Published in final edited form as:

Pharmacogenet Genomics. 2009 November ; 19(11): 852–863. doi:10.1097/FPC.0b013e3283313aa5.

Mechanisms of Differential Expression of the *CYP2A13* 7520C and 7520G Alleles in Human Lung: Allelic Expression Analysis for *CYP2A13* Heterogeneous Nuclear RNA, and Evidence for the Involvement of Multiple *Cis*-Regulatory Single Nucleotide Polymorphisms

Hong Wu, Xiuling Zhang, Guoyu Ling, Jaime D'Agostino, and Xinxin Ding

Wadsworth Center, New York State Department of Health, and School of Public Health, State University of New York at Albany, Albany, NY 12201, USA

Abstract

Objective—To identify the mechanisms underlying the decreased allelic expression of a common *CYP2A13* allele (7520C>G) in the human lung; *CYP2A13* is expressed selectively in the respiratory tract, and is highly efficient in the metabolic activation of several chemical carcinogens.

Methods—The 7520C/G alleles were compared for mRNA stability in cells and relative heterogeneous nuclear RNA (hnRNA) levels in human lungs. Promoter region SNPs were identified and analyzed through in vitro reporter gene assays and gel-shift assays, in order to uncover the causative SNPs responsible for the decreased allelic expression.

Results—1) The 7520C>G SNP does not influence *CYP2A13* mRNA stability in *CYP2A13*-transfected human lung or nasal epithelial cells; 2) levels of the 7520G hnRNA were consistently lower (<10%) than the levels of the 7520C hnRNA in lung samples from nine heterozygous individuals; 3) three SNPs (–1479T>C, –3101T>G, and –7756G>A) in linkage disequilibrium with the 7520C>G variation were found to cause altered interaction with DNA-binding proteins and decreases in promoter activity; 4) the suppressive effects of the –1479T>C, –3101T>G, and –7756G>A SNPs on the *CYP2A13* promoter were additive, while the negative effects of the –1479T>C SNP were enhanced by methylation of the –1479C.

Conclusion—The decrease in expression of the 7520G allele was due to cumulative suppressive effects of multiple SNPs, with each by itself having a relatively small effect on *CYP2A13* transcription.

Keywords

CYP2A13; Single Nucleotide Polymorphisms; Allelic expression; Transcriptional regulation

Introduction

Cytochrome P450 (CYP or P450) enzymes play an important role in the metabolism of a variety of compounds, including therapeutic agents, environmental toxicants, and chemical

carcinogens [1,2]. Variations in P450 expression or activity can contribute substantially to inter-individual differences in therapeutic efficacy or toxicant susceptibility [3]. CYP2A13 is a human P450 enzyme selectively expressed in the respiratory tract [4-6]. CYP2A13 is the most active enzyme in the metabolic activation of a tobacco-specific carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), and is also active toward a number of other respiratory-tract toxicants [4,7-9]. CYP2A13 has been proposed to play a critical role in tobacco-related lung carcinogenesis [4,7]. The *CYP2A13* gene is polymorphic; a total of ~20 *CYP2A13* haplotypes have been identified to date (<http://www.cypalleles.ki.se/cyp2a13.htm>). A recent epidemiological study in a Chinese population indicated that an Arg257Cys variation in CYP2A13, leading to decreases in CYP2A13 enzyme activity [10], is associated with a reduced risk of smoking-induced lung adenocarcinoma [11].

Large inter-individual differences in the expression of CYP2A13 mRNA and protein in human lung have also been documented [6,12]. However, mechanisms that regulate CYP2A13 expression are largely unknown. In a recent study, we found that the transcriptional regulation of *CYP2A13* involves C/EBP transcription factors, and that CYP2A13 expression can be influenced by DNA methylation [13]. Furthermore, through an allelic expression analysis of CYP2A13 mRNAs from heterozygous individuals, we identified a low-expressing *CYP2A13* allele, marked by a single nucleotide polymorphism (SNP), 7520C>G, in the 3'-untranslated region (3'-UTR); mRNA from the 7520G allele was found to be expressed at 10-fold lower levels than mRNA from the 7520C allele [12]. This remarkable expression difference provides unique opportunities for identification of the mechanisms that are involved in the regulation of *CYP2A13* expression, as well as for further study of the potential associations between CYP2A13 expression and risks of smoking-induced lung cancer.

In the present study, we have performed a series of experiments to identify the mechanisms responsible for the low expression of the 7520G allele. Given the location of the 7520C>G in the 3'-UTR, we first tested the hypothesis that the 7520C>G change renders the CYP2A13 mRNA unstable. We then tested the alternative hypothesis that the allelic expression difference was due to a decrease in transcriptional activity of the 7520G allele. We developed a protocol for quantitative allelic expression analysis of heterogeneous nuclear RNA (hnRNA), the primary transcript; the underlying rationale is that abundance of hnRNA is an indicator of transcriptional activity of a given promoter in vivo [14]. Through allelic expression analysis of hnRNA, we confirmed that the 7520G allele had much reduced transcriptional activity in human lung, compared to the transcriptional activity of the 7520C allele. Subsequent studies were designed to identify additional SNPs, located in the 5'-flanking region of the *CYP2A13* gene, that are in linkage disequilibrium (LD) with the 7520C>G SNP, and that could be potentially responsible for the expression differences between the 7520C/G alleles. The 5'-flanking region SNPs were characterized using several experimental approaches, including computational analysis for potential transcription factor binding sites, gel-shift assays with methylated or un-methylated DNA probes, and reporter gene assays with single or composite SNP sites. Our results indicate that the decrease in expression of the 7520G allele was due at least partly to cumulative, suppressive effects of multiple SNPs, including -1479T>C, -3101T>G, and -7756G>A, with each by itself having a relatively small effect on *CYP2A13* transcription.

