

Technical advance

Open Access

Rapid detection of the *CYP2A6*12* hybrid allele by Pyrosequencing[®] technology

Deborah A Koontz*¹, Jacqueline J Huckins², Antonina Spencer² and Margaret L Gallagher¹

Address: ¹Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Hwy NE, Atlanta, GA 30341, USA and ²Qiagen, Inc., 19300 Germantown Rd, Germantown, MD 20874, USA

Email: Deborah A Koontz* - duk5@cdc.gov; Jacqueline J Huckins - jacki.huckins@qiagen.com; Antonina Spencer - antonina.spencer@qiagen.com; Margaret L Gallagher - mxg2@cdc.gov

* Corresponding author

Published: 24 August 2009

Received: 11 June 2009

BMC Medical Genetics 2009, **10**:80 doi:10.1186/1471-2350-10-80

Accepted: 24 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2350/10/80>

© 2009 Koontz et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Identification of *CYP2A6* alleles associated with reduced enzyme activity is important in the study of inter-individual differences in drug metabolism. *CYP2A6*12* is a hybrid allele that results from unequal crossover between *CYP2A6* and *CYP2A7* genes. The 5' regulatory region and exons 1–2 are derived from *CYP2A7*, and exons 3–9 are derived from *CYP2A6*. Conventional methods for detection of *CYP2A6*12* consist of two-step PCR protocols that are laborious and unsuitable for high-throughput genotyping. We developed a rapid and accurate method to detect the *CYP2A6*12* allele by Pyrosequencing technology.

Methods: A single set of PCR primers was designed to specifically amplify both the *CYP2A6*1* wild-type allele and the *CYP2A6*12* hybrid allele. An internal Pyrosequencing primer was used to generate allele-specific sequence information, which detected homozygous wild-type, heterozygous hybrid, and homozygous hybrid alleles. We first validated the assay on 104 DNA samples that were also genotyped by conventional two-step PCR and by cycle sequencing. *CYP2A6*12* allele frequencies were then determined using the Pyrosequencing assay on 181 multi-ethnic DNA samples from subjects of African American, European Caucasian, Pacific Rim, and Hispanic descent. Finally, we streamlined the Pyrosequencing assay by integrating liquid handling robotics into the workflow.

Results: Pyrosequencing results demonstrated 100% concordance with conventional two-step PCR and cycle sequencing methods. Allele frequency data showed slightly higher prevalence of the *CYP2A6*12* allele in European Caucasians and Hispanics.

Conclusion: This Pyrosequencing assay proved to be a simple, rapid, and accurate alternative to conventional methods, which can be easily adapted to the needs of higher-throughput studies.

Background

Cytochrome P450 (CYP) is a diverse superfamily of enzymes that catalyzes the metabolism of a wide variety of compounds, including drugs and xenobiotics [1]. *CYP2A6*

is a member of this superfamily. It was first identified as the human coumarin 7-hydroxylase enzyme [2]. *CYP2A6* is also the primary enzyme responsible for the metabolism of nicotine to cotinine. Substantial inter-individual

and inter-ethnic differences in enzyme activity are thought to be attributable to genetics as well as to environmental factors [3]. Individuals lacking a functional *CYP2A6* gene have impaired nicotine metabolism and are therefore less prone to nicotine addiction and possibly tobacco-related cancers [4].

The *CYP2A6* gene is located on chromosome 19, adjacent to the inactive and highly homologous *CYP2A7* gene. Of the 17 different allelic variants in *CYP2A6* that are reported to abolish or decrease enzymatic activity, five are the result of unequal crossover and gene conversion events between *CYP2A6* and *CYP2A7* [5]. The hybrid *CYP2A6*12* allele is the product of such a crossover event; it contains the 5' regulatory region and exons 1–2 of the *CYP2A7* gene and exons 3–9 of the *CYP2A6* gene. The overall effect is a change in 10 amino acids and a reduction in enzyme activity [6].

