Polymorphisms in the human aromatase cytochrome P450 gene (CYP19) and breast cancer risk

Catherine S. Healey, Alison M. Dunning, Francine Durocher, Dawn Teare, Paul D. P. Pharaoah, Robert N. Luben, Douglas F. Easton and Bruce A. J. Ponder

Carcinogenesis vol.21 no.2 pp.189–193, 2000

The aromatase enzyme catalyses the conversion of androgens to oestrogens in the oestrogen biosynthesis pathway. Because increased exposure to oestrogens is considered to be a risk factor for breast cancer, the human aromatase gene (CYP19) is a plausible candidate for low penetrance breast cancer susceptibility. Preliminary reports have suggested that specific alleles of a TTTA repeat may be associated with differences in breast cancer risk. We have identified two new polymorphisms in the CYP19 gene: a TCT insertion/deletion in intron 4 and a G→T substitution in intron 6, which have rare allele frequencies of 0.35 and 0.45, respectively, in the British population. Comparison was made between the frequencies of these alleles and those of the TTTA repeat in up to 599 breast cancer cases and 433 normal controls from the East Anglian British population. We found strong linkage disequilibrium between the alleles of these three loci, but no significant association of any alleles with breast cancer risk. The maximum odds ratios observed were: 1.03 (95% CI 0.68–1.55) for the intron 4 TCT insertion/deletion polymorphism [del/del versus ins/ins]; 1.56 (95% CI 0.68–1.55) for the intron 4 [TTTA]10 allele; 1.29 (95% CI 0.99–2.21) for the intron 6 G→T polymorphism [TT versus GG]. We conclude that the CYP19 gene has no major role in common breast cancer incidence in the British population.

Introduction

Breast cancer is a common disease that will affect 8% of British women. However, less than 5% of breast cancer incidence can be explained by rare, highly penetrant genes such as BRCA2 and TP53 (1,2). In principle, common, low penetrance genes could explain the majority of breast cancer cases.

It is widely accepted that oestrogens are involved in the development of breast tumours and increased lifetime exposure to endogenous oestrogens is known to increase risk of breast cancer (3,4). Epidemiological studies have identified several factors affecting lifetime exposure that increase the risk of breast cancer. These include: early age at menarche; late age at first pregnancy; late age at menopause; nulliparity; and post-menopausal obesity (oestrogen biosynthesis in post-menopausal women occurs primarily in adipose tissue). In addition to length of exposure to oestrogens, exposure to high levels of oestrogens may also increase breast cancer risk. Toniolo et al. (5) demonstrated that women with high levels of serum oestrogens, particularly oestradiol, the most biologically active oestrogen, have an increased risk of breast cancer. Furthermore, oestrogens in breast adipose tissue may act locally to promote the growth of breast carcinomas (6–8).

The CYP19 gene, located on chromosome 15, encodes the enzyme P450 aromatase, which catalyses the biosynthesis of oestrogens from androgens in three sequential steps (described in ref. 9). Aromatase P450 is present in the endoplasmic reticulum of cells in which it is expressed, including the granulosa cells and corpus luteum of the ovary, the Leydig cells of the testis, the placenta, various sites in the brain and in adipose tissue. Different forms of oestrogens are synthesized from different androgen substrates in the various tissues. For instance, oestrone is synthesized from androstenedione in adipose tissue, oestradiol from testosterone in granulosa cells and oestriol from 16α-hydroxylated androgens in the placenta.

The CYP19 gene has 10 exons and transcription begins in exon 2. Tissue specificity is achieved using alternative splicing of exon 1 combined with different tissue-specific promoters (reviewed in ref. 10). Several mutations in the CYP19 gene have been identified which result in an autosomal recessive form of female pseudohermaphroditism and virilization of the mother during pregnancy due to impaired or absent aromatase activity. In addition, at puberty the affected female may show signs of virilization and have pubertal failure, hypergonadotrophic hypogonadism, polycystic ovaries and tall stature (11–14).