Methods

Plasmid construction

Detailed methods for the preparation of various expression vectors and reporter gene constructs are described in Supplemental Materials. The pCMV_2A13_UTRwt and pCMV_2A13_UTRmut expression plasmids contained the full-length CYP2A13*1 cDNA, with either 7520C or 7520G, respectively, in the 3'-UTR, preceded by a CMV promoter. The

p2A13_-1479T plasmid, containing a 2-kb *CYP2A13*1* promoter fragment in a pGL3_Basic (Promega, Madison, WI) luciferase reporter vector, was originally named p2A13_2k [13]; the p2A13_-1479C plasmid contains a C at -1479, instead of a T. The p2A13_-1479C and p2A13_-1479T constructs were methylated in vitro with CpG methylase M. SssI (New England Biolabs, Ipswich, MA), in the presence of 16 $\mu\text{mol/L}$ S-adenosylmethionine (New England Biolabs). The reaction was carried out at 37°C for 6 h. Complete methylation was confirmed by digestion with methylation-sensitive BstUI (New England Biolabs). The p2A13_-3101T, p2A13_-3101G, p2A13_-7756G, p2A13_-7756A, p2A13_-7756G/-3101T, p2A13_-7756A/-3101G, p2A13_-7756G/-3101T/-1479T, and p2A13_-7756A/-3101G/-1479C constructs each contained a 245-bp (-6 to -250) *CYP2A13*1* promoter fragment in the pGL3_Basic vector, with additional one, two, or three SNP-containing DNA fragments inserted further upstream of the 245-bp *CYP2A13* promoter. All promoter constructs were confirmed by DNA sequencing.

Preparation of DNA and RNA from human lung samples, and first-strand DNA synthesis

DNA and RNA were isolated from resected normal lung biopsy or autopsy tissues (adult tissues were provided by the National Cancer Institute Cooperative Human Tissue Network; fetal tissues were provided by the University of Washington Birth Defects Research Laboratory, Seattle, WA). Individuals selected for analysis were previously genotyped as heterozygotes for the 7520C>G variation [12]. DNA samples used for haplotype analysis, *cis*-variation identification, and methylation status analysis were isolated from normal lung tissues with the DNeasy tissue kit (Qiagen). Total RNA used for allelic expression analysis was isolated from the same individuals with use of an RNeasy kit (Qiagen). Prior to RT, total RNA samples were first treated with RNase-free DNase I (Invitrogen) for 15 min at room temperature, in order to eliminate potential genomic DNA contamination. The RNA samples were then reverse-transcribed into first-strand DNA with 2 μg of total RNA in a total volume of 20 μl RT reaction, using the SuperScript III first-strand synthesis system and either a random primer or an oligo dT primer (Invitrogen). Reverse transcriptase was omitted in negative control reactions.

Allelic expression analysis of hnRNA

The allelic expression analysis of *CYP2A13* (7520C/G) hnRNAs was carried out using the general strategy described previously for allelic expression analysis of *CYP2A13* (7520C/G) mRNAs [12], but with PCR primers that are specific for hnRNA. The first PCR was performed to amplify a 511-bp fragment (from intron 8 to 3'-UTR) of *CYP2A13* primary transcript, using primers 2A13HnI8F1 (5'-gaggatcggagtctcctctg-3') and 2A13R-RT2 (5'-tcacagctctgaaggacatc-3'). Reaction mixtures, in a total volume of 25 μl , contained 2 μl of RT product, 2.5 mM MgCl_2 , 0.32 μM of each primer, 0.2 mM dNTP, 2.5 μl 10 $\times$ PCR buffer and 2.5 U HotstarTaq polymerase (Qiagen). After a preincubation at 95°C for 15 min, the reaction was carried out for 55 cycles, with each cycle consisting of a denaturation at 95°C for 30 s, annealing at 64°C for 45 s, and extension at 72°C for 45 s, followed by a final extension at 72°C for 10 min. The second PCR was carried out using 2A13HnI8F2 (5'-tccctagagagtgacgccg-3') and 2A13R-RT1 (5'-tgttcctctaaccacctct-3') to amplify a 398-bp fragment. PCR was performed on the LightCycler using allele-specific hybridization probes designed and synthesized by TIB MolBiol (Adelphia, NJ), as described previously [12]. The reaction mixtures, in a total volume of 10 μl , contained 2 μl of 10-fold diluted 1st PCR product, 2.0 mM MgCl_2 , 0.3 μM each primer, 0.3 μM Sensor [C] or Sensor [G] probe, 0.3 μM Anchor probe, and 1 μl LightCycler FastStart hybridization probe reagent mix (Roche Applied Science). The second PCR and melting curve analysis were performed essentially as previously described, except that PCR reactions were carried out for 60 cycles (each cycle consisting of denaturation at 95°C for 10 s, annealing at 66°C for 10 s, and extension at 72°C for 15 s) [12]. Purified *CYP2A13* (7520C/G) fragments, at various molar ratios (10:1; 5:1; 3:1; 1:1; 1:3; as well as 7520C only and 7520G only), were used as PCR templates for generation of standard curves. For each sample, PCR

was run in quadruplicate and included negative control reactions, in which the reverse transcriptase was omitted during the RT step; the controls enabled us to detect potential contamination by genomic DNA.

Haplotype analysis for known *CYP2A13* SNPs, methylation analysis of *CYP2A13* promoter, and identification of additional SNPs in *CYP2A13* 5'-intergenic region

Molecular cloning and DNA sequencing were performed to identify *CYP2A13* haplotypes and known SNPs (between –2 kb and 3'-UTR) that are physically linked with the 7520G variation. Additional SNPs in *CYP2A13* 5'-intergenic region were identified by PCR-based sequencing analysis of genomic fragments from 7520C/G heterozygotes. Details of these studies are described in the Supplemental Materials. Procedures for methylation analysis of the *CYP2A13* promoter at the –1479 SNP site are also described in the Supplemental Materials.

Reporter gene assays in H441 cells

NCI-H441 human lung cancer cells (ATCC) were cultured at 37°C under 5% CO₂ in RPMI1640 medium (Invitrogen) supplemented with 10% fetal bovine serum. Cells (2×10^5) were seeded into 12-well plates 24 h before transfection. For transfection, cells were washed two times with serum-free MEM and incubated with Opti-MEM (Invitrogen), prior to the addition of a luciferase reporter construct or the pGL3-Basic vector control (0.5 or 0.8 µg), and an internal control reporter plasmid (pRL-TK or pRL-SV40; 0.1 µg), premixed with serum-free MEM containing Lipofectamine 2000 (Invitrogen) or Fugene 6 reagent (Roche Applied Sciences). After 6 h, the culture medium was replaced with RPMI1640 medium containing 10% fetal bovine serum. Luciferase activities were determined 24 h after transfection, using the Luciferase Reporter Assay System (Promega).