The method currently used to genotype *CYP2A6*12* involves a two-step PCR [6]. The first step (PCR I) amplifies a 2.3 kb fragment encompassing a region from 240 bp upstream of the start codon to exon 3. This method uses a forward primer that binds to a site that is identical in both *CYP2A6* and *CYP2A7* gene sequences. However, amplification of the *CYP2A7* gene is prevented by use of a reverse primer that is specific for *CYP2A6* exon 3 sequence. Thus, the primer set can amplify both *CYP2A6*1* wild-type and *CYP2A6*12* variant but not the *CYP2A7* gene. The PCR I product serves as a template for two separate allele-specific PCR reactions (PCR II). Genotype determination is by gel electrophoresis of allele-specific products for *CYP2A6*1* and *CYP2A6*12*. Although accurate and reliable, this method is not practical for use in large population screening.

The aim of this study was to design a Pyrosequencing assay to genotype the *CYP2A6*12* hybrid allele that could be adapted to higher-throughput studies. The unique sequence structure of the hybrid allele lends itself to this technology. The Pyrosequencing method involves solution-based fluorescent real-time DNA sequencing whereby a sequencing primer is hybridized to a single-stranded PCR amplicon that serves as a template for sequential polymerase-catalyzed incorporation of nucleotides. Each incorporation event facilitates a cascade of enzymatic reactions that generates light displayed as a series of peaks (Pyrogram® traces) representative of the nucleotide sequence. The Pyrosequencing method has been used for many applications, such as SNP genotyping [7], DNA methylation analysis [8], and gene copy number assessment [9,10].

Methods

Subjects

DNA samples from 90 individuals of the Polymorphism Discovery Resource (PDR) and 14 individuals of a multi-generational Centre d'Etude du Polymorphisme Humain (CEPH)/Utah family (Coriell Cell Repositories, Camden, NJ) were used to optimize and validate the Pyrosequencing assay. An additional 181 samples from the Coriell SNP500 panel and the NIGMS Human Variation panels HD24AA, HD24EC, HD24CHI, and HD07 Japanese <http://ccr.coriell.org/Sections/Collections/NIGMS/Populations.aspx?PgId=177> were genotyped using the validated Pyrosequencing assay to assess multi-ethnic variation in frequency of the *CYP2A6*12* allele. The combined panels included African Americans (n = 47), European Caucasians (n = 48), Pacific Rim (n = 54), and Hispanics (n = 32). The repository retains no links to individuals, and the Coriell Institutional Review Board reviews the procedures of the repository annually. In addition, 7,159 participants aged 12 years or older from Phase 2 (1991–1994) of the Third National Health and Nutrition Examination Survey (NHANES III) [11] were screened using the Pyrosequencing assay to facilitate automation into the workflow. Use of these samples for this study was approved by the Centers for Disease Control and Prevention (CDC)/National Center of Health Statistics (NCHS) Institutional Review Board. Since the initial informed consent for use of participant blood for research did not include genetics, the CDC/NCHS Ethics Review Board approved a revised plan in 2001 which allows for the linkage of genetic results to NHANES data through the NCHS Research Data Center to ensure that confidentiality of study participant's identity is maintained [12].

Assay Design and Validation

The Pyrosequencing assay was designed on the basis of sequences for the *CYP2A6* and *CYP2A7* genes obtained from the Ensembl Gene ID ENSG00000213052 and ENSG00000198077 databases, respectively [13]. The two sequences were aligned from 240 bp upstream of the start codon up to and including 25 bp of exon 3 to locate suitable PCR primer regions to specifically amplify *CYP2A6*1* and *CYP2A6*12* alleles but not the *CYP2A7* gene. The forward PCR primer is located upstream of exon 2 and targets a region that is identical in both *CYP2A6* and *CYP2A7* sequences. The 3'-biotinylated reverse primer is specific to *CYP2A6* sequence and binds to the beginning of exon 3. The reverse primer does not bind to the *CYP2A7* sequence because of the presence of two neighboring nucleotide mismatches in the primer binding site. The specificities of the PCR primers and the primer melting temperatures were verified by use of the NCBI BLAST server <http://www.ncbi.nlm.nih.gov/BLAST/>.

