Genetics of CYP1A1: Coamplification of Specific Alleles by Polymerase Chain Reaction and Association with Breast Cancer

Timothy R. Rebbeck, Elizabeth A. Rosvold, David J. Duggan, John Zhang, and Kenneth H. Buetow
Division of Population Science, The Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Abstract
CYP1A1 is a gene of the cytochrome P-450 family that has been proposed to be a biomarker of cancer risk. We introduce a polymerase chain reaction-based assay to measure allelic variability in exon 7 of the CYP1A1 gene. This genetic variant is associated with an amino acid change at residue 462 in the aryl hydrocarbon hydroxylase protein product. Previously, measurement of CYP1A1 genotypes at this variant site required two assays, one to detect each allele. By using three primers in a single polymerase chain reaction rather than two primers in each of two polymerase chain reactions, the proposed assay may facilitate population-based study protocols. We estimate the frequency of this polymorphism in a Caucasian population to be 0.03, with an observed heterozygosity of 0.06. We have also confirmed the Mendelian segregation of this polymorphism in four multigeneration Centre d’Étude du Polymorphisme Humain families and have placed this locus in a multilocus linkage map on chromosome 15q. The distribution of this polymorphism was the same in breast cancer cases as in two sets of healthy controls.

Introduction
The CYP1A1 locus encodes AHH. AHH is involved in the metabolism of environmental carcinogens and estrogens (1–3). Therefore, CYP1A1 genotypes may be biomarkers of cancer risk. The CYP1A1 gene has been cloned and sequenced (3, 4). Several DNA polymorphisms at this locus have been reported (5–8). Of these, an Mspl polymorphism in the 3’-flanking region of the CYP1A1 gene (7, 9) is correlated with AHH enzyme inducibility (10). This Mspl polymorphism is in linkage disequilibrium with an allelic variant in exon 7 at the site of a single A to G base pair substitution. This substitution results in an isoleucine to valine amino acid sequence change at residue 462 of the heme-binding region in the cytochrome P4501A1 protein (11).

The objective of this paper is to present a new PCR-based CYP1A1 assay to measure allelic variability in exon 7 of the CYP1A1 gene, to describe the genetic epidemiology of this polymorphism in Caucasians, and to examine the association of this polymorphism with breast cancer. A single PCR reaction is described here that coamplifies two allele-specific DNA sequences in the CYP1A1 gene using three synthetic primer strands. We have measured this polymorphism in a sample of CEPH families, unrelated CEPH adults, healthy controls, and FCCC breast cancer cases.

Materials and Methods
Study Subjects. Two samples were considered in the present analysis. (a) Members of CEPH reference families were used to study the pattern of segregation of the CYP1A1 polymorphism and to place this locus on a multipoint linkage map. In addition to these families, a sample of 90 unrelated CEPH adults were studied to estimate the frequencies of the CYP1A1 alleles; and (b) breast cancer cases and healthy controls ascertained through the FCCC and its affiliated hospitals were studied. A consecutive set of 96 incident breast cancer cases diagnosed at the FCCC between 1991 and 1993 were considered. A sample of 146 healthy controls were selected for comparison to the breast cancer cases. These controls were individuals undergoing routine cancer screening at the FCCC, FCCC employees, or FCCC volunteers. Employees and volunteers were drawn from the same socioeconomic and geographical area as the breast cancer patients (Northeast Philadelphia and surrounding Montgomery County).

Biosamples and PCR Assay. DNA was extracted from peripheral blood samples obtained from all study subjects. Nucleic acid extraction was performed by using the method of Miller et al. (12) or by an ABI 340 Nucleic Acid Extraction Machine using the protocols and reagents suggested by the manufacturer.

Using published sequence data, three primer sequences were designed using the computer software PRIMER (Ver. 0.5; Whitehead Institute for Biomedical Research). The primers synthesized were:

A allele-specific sequence:

5′GAA GTG TAT CCG TGA GAC CA-3′

G allele-specific sequence:

5′-GTG TAT CCG TGA GAC CG-3′

Received 3/9/94; revised 5/26/94; accepted 5/26/94.
1 This work was supported by grants from the Lucille P. Markey Charitable Trust, the Margaret Dyson Foundation, USPSH Grants HG00355 and HG00206, NIH Grants CA60798, CA57708, CA06927, and RR05895, and an appropriation from the Commonwealth of Pennsylvania.
2 To whom requests for reprints should be addressed, at University of Pennsylvania, School of Medicine, NEB 229L-6095, 420 Service Drive, Philadelphia, PA 19104.
3 The abbreviations used are: AHH, aryl hydrocarbon hydroxylase; PCR, polymerase chain reaction; CEPH, Centre d’Étude du Polymorphisme Humain; FCCC, Fox Chase Cancer Center; CE, catecholestrogen.
Opposing strand sequence:

5′-GTA GAC AGA GTC TAG GCC TCA-3′

The single 5-μl PCR amplification reaction mixture consisted of 0.51 μl of double distilled water, 0.5 μl of 10× reaction buffer [67 mM counter ion buffer (pH 8.8) containing 6.7 mM MgCl2, 16 mM ammonium sulfate, 10 mM 2-mercaptoethanol, and 10% dimethyl sulfoxide], 0.1 μl of 50 mM Mg2+, 0.8 μl of 0.25 mM deoxynucleotide triphosphate, 0.5 μl each of 5-μm primers, 0.05 μl [35S]dATP, 1.5 μl of 10 ng/μl template DNA, and 0.04 μl of Taq polymerase (Amplicon). After an initial denaturation step of 95°C for 5 min, the amplification profile consisted of 35 cycles of 95°C for 1 min, 58°C for 1 min, and 72°C for 1 min. A final extension period of 10 min at 73°C was carried out at the end of the PCR cycling.

Electrophoresis was carried out by using a 6% denaturing polyacrylamide gel (20 ml of 30% acrylamide, 26.6 ml distilled H2O, 20 ml of 5× Tris-borate EDTA buffer, and 42 g of urea). The gel mixture was vacuum degassed for 15 min with 340 μl of 10% ammonium persulfate. Polymerization was aided by the addition of 34 μl of N,N,N′,N′-tetramethylethlenediamine. Five μl of denaturing loading buffer [95 ml filtered formamide-4 ml 0.5 M phosphate, 0.5 μl each of 5× counter ion buffer (pH 8.8)-0.1 g bromophenol blue-0.1 g xylene cyanol] was added to the 5 μl DNA amplification product samples. These samples were denatured at 95°C for 5 min prior to electrophoresis for 3 h at room temperature. The gel was then transferred to Whatman filter paper and exposed on Kodak X-OMAT AR film for 2 days.

**Statistical Analysis.** Estimates of CYP1A1 allele frequencies were made using the samples of unrelated CEPH adults and the FCCC control population. A χ2 test was used to identify deviations from Hardy-Weinberg proportions (13). The chromosomal location of the CYP1A1 gene was determined by computing multipoint linkage statistics using maximum likelihood methods as implemented in the computer software CRIMAP (Ver. 2.4). The reference genetic map was generated by the Cooperative Human Linkage Center using Ver. 2.0 of combined CEPH data. These data are available by request from CEPH, electronically via anonymous file transfer program (FTP) at ftp.chlc.org, or via Gopher server at gopher.chlc.org. Two-tailed exact tests for 2 x 2 tables (14) were used to compare the distribution of CYP1A1 alleles among breast cancer cases and controls. Gene frequency and disease association computations were undertaken using SAS Release 6.03 (SAS Institute, Carey NC).

**Results**

**Polymorphism Frequencies.** The single PCR reaction produced two amplified sequences; the A allele-specific PCR product was 208 base pairs and the G allele-specific PCR product was 211 base pairs. The migration patterns of these alleles on denaturing gel electrophoresis are presented in Fig. 1. Using 90 unrelated CEPH parents, 85 AA, 5 AG, and 0 GG genotypes were observed. This corresponded to frequencies of 0.97 for the A allele and 0.03 for the G allele. Observed heterozygosity was computed to be 0.06. These genotypes were confirmed using the assay of Hayashi et al. (Ref. 11; results not shown). There was no deviation from expected Hardy-Weinberg frequencies (χ2 = 0.002; df = 2; P > 0.9).

**Mendelian Inheritance and Linkage.** Codominant segregation was confirmed using multigeneration CEPH families.

![AA AG GG](image)

Fig. 1. Alleles of the CYP1A1 polymorphism.

The results of these analyses are presented in Table 1. A set of four three-generation CEPH families was typed at the CYP1A1 locus. These families were genotyped because the parents represented different mating types that would be expected to demonstrate Mendelian segregation. The observed genotypes were consistent with Mendelian segregation in the four families (Table 1). A fifth (control) family was genotyped in which parents were monomorphic at the CYP1A1 locus. As expected, no Mendelian segregation was observed in this family (1334, Table 1).