Gel Shift Assays

Nuclear extracts from H441 cells were prepared using a CellLytic NuCLEAR Extraction Kit (Sigma, Saint Louis, MO). The resulting nuclear pellet was resuspended in 20 mM HEPES, pH 7.9, 1.5 mM MgCl₂, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol, 1.0 mM DTT and a protease inhibitor cocktail (Sigma Cat.# P 8340). The nuclear extracts were stored in small aliquots at –80 °C before use. Nuclear extracts from fresh human lung tissues were prepared as described previously [13]. Oligonucleotide probes, including those with a methylated C at the –1479T>C SNP site, were synthesized by Integrated DNA Technologies. Complementary oligonucleotides were annealed, and the resulting double-stranded oligonucleotide probes were end-labeled with γ -³²P-ATP (Amersham Biosciences Corp., Piscataway, NJ). Gel shift assays were carried out by incubation of 4 or 5 µg of a nuclear extract with the γ -³²P-labeled DNA probes (~100,000 cpm) in a 25-µl incubation mixture containing 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, 1 mM EDTA, 12% glycerol, and 1.0 µg of poly(dI-dT) (Amersham Pharmacia Biotech). Unlabeled, double-stranded oligonucleotides were added, before the addition of a labeled probe, at 5-to 50-fold molar excess, as potential competitors. After incubation at room temperature for 30 min, the samples were loaded on 5% non-denaturing polyacrylamide gels that had been pre-run at 150 V for 1 h in Tris-glycine buffer. After electrophoresis at 180 V for 1 h, the gels were dried, and radioactive signals were visualized by exposure to an X-ray film (Kodak X-AR; Sigma) at –80°C for 24-36 h.

Results

The 7520C>G variation does not change the stability of *CYP2A13* mRNA

The 7520C>G variation, which was previously found to be strongly associated with a low allelic expression in human lung [12], is located in the last exon of the *CYP2A13* gene, and in the 3'-UTR of the *CYP2A13* mRNA. To determine the cause of the low allelic expression of

the *CYP2A13* 7520G alleles, we first tested the hypothesis that the low-expressing, 7520G-containing *CYP2A13* mRNA has lower stability than does the high-expressing, 7520C-containing *CYP2A13* mRNA. The rates of mRNA degradation were determined, following the addition of an RNA polymerase inhibitor, actinomycin D, in cells transiently transfected with an expression vector for *CYP2A13*. The results of this study (see Supplemental Materials) suggested that the 7520C>G change does not alter the stability of *CYP2A13* mRNA in human lung or nasal mucosa.

Decreased allelic expression of the *CYP2A13* 7520G allele occurred at the transcriptional level

To test the hypothesis that the decreased allelic expression of the *CYP2A13* 7520G allele is due to a lowered rate of transcription, we determined the relative abundances of the primary transcripts (hnRNA) for the 7520C and 7520G alleles. The strategy for allelic-expression analysis of *CYP2A13* hnRNA is outlined in Figure 1A. Because of the low abundance of hnRNAs, nested-PCR was carried out, to increase detection sensitivity. The allelic hnRNA expression analysis was performed for RNA samples isolated from normal lung tissues of nine 7520C/G subjects (Table 1); the tissue samples were previously found to exhibit substantially decreased expression of the 7520G allele at the mRNA level [12]. As shown in Table 1, the ratios of the relative hnRNA levels of the two alleles (7520C:7520G), as estimated from the peak-height ratios of the two respective PCR products, were greater than 10 for all nine lung samples tested using either the 7520C or the 7520G hybridization probe; this result indicated essentially monoallelic expression of the 7520C allele. An example of the melting curves of hnRNA-PCR products generated from the 7520C/G samples is shown in Figure 1 (B and C). In control experiments, PCR products derived either from genomic DNA of 7520C/G heterozygous lung samples (Fig. 1B) or from a 1:1 mixture of standard 7520C and 7520G templates (Fig. 1B and 1C), were also analyzed; these control samples each yielded two peaks representing the two alleles, in contrast to the single peak detected in the hnRNA-PCR products. In negative control experiments (not shown), in which the reverse transcriptase was omitted, no product was detected, indicating negligible levels of genomic DNA contamination in the reverse-transcribed cDNA preparation. Plots of the ratios of the maximum 7520G and 7520C peak heights on the melting curve versus the ratios of standard 7520G and 7520C template concentrations were linear (not shown), with R^2 values being greater than 0.997 for all analyses. Taken together, these data demonstrated that the decreased expression of the 7520G allele is manifested at the transcriptional, rather than the post-transcriptional, level.

Functional Analysis of the -1479T>C variation

Haplotype analysis for SNPs previously identified within the *CYP2A13* gene and the proximal 2-kb promoter region of 16 7520C/G heterozygotes [12] indicated that the 7520C>G variation is physically linked to -1479T>C and -1240A>G in the *CYP2A13**1H, *1N, *1P, and *8B alleles (see Supplemental Materials). Database analysis (TFsearch: <http://www.cbrc.jp/research/db/TFSEARCH.html>) for potential transcription factor binding sites in *CYP2A13* 5'-flanking region indicated that -1479T>C, but not -1240A>G, lies within the sequence that composes putative binding sites for several regulatory proteins, including the POU factor Brn-2 [15], at 81.2% identity, the cut-like homeodomain protein [16], at 79% identity, and the transcriptional repressor CDP [17], at 76.5% identity. Therefore, -1479T>C was selected for subsequent functional analysis.

To determine whether the -1479T>C change alters the binding of putative transcription factors, we performed gel-shift assays using nuclear extracts prepared from human lung. A broad, but faint, shift band, suggesting the occurrence of multiple protein binding complexes, was detected with the -1479T probe, whereas a single, strong, shift band, with a size approximating

that of the band detected by the –1479T probe, was detected with the –1479C probe (Fig. 2A). Binding specificity was confirmed by competition with excess unlabeled probe.