The internal forward sequencing primer is nested just downstream of the forward PCR primer in intron 1. It contains two degenerate nucleotides to enable the primer to hybridize to both *CYP2A6*1* and *CYP2A6*12*-specific PCR products that differ from each other by two nucleotides at the primer binding site. The sequencing primer is positioned to interrogate two key nucleotide sites that allow for clear distinction between *CYP2A6*1* and *CYP2A6*12* alleles. Primer location and sequences are described in Figure 1.

PCR amplification was performed by use of the HotStarTaq Master Mix kit with Q solution (Qiagen, Valencia, CA). The 25 µl PCR reaction contained 20 ng DNA, 0.4 µM of each primer, 1× HotStarTaq Master Mix, and 1× Q-solution. Amplification was performed with initial denaturation at 95°C for 15 min, followed by 40 cycles of 95°C for 30 sec, 58°C for 30 sec, 72°C for 90 sec, and a final 72°C extension for 2 min 30 sec. PCR products (15

ul) were processed in a 96-well format for Pyrosequencing analysis by use of a Vacuum Prep Workstation, following the standard manufacturer's protocol (Biotage AB, Uppsala, Sweden). Single-stranded DNA was prepared by capture of the biotinylated strands onto streptavidin-coated Sepharose beads (GE Healthcare, Piscataway, NJ) with subsequent removal of non-biotinylated single strands. The bead-DNA complex was treated with ethanol, denaturation buffer, and wash buffer and released into a PyroMark Q96 plate containing a mixture of annealing buffer and 0.8 µM sequencing primer. The plate was incubated at 80°C on a heat block for 2 minutes, followed by a slow cool for an additional 5 minutes by turning off the heat block. The plate was transferred directly into a Biotage PSQ 96HSA System for sequence determination. Data were analyzed with PSQ HS 96 SNP software that automatically scores genotypes by using a novel SNP algorithm based on comparison of real-time sequence output (Pyrogram traces) with a theoretical histogram.