Two polymorphic sites have been used in breast cancer studies: (i) a tetranucleotide repeat polymorphism, [TTTA]n, located in intron 4 of the CYP19 gene (15); and (ii) a C826T variation in exon 7 which gives rise to the amino acid substitution Arg264Cys and is observed by sequencing and SSCP analysis (9,16–18). There have been numerous reports of different TTTA repeat alleles being associated with variation in breast cancer risk: Kristensen et al. (19) found that carrying the [TTTA]12 allele was associated with an increased risk of breast cancer in their Scandinavian population [odds ratio (OR) 2.42, 95% confidence interval (95% CI) 1.03–5.80]. This was confirmed in a study by Haiman and co-workers (20) where the [TTTA]12 allele was over-represented in breast cancer cases (OR 1.84, 95% CI 1.02–3.32), but refuted by Siegelmann-Danieli and Buetow (21), who found that [TTTA]12 occurred at a greater frequency in their control population. Haiman et al. (20) also found an increase in breast cancer risk associated with the [TTTA]10 allele (OR 4.03, 95% CI 1.52–10.67) and Siegelmann-Danieli and Buetow (21) reported a greater frequency of the [TTTA]n allele measuring 171 bp (allele 2) in cases compared with controls (18.5 versus 13.38%, OR 1.47, 95% CI 0.99–2.17).
Sourdaine et al. (17) found one Arg264Cys heterozygote in five breast tumours, but no polymorphism was seen in an equal number of controls, suggesting a low Cys264 allele frequency in the British population. In contrast, Watanabe et al. (18) found the Cys264 allele frequency to be 30% in a Japanese population sample. Seigelmann-Danieli and Buetow (21) also reported the variant in US breast cancer cases. Although creating a non-conservative amino acid substitution, this polymorphism has no apparent effect on aromatase activity or response to aromatase inhibitors (17,18), nor is it associated with breast cancer risk in Japanese women (18). In this present study we set out to investigate possible associations between common CYP19 polymorphisms and breast cancer risk in the British population.

Materials and methods

Subjects

All patients and controls in this study are Caucasian females from the East Anglian region of the UK. Two separate series (strata) were used to overcome the need to increase thresholds of statistical significance when carrying out multiple tests. The selection criteria for the two series differed slightly and the sets of results were initially examined individually. They were found to be similar and so were combined.

The first series is a prospectively ascertained group of 288 incident patients, attending the Addenbrooke's Hospital (East Anglia) for treatment between 1992 and 1995 and diagnosed below age 71 years (mean age 52.5, standard deviation 12.8, range 28.6–70.8 years). These are compared with a group of 288 randomly selected, anonymous controls taken from the EPIC study (22), a population-based cohort study of diet and health (mean age 58.9, standard deviation 9.2, range 44.7–75.6 years). This cohort contains ~25,000 individuals resident in Norfolk (East Anglia) recruited from 1992 to 1996.

The second series is a group of 384 patients retrospectively ascertained through the East Anglian Breast Cancer Registry, as part of the Anglian Breast Cancer Study, comprising all patients diagnosed below age 55 years since 1991 and still alive in 1996 (mean age 46.6, standard deviation 5.7, range 25.0–54.9 years). These are compared with a second group of 384 random controls also from the EPIC cohort (mean age 55.6, standard deviation 8.1, range 39.9–69.9 years).

Genotyping

Microsatellite analysis of the intron 4 TTTA repeat was performed on 20–100 ng of DNA as follows: 200 mM PolyF (CCAGTTAATGTG-GAGC) and PolyR (TTACGGTACCAAGGATC) (15), 0.4 U red hot polymerase and 10× buffer (Advanced Biotechnologies), 1.5 mM MgCl₂, dNTPs (200 µM A, T and G and 10 µM dCTP) and 0.3 µCi [α-³²P]dCTP in a 30 µl volume. Amplification conditions were: 1 cycle of 95°C for 5 min; 40 cycles of 95°C for 1 min, 55°C for 1 min, 72°C for 1 min; 1 cycle of 72°C for 10 min in an MJ Tetrad PCR machine (GRI, UK). The PCR was then electrophoresed through a sequagel-6 matrix (National Diagnostics) at 120 W until the xylene cyanol dye reached the bottom. The gel was exposed to X-ray film at ~70°C overnight. [TTTA]₈ gives a fragment of 175 bp.

The same fragment was amplified for sequence analysis using the same PCR conditions as above, but excluding [α-³²P]dCTP and using 200 µM dNTPs. The Sequenase PCR product sequencing kit (US Biochemical) was used according to the manufacturer’s instructions.

A second set of primers for [TTTA]₈ microsatellite analysis in intron 4 were designed to exclude the TCT insertion/deletion (Ins/Del site): I₄F, and the [TTTA]ₙ tract (I₄R, CTGGGTGATAGAGTCAGAGC. These were used in microsatellite analysis as described above. [TTTA]₈ in this instance gives a fragment of 105 bp.