The –1479T>C change creates a new CpG site, a potential target for DNA methylation. To determine whether methylation of the allele-specific CpG site at –1479 alters the binding of transcription factors, we also performed gel-shift assays using labeled, methylated –1479C probe and corresponding unlabeled, methylated competitor (Fig. 2B). Similar to the unmethylated probe, the methylated probe formed specific DNA-protein complexes with human lung nuclear extract; however, the complexes formed appeared as faint duplex bands, in contrast to the single predominant band formed with the unmethylated –1479C probe (Fig. 2A). These results suggested that methylation of –1479C interferes with interactions of the site with nuclear proteins in human lung.

To confirm the anticipated negative effects of the –1479T>C variation on promoter activity, we performed reporter gene assays in H441 cells, a human lung adenocarcinoma epithelial cell line, using promoter constructs containing either –1479T or –1479C (Fig. 2C). Furthermore, to determine whether the methylation status of the allele-specific CpG site at –1479 affects *CYP2A13* promoter activity, we also conducted reporter gene assays using a methylated *CYP2A13* promoter construct containing either a T or a C at –1479 (Fig. 2C). We found the two constructs to be completely methylated in vitro, at all C-nucleotides, as confirmed by restriction analysis with a methylation-sensitive endonuclease BstU I, which cut the unmethylated constructs at multiple sites, but did not cut the methylated constructs (data not shown). As shown in Figure 2D, the –1479T>C variation led to a small (15%), but statistically significant ($p<0.05$), decrease in promoter activity in H441 cells, in the absence of promoter methylation. General C-methylation of the promoter constructs led to substantial decreases in promoter activity, regardless of whether the nucleotide at –1479 was a T or a C. However, the –1479T>C change was accompanied by a 32% decrease in promoter activity when the promoter was methylated; this extent of decrease was 2-fold higher than that seen with unmethylated promoter constructs.

To determine whether the allele-specific CpG site at –1479 is methylated in human lung, we first performed methylation-sensitive sequencing for DNA samples from four –1479T/C individuals. The bisulfite-induced change of un-methylated C to T was complete under the assay conditions used, as indicated by the total conversion of –1481C, –1472C, and –1471C to T (Fig. 3, panels A and B). However, –1479C persisted after bisulfite treatment (Fig. 3, panel B). We conclude that the –1479C was at least partially methylated in human lung.

Taken together, these findings indicate that the allele-specific CpG site at –1479, which is at least partially methylated in human lung, contributes to the decreased allelic expression of the *CYP2A13* gene. Nevertheless, it was clear that the observed extent of decrease in promoter activity could not totally account for the more than 10-fold difference in allelic expression of the 7520C/G alleles in human lung. Furthermore, the –1479T>C change was also a part of the *10 allele (see Supplemental Materials, Table S2); the individual harboring the *10 allele had a *10/*1N genotype, and the 7520G-containing *1N allele had lower expression than did the *10 allele. Therefore, the –1479T>C change alone appears insufficient to decrease the allelic expression to the extent observed for 7520G-containing alleles. Consequently, we considered, in our further studies, the possibility that additional *cis*-regulatory polymorphisms, located in the regions farther upstream of the 2-kb 5'-flanking region of the *CYP2A13* gene, contribute to the decreased expression of the 7520G allele.

Identification of additional polymorphisms linked to the 7520G allele

To identify additional polymorphisms in the *CYP2A13* 5'-flanking region, we sequenced the region from –1676 to –25723 bp, for three 7520C/G individuals (P2, P5, and A10 in Table 1)

and, as a control, one 7520C/C individual. A total of 10 common SNPs were identified for the three heterozygotes (Table 2); none of the 10 SNPs was found in the 7520C/C individual. Database analysis for potential transcription factor binding sites, with a threshold consensus score of 70, revealed that each of the 10 common SNPs could lead to allele-differential binding for one or more transcription factors, some of which are lung-enriched transcription factors (Table 2).

The 10 SNPs are all found in the DB-SNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), and two of them (–3101T>G and –7756G>A) are also represented on the HapMap (http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B36/). The –3101T>G and –7756G>A SNPs are in linkage disequilibrium (LD) (with a LOD score of 12.2; HapMap). Physical linkage between these SNPs and the 7520C/G variation has not been examined; the 7520C>G variation is not yet represented on the current version of HapMap. However, the common occurrence of the 10 SNPs among the three 7520C/G individuals, and the SNPs' absence in the 7520C/C individual, suggested that these polymorphisms are likely in LD with the 7520C>G variation. Further genotypic analysis of nine additional 7520C/G individuals and three additional 7520C/C individuals for the –3101T>G and –7756G>A SNP sites, sites for which allele frequencies had been found previously to vary from 0% to 29% in differing populations (HapMap), confirmed significant association among –3101T>G, –7756G>A, and 7520C>G SNPs ($P<0.001$; Fisher's Exact test). Taken together, these data suggested that the –3101T>G and –7756G>A SNPs, and most likely also the other eight common SNPs, are linked with the 7520C>G SNP.

Allele-differential binding of nuclear proteins to –3101T/G and –7756G/A in gel-shift assays

To determine whether the common SNPs identified in the 5'-flanking region alter the binding of transcription factors, we performed gel-shift assays for the 10 SNPs using nuclear extracts prepared from H441 cells, which are known to support *CYP2A13* expression after a combined treatment with AzaC and TSA [13]. Probes containing –3101T>G, –7756G>A (Fig 4), –4729 C>T, or –6495C>T (data not shown) revealed allelic differential binding by nuclear proteins, whereas probes containing the other SNPs either did not form any protein-binding complex under the conditions used, or else did not show differential binding by nuclear proteins (data not shown). The –3101T probe formed three binding complexes, whereas the –3101G probe generated binding complexes that were barely detectable (Fig. 4A), indicating greatly diminished binding affinity. The three complexes formed by the –3101T probe were all competed by the addition of a 50-fold excess amount of unlabeled probe (data not shown), which confirmed that the three complexes detected were all specific. Both –7756A and –7756G probes gave rise to a single gel-shift band; although the band intensity was lower for the –7756A probe than for the –7756G probe (Fig 4B), the allele-related differential binding was not as remarkable as was observed for the –3101T/G probes. In order to rule out the possibility that the difference in the band intensities is due to variations in labeling efficiency, rather than differences in intrinsic binding affinity, we performed reciprocal competition assays. We found that, no matter which labeled probe was used, the –7756G competitor was always more effective than was the –7756A competitor (Fig. 4B). The DNA probes for the –7756 SNP included sequences that resemble a cAMP-response element binding (CREB) sequence. However, an oligonucleotide containing the CREB consensus sequence was not able to compete for nuclear protein binding to the –7756 probes, a result indicating that other, unidentified transcription factors must be involved (data not shown).