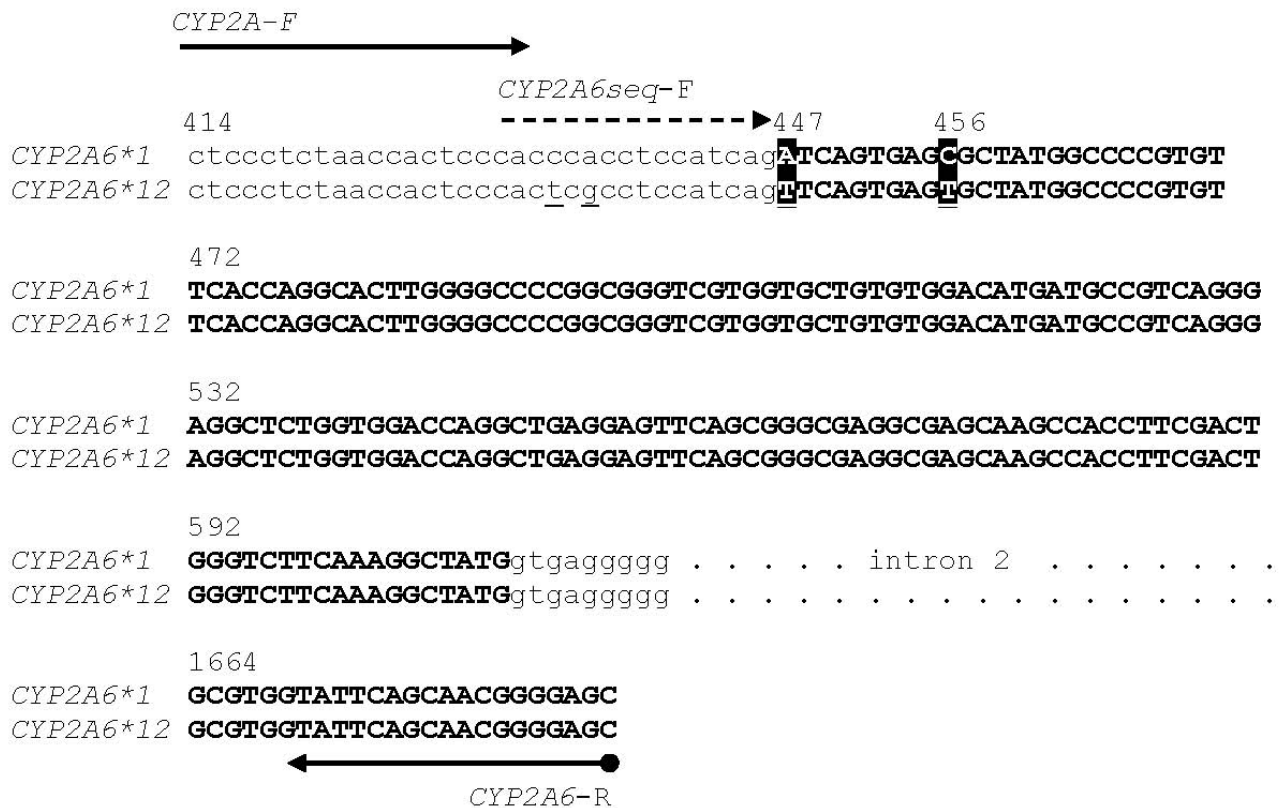


Figure 1
Alignment of DNA sequence for *CYP2A6*1* and *CYP2A6*12*. Arrows indicate forward PCR and sequencing primers (5'-CTCCCTCTAACCCTCCAC-3' and 5'-CACYCRCTCCATCAG-3') and 3'-biotinylated reverse primer (5'-GCTC-CCCGTTGCTGAATA-3'). The resulting 1274 bp PCR product is specific to *CYP2A6*1* and *CYP2A6*12* alleles. Exon 2 and 3 sequences are shown in bold, and the black boxes represent sequence differences between the two alleles that generate allele-specific peaks in the pyrosequencing reaction. Nucleotide numbering positions for both introns and exons are in reference to nucleotide 1, which is the A of the ATG start codon.

A two-phase approach was used to genotype the 90 Coriell PDR DNA samples prior to analysis by the Pyrosequencing assay. First, the conventional two-step PCR method was performed as previously described [6]. The PCR I product from this method was also directly sequenced through use of the ABI PRISM BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). A reverse sequencing primer (5'-AGACTCTGGTCCACACTGGTCAA-3') in intron 1 was designed to capture all sequence differences between *CYP2A6**1 and *CYP2A6**12 in exon 1.

CYP2A6*12 Allele Frequency Determination

The Pyrosequencing assay was used to assess multi-ethnic variation in frequency of the *CYP2A6**12 allele by genotyping individuals from the African American, European

Caucasian, Pacific Rim, and Hispanic population groups. For accurate assessment of allele frequency, it was important to look at *CYP2A6* copy number, since the entire *CYP2A6* gene can be deleted [14]. The gene copy number was determined by use of a TaqMan assay, as previously described [15]. This assay uses a FAM-labeled probe specific to *CYP2A6* and a VIC-labeled probe specific to *CYP2A7* that serves as an internal reference for 2 copies.

Automation of the Pyrosequencing Workflow

The Pyrosequencing assay was used to screen 7,159 individuals from the Third National Health and Nutrition Examination Survey (NHANES III) as part of a larger study of genetic variation in xenobiotic metabolizing enzymes (*manuscript in preparation*). To accommodate the large number of samples, robotics were integrated into the

