobiotic toxicity and tobacco-related tumorigenesis in the respiratory tract.

Genetic polymorphisms that affect the metabolic activity or expression of biotransformation enzymes may be important contributors to interindividual differences in susceptibility to environmental diseases. In a previous study, we used polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) analysis to investigate the genetic polymorphisms of the CYP2A13 gene (Zhang et al., 2002). A total of seven variant alleles was detected, but only four represented single nucleotide polymorphisms (SNPs) (i.e., variants having allele frequencies greater than 1%), and only one SNP was detected in the coding region, in exon 5, leading to an Arg257Cys amino acid change. The Arg257Cys variant has been characterized following heterologous expression; it was 37 to 56% less active than the 257Arg protein toward all substrates tested in a reconstituted system, and it displayed a >2-fold decrease in catalytic efficiency toward NNK (Zhang et al., 2002).

Because PCR-SSCP may not detect all mutations, it is possible that additional variant alleles occur in the CYP2A13 gene, which remain to be identified by the use of more sensitive techniques. In the present study, we have developed a long-PCR protocol to amplify the CYP2A13 gene for sequence analysis. A total of 32 samples from Chinese donors was sequenced, which led to the identification of 13 previously unidentified SNPs in the CYP2A13 gene. Three SNPs in the coding region, one of which leads to a nonfunctional allele. The frequencies of the coding region mutations were further examined using random samples of white, black, Hispanic, and Asian newborns from New York. In addition, haplotype analysis was performed to identify linkages between the various mutations.
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 long CYP2A13 gene. PCR amplification was carried out in a PerkinElmer thermal cycler 9600 (Applied Biosystems, Foster City, CA). The reaction mixtures, in a volume of 50 μl, contained 10 mM Tris-SO₄ (pH 9.1), 1.5 mM MgSO₄, 60 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 8 min, were carried out, followed by a final extension at 72 °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 sequenced using primers as shown in Table 1. The sequenced regions covered about 1 kilobase before the ATG start codon, all exons and exon-intron junctions, and about 100 bp after any corresponding CYP2A6 or CYP2A7 sequences (data not shown). 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 10 are located in noncoding regions: two in the 5'-flanking region, two in the 3'-flanking region, and 10 in intronic regions.
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>T missense mutation, leading to a predicted amino acid change from Arg to Gln at position 101. This SNP was detected in two subjects, with an apparent allele frequency of 21.8% 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 (Zhang et al., 2002). 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 257Cys 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
The Arg257Cys mutation is known to cause a general reduction in CYP2A13 activity for all substrates tested, including NNK, 2′-methoxyacetylphenone, coumarin, hexamethylene phosphoramide, N,N-dimethylaniline, 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 (Szczesna-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 CYP2A, 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 premature 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 monooxygenase. The clear-cut functional consequence of this mutation will make it particularly interesting for future studies that correlate 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; Oscarnson 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., Oscarnson, 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. Akiyama: TFSEARCH: Searching Transcription Factor Binding Sites, http://www.crbc.jp/papia.html). None of the intron mutations occur at a known splicing site (Senapathy et al., 1990) or would generate a new splice site, as determined by an analysis using the Omiga program, version 2.0 (Accelrys, Cambridge, UK). The two SNPs in the 3′-untranslated region are not expected to affect RNA folding, according to an analysis using mfold, version 3.1 (http://www.bioinfo.rpi.edu/applications/mfold) (Mathews et al., 1999; Zuker et al., 1999). Further studies on the impact of these mutations on the expression of the CYP2A13 gene are warranted, since significant interindividual differences in the level of CYP2A proteins have been found in microsomes from fetal nasal mucosa (Gu et al., 2000).

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 manuscript.

References


activation of a tobacco-specific carcinogen, 4-(methylamino)-1-(3-pyridyl)-1-butanol.

Cancer Res 60:5074–5079.