Additive effects of -1479T>C, -3101T>G, and -7756G>A variants on CYP2A13 promoter activity

We used transient transfection and reporter gene assays to investigate the individual effects of the -3101T>G and -7756G>A variations, as well as combined effects of the -1479T>C, -3101T>G, and -7756G>A variations, on *CYP2A13* promoter activity in H441 cells. A DNA fragment containing one, two, or all three of the SNP sites was inserted upstream of a 245-bp *CYP2A13* promoter-luciferase vector [13], and the luciferase activity was then compared among cells transfected with differing *CYP2A13* promoter constructs. As shown in Figure 5, the construct containing -3101G or -7756A had, respectively, 19% or 25% lower promoter activities than did the construct containing -3101T or -7756G, whereas the construct containing the -3101G/-7756A double mutation had ~45% lower promoter activity than did the construct containing -3101T/-7756G. Furthermore, the construct containing the -1479C/-3101G/-7756A triple mutation had 56% lower activities than did the construct containing -1479T/-3101T/-7756G. Together with the aforementioned demonstration of allelic differential protein binding, these data suggested not only that the three SNPs are located in or near the sequence comprising a transcriptional factor binding site, but also that the sequence changes hinder the function of the relevant transcription factors. Moreover, the resultant negative effects of the three SNPs on *CYP2A13* promoter activity are additive.

Discussion

Allelic variation in gene expression appears to be common in the human genome [18,19]. For P450 genes, allelic differential expression has been reported for CYP3A4 [20] and CYP3A5 [21], in addition to CYP2A13 [12,22]. Knowledge of SNPs uniquely associated with a high- or low-expressing allele (i.e., marker SNPs) could provide a means by which to predict, based on genotyping analysis, the relative total expression levels of the gene in a study population. Furthermore, information on the identity of the SNPs that are responsible for the differing expression levels between two given alleles (i.e., causative SNPs) would shed light on the molecular mechanisms underlying the variable gene expression.

Previously, we had developed a PCR-based allelic expression assay, through which we discovered that the 7520C>G variation is associated with decreased (>10-fold) CYP2A13 allelic expression [12]. In the present studies, we have further extended the assay to allelic expression analysis of the CYP2A13 hnRNA, an approach that allowed us to conclude that the decreased expression of the 7520G allele is occurring at the level of transcription. Our real-time PCR-based method for quantitative allelic mRNA or hnRNA expression analysis combines melting-curve analysis with the use of allele-specific hybridization probes [12]. Several other methods have been developed for allelic mRNA expression analysis [e.g., 23, 24]. A semi-quantitative method for allelic hnRNA expression analysis was described recently for determination of relative expression levels of two CYP3A4 alleles in human liver [20]. Notably, allelic hnRNA expression analysis is technically challenging, given the much lower abundance of hnRNA than that of mRNA; however, it can substantially enhance the capability to perform allelic expression analysis, in that it makes it possible to determine relative expression levels for most, if not all, SNPs located either in the exons or in the introns. Additional discussions on allelic expression analysis are included in Supplemental Materials.

Our finding, that the 7520C>G change, which occurs in the 3'-UTR, does not lead to altered mRNA stability, suggests that the 7520C>G SNP is a marker for, rather than a cause of, the decreased expression of the 7520G allele. This notion is consistent with the finding of allelic hnRNA expression analysis, namely, that the decreased expression of the 7520G allele occurred at the level of transcription. Subsequent search for causative SNPs in the 5'-flanking region led to the identification of -1479T>C, -3101T>G, and -7756G>A, all of which are associated with altered DNA-protein binding in gel-shift assays and decreased *CYP2A13*

promoter activity in reporter gene assays. It is noteworthy that, although the extents of decreases in promoter activities in reporter constructs containing a single SNP were all relatively low, an additive effect was seen among -1479T>C, -3101T>G and -7756G>A, while the effect of the -1479T>C change was enhanced by methylation of the -1479C. Thus, the evidence supporting the designation of -1479T>C, -3101T>G, and -7756G>A as causative SNPs for the low expression of the 7520G allele is compelling, even though we cannot exclude the possibility that additional causative SNPs occur further upstream or elsewhere in the surrounding genomic regions.

Based on our present findings, we propose that the *CYP2A13* promoter in the 7520G allele is suppressed by the cumulative effects of several causative SNPs located in proximal, as well as distal, 5'-flanking regions of the gene. It seems reasonable to assume that occurrence of only a single causative SNP is insufficient to induce a substantial decrease in *CYP2A13* expression, an assumption supported by the fact that the -1479T>C change is also a part of the *1O allele; the latter, which does not contain the 7520C>G change, has higher expression than does the 7520G-containing *1N allele. It remains to be determined whether this model of transcriptional modulation by multiple *cis*-regulatory SNPs is operative in other *CYP2A13* alleles or other *CYP* genes. Unquestionably, for a mechanism involving the interplay of multiple SNPs, causality will be much harder to establish than in a situation in which only a single SNP is operative. Furthermore, the protein factors that interact with the identified causative SNPs have yet to be identified. In that context, it will be interesting to determine whether proteins binding at these three causative SNPs interact with the C/EBP factors that were recently found to activate regulatory elements at the proximal region of the *CYP2A13* promoter [13].