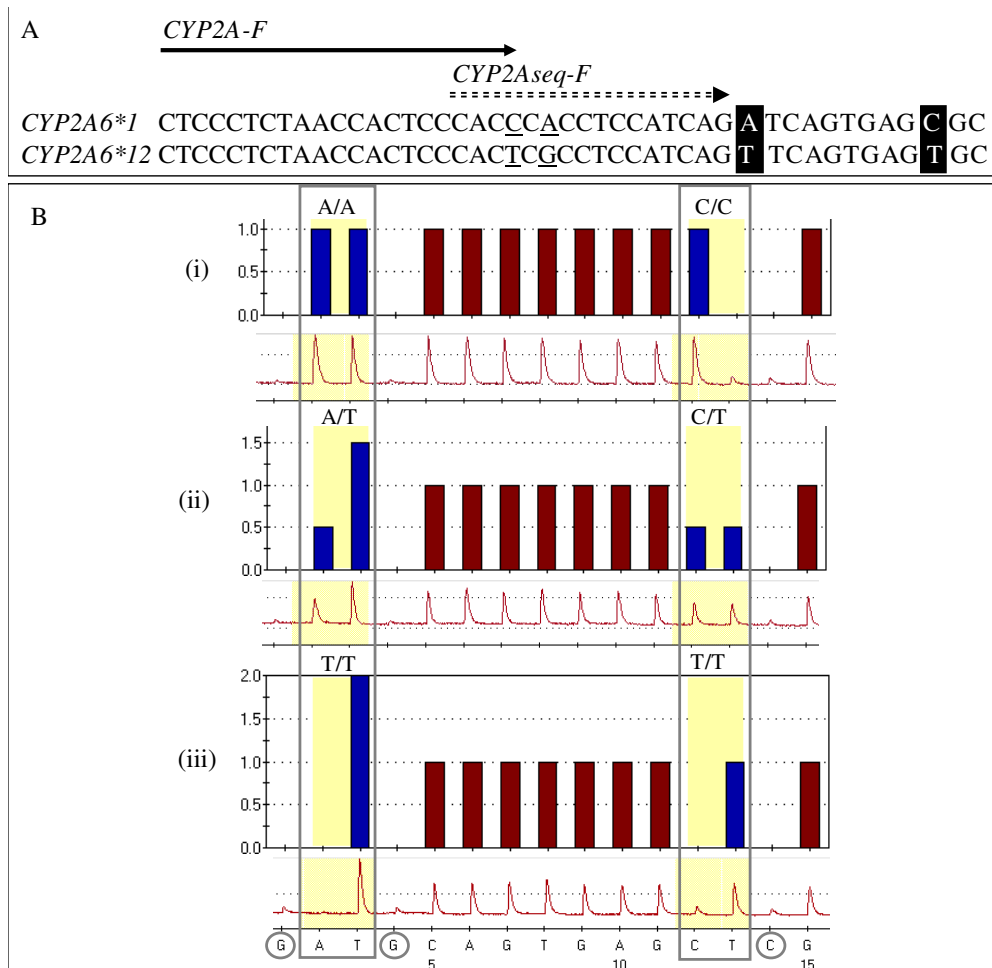


Figure 2

Pyrosequencing strategy and results for *CYP2A6* *1 and *CYP2A612 alleles.** A) Schema of *CYP2A6* *1 and *CYP2A6**12 template sequences; arrows indicate forward PCR and sequencing primers, with degenerate nucleotide positions underlined. The two allele-specific nucleotides are boxed. B) Predicted histogram and pyrograms for *CYP2A6**1/*1 (i), *CYP2A6**1/*12 (ii), and *CYP2A6**12/*12 (iii). Nucleotide dispensation order is shown at the bottom of pyrogram iii. Dispenses 2 and 3 interrogate the first allele-specific nucleotide (A in *CYP2A6* *1 or T in *CYP2A6**12), and dispenses 12 and 13 interrogate the second allele-specific nucleotide (C in *CYP2A6* *1 or T in *CYP2A6**12). Dispenses 1, 4, and 14 (circled) serve as negative background controls.

Table 1: CYP2A6*12 genotype and allele frequencies by ethnicity as determined by Pyrosequencing

Population	Genotypes			Alleles ^b		Published CYP2A6*12 allele frequencies (# alleles tested) ^{ref.}
	CYP2A6 *11/*1a	CYP2A6 *11/*12	CYP2A6 *12/*12	CYP2A6*1	CYP2A6*12	
African American (n = 47)	100% (n = 47)	0% (n = 0)	0% (n = 0)	100% (n = 92)	0% (n = 0)	0.4% (n = 268) ¹⁷
European Caucasian (n = 48)	98% (n = 47)	2% (n = 1)	0% (n = 0)	99% (n = 94)	1% (n = 1)	2% (n = 1440) ¹⁷
Pacific Rim ^c (n = 54)	100% (n = 52)	0% (n = 0)	0% (n = 0)	100% (n = 97)	0% (n = 0)	0% (n = 194) ⁶ 0.3% (n = 354) ¹⁷
Hispanic ^d (n = 32)	90.60% (n = 29)	9.40% (n = 3)	0% (n = 0)	95.30% (n = 61)	4.70% (n = 3)	2.2% (n = 184) ⁶