Single-strand conformation polymorphism (SSCP) and sequence analysis of intron 6/ exon 7 were performed as for intron 4 but using the primers I₄F (TCTTTAAGTACTGCTGGC) and I₄R (TTACGGTATAGGATCAGAC). These were used in microsatellite analysis as described above. [TTTA]₈ in this instance gives a fragment of 105 bp.

Statistics

Associations between polymorphisms and breast cancer were analysed by logistic regression using the program S-plus. Cases and controls were genotyped in two groups where possible, stratum 1 (up to 288 cases and 288 controls) and stratum 2 (up to 384 cases and 384 controls), and all analyses allowed for strata as a covariate. The effects of the TTTA repeats were assessed by testing for a trend in breast cancer risk with repeat length by fitting a parameter for repeat length (averaged over the two chromosomes) in logistic regression. Linkage disequilibrium between pairs of allelic sites was estimated using the correlation coefficient D (23). Meta-analysis was performed by log linear regression. Odds ratios were estimated for specific alleles versus all others and each study was treated as a separate stratum.

Results

Polymorphism discovery

Initially all samples were genotyped for the [TTTA]ₙ tract in intron 4 with the PolyF and PolyR primers. TTTA repeats ranging from 7–13 were observed and the frequencies are given in Table I. In order to correlate the microsatellite band sizes with TTTA repeat length, homozygotes for the two most common alleles (Figure 1a), allele 1 (the smallest) and allele 6, were sequenced. Allele 6 had 11 TTTA repeats, but allele 1 had seven repeats, not six as predicted from its size on the microsatellite gel. However, when the sequence was examined further, a 3 bp TCT deletion was found 50 bp upstream of the [TTTA]ₙ tract in allele 1 but not in allele 6. This deletion explained the apparent discrepancy between band size and repeat number and represents an insertion/deletion polymorphism. The rarer TCT deletion allele has a frequency of 0.35 in the control population. We have subsequently found that this polymorphism has been independently discovered in the Japanese population (24).

We searched for the reported Arg264Cys (9,16–18) polymorphism in exon 7 by sequencing in 17 breast cancer samples (34 alleles), to look for the C→T change. No such polymorphism was found in these samples (i.e. we observed a Cys allele frequency of 0; exact 95% CI 0–8%). However, in the same PCR fragment, a novel G→T change in intron 6 (Figure 2), 105 bp upstream from exon 7, was observed with a T allele frequency of 0.45. This change creates a SSCP. From sequence analysis of intron 6, an extra adenine base (cagAtca, Figure 2) was also found 89 bp upstream of exon 7 that was missing from the published sequence (GenBank/EMBL accession no. J05105). This was universally present (i.e. not polymorphic) in our population sample.

Linkage disequilibrium

In order to establish linkage disequilibrium between the [TTTA]₈ tract and the TCT Ins/Del polymorphism, the relative sizes of the bands from the two PCR fragments including and excluding the TCT (PolyF and PolyR versus I₄F and I₄R, respectively) were compared. Any haplotype with the TCT deletion stayed in the same relative position (Figure 1a and b). The second series is a group of 384 patients retrospectively ascertained through the East Anglian Breast Cancer Registry, comprising all patients diagnosed below age 55 years since 1991 and still alive in 1996 (mean age 46.6, standard deviation 5.7, range 25.0–54.9 years). These analyses are compared with the second group of 384 random controls also from the EPIC cohort (mean age 55.6, standard deviation 8.1, range 39.9–69.9 years).

Association with breast cancer

Each polymorphism has been examined for any association with an increased risk of breast cancer. For the [TTTA]₈ polymorphism we compared allele frequencies between cases and controls to look for any association with breast cancer risk and no significant differences were observed (Table I).
CYP19 polymorphisms and breast cancer risk

Table I. CYP19 intron 4 TTTA repeat length alleles in cases and controls

<table>
<thead>
<tr>
<th>Allele</th>
<th>Total alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>195</td>
</tr>
<tr>
<td>7</td>
<td>191</td>
</tr>
<tr>
<td>6</td>
<td>187</td>
</tr>
<tr>
<td>5</td>
<td>183</td>
</tr>
<tr>
<td>4</td>
<td>179</td>
</tr>
<tr>
<td>3</td>
<td>175</td>
</tr>
<tr>
<td>2</td>
<td>171</td>
</tr>
<tr>
<td>1</td>
<td>168</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fragment size (bp)</th>
<th>195</th>
<th>191</th>
<th>187</th>
<th>183</th>
<th>179</th>
<th>175</th>
<th>171</th>
<th>168</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases (%)</td>
<td>7</td>
<td>31</td>
<td>458</td>
<td>15</td>
<td>6</td>
<td>107</td>
<td>169</td>
<td>405</td>
</tr>
<tr>
<td>Controls (%)</td>
<td>6</td>
<td>18</td>
<td>328</td>
<td>7</td>
<td>5</td>
<td>88</td>
<td>108</td>
<td>306</td>
</tr>
</tbody>
</table>

Alleles observed using primers Poly F and R (16).