The results of multipoint linkage analyses using CEPH data confirmed the location of CYP1A1 to chromosome 15 (compare with Ref. 15). The best location for the CYP1A1 locus was in the interval between D15S117 and D15S108. No recombinants were observed between the CYP1A1 polymorphism and these markers. Two-point linkage analysis indicated that three chromosome 15 Cooperative Human Linkage Center markers showed the most significant linkage with the CYP1A1 polymorphism: D15S108 (lod score = 3.31); D15S33 (lod score = 3.20); and D15S26 (lod score = 3.20). Other regions of chromosome 15q bounded by D15S97 and D15S175 could not be excluded from linkage with CYP1A1.

**Association with Breast Cancer.** A comparison of the distribution of CYP1A1 alleles in breast cancer cases and controls is presented in Table 2. Two G alleles of 192 total alleles (0.5%) were observed in breast cancer cases, and no G alleles of 252 total alleles were observed in FCCC controls. This relationship was not statistically significant (exact test significance = 0.186). In CEPH controls, a higher proportion of G alleles (5 of 180; 3%) was observed than that in breast cases (0.5%). This relationship was not statistically significant (exact test significance = 0.271). When FCCC and CEPH controls were pooled, the proportion of G alleles (5 of 432; 0.2%) was lower than that in breast cases (0.5%). Again, this difference was not statistically significant (exact

---

**Table 1**

<table>
<thead>
<tr>
<th>CEPH family</th>
<th>Parental mating type (father x mother)</th>
<th>Number of offspring with genotypes</th>
<th>Paternal, maternal grandparents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1333</td>
<td>AG x AG</td>
<td>1 6 2</td>
<td>AA x AA, AA x AG</td>
</tr>
<tr>
<td>1334</td>
<td>AA x AA</td>
<td>7 0 0</td>
<td>AA x AA, AA x AA</td>
</tr>
<tr>
<td>1354</td>
<td>AG x AA</td>
<td>4 7 0</td>
<td>AG x AA, AA x AA</td>
</tr>
<tr>
<td>1454</td>
<td>AA x AG</td>
<td>2 8 0</td>
<td>AA x AG, AG x AA</td>
</tr>
<tr>
<td>1582</td>
<td>AA x AG</td>
<td>5 2 0</td>
<td>AA x AA, AG x AA</td>
</tr>
</tbody>
</table>

*genotype unavailable.

---

**Table 2**

<table>
<thead>
<tr>
<th>Alleles in breast cancer cases</th>
<th>Alleles in FCCC controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA x AG</td>
<td>AA x AG</td>
</tr>
<tr>
<td>AG x AA</td>
<td>AA x AA</td>
</tr>
<tr>
<td>AA x AG</td>
<td>AG x AA</td>
</tr>
<tr>
<td>AG x AA</td>
<td>AA x AA</td>
</tr>
<tr>
<td>AA x AG</td>
<td>AA x AG</td>
</tr>
</tbody>
</table>

*genotype unavailable.

---

**Fig. 1.** Alleles of the CYP1A1 polymorphism.
test significance = 0.688). Finally, G alleles were observed in CEPH controls (3%), while no G alleles were observed in FCCC controls. This relationship was statistically significant (exact test significance = 0.012).

The lack of statistical significance reported here may in part be explained by a lack of statistical power resulting from the low G allele frequency. We computed the approximate statistical power of the present sample to detect a difference in allele frequency in breast cancer cases compared with controls using the SOLO Power Analysis software (BMDP Statistical Software, Inc., Los Angeles, CA). We assumed a G allele frequency estimate of 0.005 in breast cancer cases (as observed in Table 2) and 0.03 in controls (as observed for CEPH controls in Table 2). At an $\alpha$ level of 0.05, a sample of 175 case alleles and 175 control alleles would have 78% power to detect an allelic association of CYP1A1 with breast cancer. This result suggests that the sample size used in the present analyses might not be sufficient to have detected an association of the CYP1A1 genotype with breast cancer.

**Discussion**

We have introduced a PCR-based assay to measure allelic variability associated with an amino acid change in the protein product of the CYP1A1 gene. Previously, measurement of CYP1A1 genotypes at the exon 7 variant site required two assays, one to detect the A allele, and one to detect the G allele. By using three primers in a single PCR rather than two primers in each of two PCRs, we have halved the number of PCRs required to generate genotypes at this locus. This approach eliminates the need for allele-specific PCRs. Genotype assays of the type proposed here may be preferable for population-based protocols when DNA availability is a limiting factor or when efficient, high throughput genotyping is desirable.