^a CYP2A6*1 refers to alleles without CYP2A6*12. ^bNumber of alleles based on copy number for CYP2A6. There were individuals with only one copy of CYP2A6 in African Americans (n = 2), European Caucasians (n = 1), and Pacific Rim (n = 7). Two individuals from the Pacific Rim were homozygous for the CYP2A6 gene deletion and were not included in the frequency data. ^cThe Pacific Rim includes Japanese, Chinese, Taiwanese, Melanesian, and Indo Pakistan. ^dThe Hispanic population is representative of the Mexican-American community of Los Angeles for this study and of a region in northern Spain for the referenced published study.

workflow. Increased genotyping throughput and accuracy of sample handling was accomplished by use of the Matrix Hydra DT (Thermo Fisher Scientific, Hudson, NH) to prepare 96-well plates of DNA. The DNA samples were allowed to dry and were stored at room temperature until needed. A Matrix WellMate (Thermo Fisher Scientific, Hudson, NH) was used for multiple dispensing steps throughout the process such as dispensing of PCR mastermix into PCR plates, dispensing of the bead mix for biotinylated PCR capture, and dispensing of the annealing buffer and sequencing primer mix into the PyroMark Q96 plate. In addition, the robotic arm of the PSQ 96HSA system was used for multiple plate testing.

Results and discussion

Pyrosequencing assay genotype results were 100% concordant with those obtained by conventional two-step PCR and cycle sequencing for the 90 PDR and 14 CEPH family members. The PDR is a multi-ethnic collection of unrelated individuals designed to serve as a resource for polymorphism discovery. No phenotypic or ethnic information is linked to the individual samples [16]. It also provides a robust source of DNA material for assay optimization and validation. Genotyping results generated for the CEPH multi-generational family pedigree displayed Mendelian inheritance patterns. This satisfied the validation of our Pyrosequencing assay and provided us with samples to use as positive controls for future CYP2A6*12 genotyping. Pyrosequencing strategy and representative Pyrogram results for each genotype are shown in Figure 2.

Results of the multi-ethnic variation panel are shown in Table 1. There was a higher allele frequency of CYP2A6*12 in Hispanics and Caucasians, with very low occurrence in

African Americans and Asians, as has been reported previously [6,17]. Use of the TaqMan assay allowed us to calculate our CYP2A6*12 allele frequency against a known copy number background. This was of particular importance in the Pacific Rim group where 7 individuals had only one copy of the CYP2A6 gene. Frequencies in all ethnic groups were in Hardy-Weinberg equilibrium. None of these samples were homozygous for CYP2A6*12. However, when the assay was used to screen the 7,159 DNA samples from the NHANES III population, five Mexican-American and non-Hispanic white participants were found to be homozygous for CYP2A6*12. These results were further confirmed by the two-step PCR method.

The addition of robotics into the workflow to facilitate screening of the NHANES III DNA samples greatly streamlined the process. Because of the 96-well format, the Pyrosequencing technology is amenable to a high degree of automation. Use of the robotic arm of the 96HSA system allowed for the transfer of up to ten 96-well plates for unattended operation. Data acquisition and analysis of each batch of ten plates subsequent to placement into the robotic mechanism was obtained in less than 2 hours.

In summary, we developed a single PCR-based Pyrosequencing assay for rapid detection of the CYP2A6*12 hybrid allele. Similar to conventional DNA sequencing, Pyrosequencing assays generate sequence data but without the need of a cycle sequencing step, expensive fluorescent dyes, or gel matrix; moreover, data are accessible in real-time. The assay, which is dependent upon amplification of highly specific PCR templates for wild-type and hybrid alleles, generates reliable substrates by which one can unequivocally identify CYP2A6*12 alleles.