Our recent study [13], also demonstrated that DNA methylation is involved in the silencing of the *CYP2A13* gene in H441 cells, a result indicating that DNA methylation may play a role in transcriptional regulation of the *CYP2A13* gene. DNA methylation has been found to be associated with differential allelic expression in other studies [20,25], although, for methylation at a given site, it is not always obvious whether the methylation actually contributes to the differential allelic expression, or else whether the methylation is merely a consequence of decreased promoter activity. Here, we have discovered not only that the allele-specific CpG site at -1479 upstream of the *CYP2A13* gene, known to be associated with the low-expressing 7520G allele, is at least partially methylated in human lung tissues, but also that *in vitro* methylation at this site leads to altered nuclear protein binding in gel-shift assays, and a reduction in relative promoter activity in reporter gene assays. Therefore, methylation at the allele-specific -1479C likely contributed to the decreased expression of the 7520G allele.

In conclusion, we have successfully adapted our original allelic gene expression analysis method to the analysis of relative hnRNA levels. Through the application of this new approach, as well as through studies on mRNA stability, we have found that the *CYP2A13* 7520C>G variation is likely a marker, but not a causative SNP, for low *CYP2A13* expression. We have further identified three causative SNPs at the 5'-flanking region of the low-expressing *CYP2A13* 7520G allele; our results indicate that the substantially decreased expression of the 7520G allele is caused by probably additive effects of these genetic variations, as well as allele-specific methylation at the -1479C. The discovery of these causative SNPs could eventually lead to the identification of additional transcription factors involved in the regulation of *CYP2A13* expression, and the knowledge of the marker and causative SNPs for this common, low-expressing *CYP2A13* allele will facilitate future epidemiological studies designed to associate genetic variations that affect *CYP2A13* expression or function and lung cancer risks. In this context, there is a remarkable interindividual difference in the susceptibility to smoking-induced lung cancer, as exemplified by the fact that only one in nine smokers develops lung cancer [26]. *CYP2A13* expression levels in human lung biopsy samples also vary, by >100 fold [6,12]. Thus, for a determination of whether *CYP2A13* plays an important role in

procarcinogen activation and smoking-induced carcinogenesis in the human lung, efforts to identify the “expression phenotypes” for other common *CYP2A13* alleles are warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the use of the services of the Molecular Genetics Core Facility of the Wadsworth Center. We thank Dr. Adriana Verschoor for reading the manuscript.

This work was supported in part by U.S. Public Health Service grant CA092596 from the National Institutes of Health.

References

1. Coon MJ, Ding X, Pernecky SJ, Vaz AD. Cytochrome P450: Progress and Predictions. *FASEB J* 1992;6:669–673. [PubMed: 1537454]
2. Guengerich FP. Cytochrome P450: what have we learned and what are the future issues? *Drug Metab Rev* 2004;36:159–197. [PubMed: 15237850]
3. Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoeigenetic and clinical aspects. *Pharmacol Ther* 2007;116:496–526. [PubMed: 18001838]
4. Su T, Bao ZP, Zhang QY, Smith TJ, Hong JY, Ding X. Human cytochrome P450 CYP2A13: Predominant expression in the respiratory tract and its high efficiency metabolic activation of a tobacco-specific carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Cancer Res* 2000;60:5074–5079. [PubMed: 11016631]
5. Zhu LR, Thomas PE, Lu G, Reuhl KR, Yang GY, Wang LD, et al. CYP2A13 in human respiratory tissues and lung cancers: an immunohistochemical study with a new peptide-specific antibody. *Drug Metab Dispos* 2006;34:1672–1676. [PubMed: 16815959]
6. Zhang X, D'Agostino J, Wu H, Zhang Q Y, von Weymarn L, Murphy SE, et al. CYP2A13: variable expression and role in human lung microsomal metabolic activation of the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *J Pharmacol Exp Ther* 2007;323:570–578. [PubMed: 17671098]
7. Ding X, Kaminsky LS. Human extrahepatic cytochromes P450: Function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. *Annu Rev Pharmacol Toxicol* 2003;43:149–173. [PubMed: 12171978]
8. Jalas JR, Hecht SS, Murphy SE. Cytochrome P450 enzymes as catalysts of metabolism of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, a tobacco specific carcinogen. *Chem Res Toxicol* 2005;18:95–110. [PubMed: 15720112]
9. He XY, Tang LL, Wang SL, Cai QS, Wang JS, Hong JY. Efficient activation of aflatoxin B-1 by cytochrome P450 2A13, an enzyme predominantly expressed in human respiratory tract. *Int J Cancer* 2006;118:2665–2671. [PubMed: 16385575]
10. Zhang X, Su T, Zhang Q-Y, Gu J, Caggana M, Li HM, et al. Genetic polymorphisms of the human CYP2A13 gene: identification of single-nucleotide polymorphisms and functional characterization of an Arg257Cys variant. *J Pharmacol Exp Ther* 2002;302:416–423. [PubMed: 12130698]
11. Wang HJ, Tan W, Hao BT, Miao XP, Zhou GQ, He FH, et al. Substantial reduction in risk of lung adenocarcinoma associated with genetic polymorphism in CYP2A13, the most active cytochrome P450 for the metabolic activation of tobacco-specific carcinogen NNK. *Cancer Res* 2003;63:8057–8061. [PubMed: 14633739]
12. Zhang X, Caggana M, Cutler TL, Ding X. Development of a real-time polymerase chain reaction-based method for the measurement of relative allelic expression and identification of CYP2A13 alleles with decreased expression in human lung. *J Pharmacol Exp Ther* 2004;311:373–381. [PubMed: 15175424]