Allelic variation at the exon 7 variant site studied here is in linkage disequilibrium with a MspI RFLP polymorphism. This polymorphism occurs with allele frequencies of 0.69 and 0.31 in a Japanese population (9) and with frequencies of 0.86 and 0.12 in a racially mixed American population (10). Estimates of allele frequencies at the variant site in exon 7 of the CYP1A1 gene have been estimated by Hayashi et al. (11) to be 0.80 and 0.20. In contrast, the estimates of the gene frequency in the present samples of Caucasians is substantially lower than either of these (0.97 and 0.03). The allele frequency estimate in the present sample of Caucasians is closer to that of Caucasians in the sample of Petersen et al. (10) than of Japanese in the samples of Kawajiri et al. (9) or Hayashi et al. (11). We conclude that variant alleles in the CYP1A1 gene may be less common in Caucasians than in Japanese.

Previous studies have examined the relationship between cancer and either phenotypic variability in AHH or genotypic variability in CYP1A1. Kellerman et al. (16) compared AHH levels in lung cancer cases and healthy controls. The low inducibility AHH phenotype occurred in 4% of lung cancer cases and 44% of controls. The high inducibility AHH phenotype occurred in 9% of lung cancer cases and 30% of controls. Similar results have confirmed this relationship (17–20). Kawajiri et al. (9) studied the relationship of CYP1A1 alleles of the MspI polymorphism between lung cancer cases and controls. Homozygotes for a rare CYP1A1 allele identified by the MspI restriction site had a 7-fold increase in lung cancer risk compared with controls. However, this relationship was not observed in other case-control studies of lung cancer to date (21, 22). No studies to date have reported the relationship of CYP1A1 allelic variants with breast cancer.

Pending the results of more comprehensive epidemiologic analyses, there is reason to consider CYP1A1 and its protein product as candidate biomarkers of breast cancer risk. AHH activity has been found in breast tumor tissue but not in normal breast epithelium (23). AHH catalyzes the monooxygenation of polycyclic aromatic hydrocarbons to phenolic products and epoxides that may be mutagenic and carcinogenic (1). For example, the products of AHH activation of benzo[a]pyrene form DNA adducts in breast tissue that are associated with DNA mutagenesis (2). AHH is also involved in the conversion of estrogen to hydroxylated CEs, particularly 2-hydroxyestradiol (24). CEs may initiate carcinogenesis through oxidation to estrogen quinones, promote DNA adduct formation, and result in DNA mutation (25). It has been suggested that CEs make a significant postmenarchal contribution to steroid-related breast carcinogenesis (26) and that metabolic competition of 2-, 4-, and 16-hydroxylated estrogens may be associated with breast carcinogenesis (27). Hydroxylation of estrogen by AHH may also act as an antiestrogen to protect against certain types of estrogen-dependent cancers (28). Studies have suggested that induction of AHH by 2,3,7,8-tetrachlorodibenzo-p-dioxin explains the antiestrogenic effect of 2,3,7,8-tetrachlorodibenzo-p-dioxin exposure (29, 30). Reports of an inverse relationship of smoking status with breast cancer risk (31) may be explained in part by the fact that AHH is activated by compounds found in cigarette smoke (32) and that increased 2-hydroxylation of estrogen has also been observed in female smokers (33). The role of AHH in both carcinogen activation and estrogen metabolism suggests that the CYP1A1 gene is a candidate genetic marker of breast cancer risk.

Biomarkers of the type reported here may ultimately be important in understanding cancer etiology and assessing the risk of an individual to develop cancer. However, it is unlikely that a single marker of this type will completely explain the cancer risk of an individual. Instead, an understanding of cancer etiology and risk will require information about multiple inherited genetic variants as well as endogenous and exogenous environmental exposures. This report represents one step in the synthesis of that information.

**Acknowledgments**

The authors wish to acknowledge Edward D. Lustbader and Katherine A. McGlynn for their contributions to the design and execution of this and related studies.
References


Genetics of CYP1A1: coamplification of specific alleles by polymerase chain reaction and association with breast cancer.

T R Rebbeck, E A Rosvold, D J Duggan, et al.


Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/3/6/511

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.