Conclusion

A two-step PCR is the predominant method of detection of the *CYP2A6*12* hybrid allele. The Pyrosequencing assay presented here offers a rapid alternative to the significantly longer conventional method. The integration of liquid handling robotics into the workflow helps reduce errors and further streamlines the process. This approach provides a substantial decrease in laboratory effort and can help facilitate analysis of this variant in large populations used in epidemiological and pharmacogenomic studies.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DK conceived of the assay design and implementation. JH and AS performed the experimental analyses. DK and MG interpreted the results and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Stanimila Nikolova for her technical assistance in confirming the *CYP2A6*12* homozygote samples. Jacqueline Huckins and Antonina Spencer were funded by the Research Participation Program at CDC, an inter-agency agreement with the U.S. Department of Energy administered by Oak Ridge Institute for Science and Education. This work was funded through allocations for the Centers for Disease Control and Prevention, Division of Laboratory Sciences.

References

- Nelson DR, Zeldin DC, Hoffman SM, Maltais LJ, Wain HM, Nebert DW: **Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants.** *Pharmacogenetics* 2004, **14**:1-18.
- Miles JS, Bickmore W, Brook JD, McLaren AW, Meehan R, Wolf CR: **Close linkage of the human cytochrome P450IIA and P450IIB gene subfamilies: implications for the assignment of substrate specificity.** *Nucleic Acids Res* 1989, **17**:2907-2917.
- Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C: **Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoeconomic and clinical aspects.** *Pharmacol Ther* 2007, **116**:496-526.
- Pianezza ML, Sellers EM, Tyndale RF: **Nicotine metabolism defect reduces smoking.** *Nature* 1998, **393**:750.
- Human Cytochrome P450 (CYP) Allele Nomenclature Committee: allele nomenclature for cytochrome P450 enzymes** [<http://www.cypalleles.ki.se>]
- Oscarson M, McLellan RA, Asp V, Ledesma M, Ruiz ML, Sinus B, Rautio A, Ingelman-Sundberg M: **Characterization of a novel hybrid allele (CYP2A6*12) that causes reduced CYP2A6 activity.** *Hum Mutat* 2002, **20**:275-283.
- Alderborn A, Kristofferson A, Hammerling U: **Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing.** *Genome Res* 2000, **10**:1249-1258.
- Yang AS, Marcos RH, Doshi K, Kondo Y, Tajara EH, Issa JJ: **A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements.** *Nucleic Acids Res* 2004, **32**:e38.
- Jansson M, Rada A, Tomic L, Larsson L, Wadelius C: **Analysis of the Glutathione S-transferase M1 gene using pyrosequencing and multiplex PCR—no evidence of association to glaucoma.** *Exp Eye Res* 2003, **77**:239-243.
- Soderback E, Zackrisson A, Lindblom B, Alderborn A: **Determination of CYP2D6 gene copy number by pyrosequencing.** *Clin Chem* 2005, **51**:522-531.
- Plan and operation of the Third National Health and Nutrition Examination Survey, 1988–94.** National Center for Health Statistics. *Vital Health Stat I* 1994(32) [http://www.cdc.gov/nchs/data/series/sr_01/sr01_032.pdf]
- National Center for Health Statistics, Centers for Disease Control and Prevention. NHANES III genetic data.** Hyattsville, MD: National Center for Health Statistics [http://www.cdc.gov/nchs/nhanes/nh3data_genetic.htm]
- Ensembl Genome Browser** [<http://www.ensembl.org>]
- Ariyoshi N, Sekine H, Saito K, Kamataki T: **Characterization of a genotype previously designated as CYP2A6 D-type: CYP2A6*4B, another entire gene deletion allele of the CYP2A6 gene in Japanese.** *Pharmacogenetics* 2002, **12**:501-504.
- Haberl M, Anwald B, Klein K, Weil R, Fuß C, Gepdiremen A, Zanger UM, Meyer UA, Wojnowski L: **Three haplotypes associated with CYP2A6 phenotypes in Caucasians.** *Pharmacogenet Genomics* 2005, **15**:609-624.
- Collins FS, Brooks LD, Chakravarti A: **A DNA polymorphism discovery resource for research on human genetic variation.** *Genome Res* 1998, **8**:1229-1231.
- Schoedel KA, Hoffman EB, Rao Y, Sellers EM, Tyndale RF: **Ethnic variation in CYP2A6 and association of genetically slow nicotine metabolism and smoking in adult Caucasians.** *Pharmacogenetics* 2004, **14**:615-626.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2350/10/80/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