13. Ling G, Wei Y, Ding X. Transcriptional regulation of human CYP2A13 expression in the respiratory tract by CCAAT/enhancer binding protein and epigenetic modulation. *Mol Pharmacol* 2007;71:807–816. [PubMed: 17148654]
14. Sharp PA. Split genes and RNA splicing. *Cell* 1994;77:805–815. [PubMed: 7516265]
15. Ramkumar T, Adler S. A requirement for the POU transcription factor, Brn-2, in corticotropin-releasing hormone expression in a neuronal cell line. *Mol Endocrinol* 1999;13:1237–1248. [PubMed: 10446900]
16. Harada R, Berube G, Tamplin OJ, Denislrose C, Nepveu A. DNA-binding specificity of the cut repeats from the human cut-like protein. *Mol Cel Biol* 1995;15:129–140.
17. Andres V, Chiara MD, Mahdavi V. A new bipartite DNA-binding domain - cooperative interaction between the cut repeat and homeo domain of the cut homeo proteins. *Genes Development* 1994;8:245–257. [PubMed: 7905452]
18. Knight JC. Allele-specific gene expression uncovered. *Trends in Genetics* 2004;20:113–116. [PubMed: 15049300]
19. Lo HS, Wang ZN, Hu Y, Yang HH, Gere S, Buetow KH, et al. Allelic variation in gene expression is common in the human genome. *Genome Res* 2003;13:1855–1862. [PubMed: 12902379]
20. Hirota T, Ieiri I, Takane H, Maegawa S, Hosokawa M, Kobayashi K, et al. Allelic expression imbalance of the human CYP3A4 gene and individual phenotypic status. *Human Mol Genetics* 2004;13:2959–2969.
21. Wojnowski L, Brockmoller J. Single nucleotide polymorphism characterization by mRNA expression imbalance assessment. *Pharmacogenetics* 2004;14:267–269. [PubMed: 15083072]
22. D'Agostino J, Zhang X, Wu H, Ling G, Zhang Q-Y, Ding X. Characterization of CYP2A13*2, a variant cytochrome P450 allele previously found to be associated with decreased incidence of lung adenocarcinoma in smokers. *Drug Metab Dispos* 2008;36:2316–2323. [PubMed: 18669584]
23. Pant PVK, Tao H, Beilharz EJ, Ballinger DG, Cox DR, Frazer KA. Analysis of allelic differential expression in human white blood cells. *Genome Res* 2006;16:331–339. [PubMed: 16467561]
24. Matyas G, Giunta C, Steinmann B, Hossle JP, Hellwig R. Quantification of single nucleotide polymorphisms: a novel method that combines primer extension assay and capillary electrophoresis. *Hum Mutat* 2002;19:58–68. [PubMed: 11754104]
25. Polesskaya OO, Aston C, Sokolov BP. Allele C-specific methylation of the 5-HT2A receptor gene: Evidence for correlation with its expression and expression of DNA methylase DNMT1. *J Neurosci Res* 2006;83:362–373. [PubMed: 16358338]
26. Wood ME, Kelly K, Mullineaux LG, Bunn PA Jr. The inherited nature of lung cancer: a pilot study. *Lung Cancer* 2000;30:135–144. [PubMed: 11086207]

Fig. 1.

Allelic expression analysis for CYP2A13 hnRNA. A. Strategy for allelic expression analysis of CYP2A13 hnRNA. DNase-treated total RNA was reverse-transcribed (RT), and the resulting first-strand cDNA fragments that contain the CYP2A13 7520C/G variation were then amplified using PCR primers F1/R1 (2A13HnI8F1/2A13R-RT2), which serve to exclude the detection of CYP2A13 mRNA. Products from the first PCR were further amplified in a nested-PCR, again using PCR primers that exclude the detection of CYP2A13 mRNA (F2/R2: 2A13HnI8F2/2A13R-RT1). PCR products from the two differing alleles were detected through melting curve analysis using allele-specific hybridization probes. The relative abundance of each of the two allelic cDNAs, which reflects the relative expression levels of the primary

transcripts arising from the two alleles, was measured by the heights of the corresponding peaks in the melting-curve analysis. Selected exons are labeled at the top. B and C. Representative results of melting-curve analysis for CYP2A13 hnRNA-PCR products are shown for one human lung sample heterozygous for the 7520C>G variation. The PCR products were detected using either the 7520C probe (B) or the 7520G probe (C). The melting curve for PCR products derived from an equimolar mixture of standard 7520C and 7520G DNA templates (labeled as G:C=1:1) is included for comparison. In B, the melting curve for PCR products amplified from the genomic DNA of a 7520C/G individual is also shown (labeled as Genomic DNA). No template was added to control reactions (H₂O).

Fig. 2.

Functional analysis of the -1479T>C variation. A. Gel shift assays. Nuclear extracts (5 μ g) from human lung were incubated with 32 P-labeled -1479T, -1479C, or methylated -1479C (-1479MeC) oligonucleotide probes. Corresponding unlabeled probes were added at 50-fold excess. The position of the free probe (arrow) and that of the specific DNA-protein complexes (*) are shown. B. Nucleotide sequences of the probes used in panel A. The nucleotide at -1479 is underlined. C. Structure of the *CYP2A13* promoter constructs used in reporter gene assays. The constructs, prepared in the pGL3_Basic vector, and each containing a 2-kb *CYP2A13* promoter fragment, included p2A13_-1479T, p2A13_-1479C, methylated p2A13_-1479T, and methylated p2A13_-1479C. Filled circles represent methylated CpG sites. D. Results of luciferase reporter gene assays. H441 cells were harvested 24 h after they had been transfected with 0.8 μ g of a reporter gene construct and 0.1 μ g of the pRL-SV40 internal control construct. The relative luciferase activity was determined as described in Materials and Methods. The data represent mean values $\pm$ S.D. from three independent experiments, and they have been confirmed in additional experiments using a different batch of plasmid DNA preparation. The

statistical significance of differences between -1479T and -1479C constructs was assessed using Student's *t*-test.

Fig. 3.

Methylation analysis of the 7520G allele-specific CpG site at -1479. The sequence was determined for a 7520C/G lung sample (P5 in Table 1). Genomic DNA was either untreated (A), or pretreated with bisulfite to convert un-methylated C to T (B), before sequencing. Arrows indicate C-peaks at -1481, -1472, and -1471 in panel A, which became T-peaks in panel B, indicating the absence of C-methylation at these positions. * indicates the -1479 position, where a C/T double peak was detected. The results are typical of lung DNA samples from four individuals analyzed.

Fig. 4. Gel-shift assays using DNA probes containing -3101T>G (A) or -7756G>A (B) variations. Nuclear extracts (4 μ g) from H441 cells were incubated with 32 P-labeled oligonucleotide probes. Unlabeled probes were added at varying amounts as competitors, to demonstrate specific binding, where indicated. The positions of the free probe (arrow) and of the specific DNA-protein complexes (*) are indicated. C. Sequences of the probes used. The variant nucleotide at -3101 or -7756 is underlined.

D

Construct	Sequence
p2A13_-3101T/G	5'-ACTCCAGGAGGCCATGAAGGAT <u>T</u> /GAATTCCTATTACAGCTCCATGG-3'
p2A13_-7756G/A	5'-TTGCCCAAACCTGGCCTTGAACCTCT <u>G</u> /AACCTCAAGGAATCCACCTACCTTGG-3'
p2A13_-7756G/-3101T	5'-TTGCCCAAACCTGGCCTTGAACCTCT <u>G</u> AACCTCAAGGAATCCACCTACCTTGACTCCAGGAGGCCATGAAGGAT <u>T</u> AATTCCTATTACAGCTCCATGG-3'
p2A13_-7756A/-3101G	5'-TTGCCCAAACCTGGCCTTGAACCTCT <u>A</u> AACCTCAAGGAATCCACCTACCTTGACTCCAGGAGGCCATGAAGGAT <u>G</u> AATTCCTATTACAGCTCCATGG-3'
p2A13_-7756G/-3101T/-1479T	5'-TTGCCCAAACCTGGCCTTGAACCTCT <u>G</u> AACCTCAAGGAATCCACCTACCTTGACTCCAGGAGGCCATGAAGGAT <u>T</u> AATTCCTATTACAGCTCCATGGCTCATCTTCAATGAGGATGATCA <u>T</u> GATGAACCAACCTCACTGTTACAA-3'
p2A13_-7756A/-3101G/-1479C	5'-TTGCCCAAACCTGGCCTTGAACCTCT <u>A</u> AACCTCAAGGAATCCACCTACCTTGACTCCAGGAGGCCATGAAGGAT <u>G</u> AATTCCTATTACAGCTCCATGGCTCATCTTCAATGAGGATGATCA <u>C</u> GATGAACCAACCTCACTGTTACAA-3'

Fig. 5.

Reporter gene assays for *CYP2A13* promoter constructs containing -1479T>C, -3101T>G, and -7756G>A variations. A. Structure of the *CYP2A13* promoter constructs. The constructs, prepared in the pGL3_Basic vector, contained one, two, or three SNP-containing oligonucleotides (shown as shaded blocks; 47-50 bp each) inserted upstream of a 245-bp *CYP2A13* promoter fragment (p2A13). B and C. Results of reporter gene assays. H441 cells were harvested 24 h after they had been transfected with 0.5 µg of a reporter gene construct and 0.1 µg of the pRL-TK internal control reporter gene construct. The relative luciferase activity was determined as described in Materials and Methods. The data represent mean values ± S.D. from three independent experiments, and they have been confirmed in additional

experiments using a different batch of plasmid DNA preparation. Statistical significance of differences between two groups was assessed using Student's *t*-test. D. Sequences (sense strand only) of SNP-containing oligonucleotides used in each *CYP2A13* promoter construct. The SNP sites are underlined.

Table 1

Relative levels of CYP2A13 –7520C and CYP2A13 –7520G primary transcripts in human lung tissues^a

Tissue ID	Age	Gender	Ethnicity	Ratio of allelic hmRNA-PCR products (7520C:7520G)	
				7520C Probe	7520G Probe
P1	3 months	Male	Unknown	>10	>10
P2	15 months	Female	African-American	>10	>10
P3	31-week gestation	Female	Unknown	>10	>10
P4	3 months	Male	African-American	>10	>10
P5	5 months	Female	Unknown	>10	>10
A10	76 yr	Female	African-American	>10	>10
A11	72 yr	Male	Caucasian	>10	>10
A12	77 yr	Male	Caucasian	>10	>10
25293	53 yr	Female	Caucasian	>10	>10

^aReverse-transcribed hmRNA from five fetal or neonatal lung samples and four adult lung samples was used for CYP2A13 allele-specific expression analysis, as described in Figure 1. All samples were 7520C/G and 3375C/C (and thus not harboring the CYP2A13*2 allele). The ratios of 7520C:7520G hmRNA-PCR products were determined with use of both 7520C and 7520G hybridization probes.

Table 2Common SNPs identified in *CYP2A13* 7520C/G individuals ^a

SNP	DB-SNP accession number	Putative transcription factor binding sites	
		Wild type allele	Variant allele
−2739 G>C	rs305990	GATA-binding factor 2 (79.8)	C/EBP (73.1); AhR/Ar (71.6)
−3101 T>G	rs305991	Pbx-1 (77.5)	NF-kappaB (81.5)
−4729 C>T	rs305993	Runt-factor AML-1 (85.4)	AP-1 (73.5)
−6495 C>T	rs305997	Serum responsive factor (75.6)	GATA-2 (73.5)
−7080 G>T	rs305998	NF-kappaB (72.8)	AhR/Ar (73.1)
−7756 G>A	rs305999	CREBP (87.5)	COUP-1 (80.3)
−9102 G>T	rs306000	GATA-2 (76.7)	
−10873 C>A	rs306001	C/EBPb (72.3)	GATA-2 (87.2)
−11445 G>A	rs10416170	octamer-binding factor 1 (75.6)	apolipoprotein AI regulatory protein 1 (76.9)
−23947 A>G	rs11669312	AP-4 (71.4)	

^aSequences between −1676 and −25723 bp were determined for three 7520C/G individuals and one 7520C/C individual, and were compared to the analogous region in the reference “wild-type” sequence (GenBank accession number NC_000008). The A of the ATG start codon is designated +1. The DB-SNP accession numbers (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) for the identified SNPs are shown. Putative transcription factor binding sites were predicted using the TFsearch program (<http://www.cbrc.jp/research/db/TFSEARCH.html>) with a threshold value of 70. Numbers in parentheses are identity scores (%) for the sites found.