Table II. Common haplotypes, in normal controls, of the three polymorphic sites: the TCT Ins/Del and [TTTA]n repeat in intron 4 and the intron 6 G→T

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>TCT Ins/Del, intron 4</th>
<th>Del</th>
<th>Ins</th>
<th>Ins</th>
<th>Other</th>
<th>G→T, intron 6</th>
<th>T</th>
<th>T</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>[TTTA]n, repeat, intron 4</td>
<td>7</td>
<td>7</td>
<td>8+</td>
<td>Other</td>
<td>7</td>
<td>7</td>
<td>T</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>G→T, intron 6</td>
<td>9</td>
<td>99 (35.4)</td>
<td>30 (10.7)</td>
<td>146 (52.1)</td>
<td>5 (1.8)</td>
<td>280</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No. observed (%)

Logistic regression was used to compare the trend in [TTTA]n genotype distribution (mean repeat length of an individuals two alleles) between cases and controls: none is apparent for each extra repeat carried (OR 1.02, 95% CI 0.97–1.07). In addition, genotype distributions between cases and controls for the intron 4 TCT Ins/Del and the intron 6 G→T polymorphisms were compared. No significant difference between the cases and controls was observed for any genotype in either polymorphism (Tables III and IV).

We particularly looked at the TTTA repeat alleles reported to have an association with increased risk of breast cancer, namely the [TTTA]12 allele carriers (19), the [TTTA]n allele 2 (21) and the [TTTA]10 allele (20); none of these associations were apparent in our population.

Four studies have now reported on the CYP19 TTTA repeat alleles. We have performed a meta-analysis on the combined results. There is borderline evidence that the [TTTA]12 carriers (19) and [TTTA]n allele 2 (21) have an increased risk of breast cancer while, on the other hand, the [TTTA]10 allele shows a significantly positive association with breast cancer (OR 2.31, 95% CI 1.35–4.10), although there is some evidence for heterogeneity, i.e. the ORs are different in each study (P = 0.07). However, the observed numbers of this allele are very small, only 18 were seen in 2892 controls (0.6%) in the three studies combined, so some of the variation could be due to small sample size.

As the synthesis of oestrogens in post-menopausal women occurs principally in the adipose tissue, variation in aromatase activity in this subset of women may be of greater importance to breast cancer risk. Hence, we examined the effects of CYP19 genotype on breast cancer risk in post-menopausal cases (defined as women >55 years old) and controls, but still no association was observed.
Discussion

Given that oestrogens affect breast cancer risk and breast tissue growth, an attractive hypothesis is that variants in the CYP19 gene which alter the ability of the enzyme to convert androgens to oestrogens might affect breast cancer risk.

There has been only one polymorphism in the CYP19 gene found to date which results in an amino acid change, the exon 7 Arg264Cys substitution (9,16–18,21). This has been shown to have no effect on breast cancer risk in the Japanese population (18). We cannot find the Cys allele of the Arg264Cys polymorphism and, in agreement with Sourdaine et al. (17), can conclude that it does not have a frequency of >8% in the British population. Our population sample size did not give us the power to look at low frequency alleles with low penetrance, hence no further studies were carried out on this polymorphism. None of the other reported polymorphisms alter the coding region of the gene: a silent G276A (Val80) in exon 3 (17); a C1558T in the non-coding region of exon 10 (17); a TTATA repeat in intron 4 (15); and numerous other intronic changes (21). Hence, these polymorphisms are unlikely to have a direct effect on the activity of aromatase, although Kristensen et al. (19) have postulated a role for the intron 4 TTATA polymorphism in differential RNA splicing. Thus far, we have no common variant of the CYP19 gene that has been reported which demonstrably affects its activity.

We have demonstrated linkage disequilibrium between two polymorphisms in intron 4 and one in intron 6 of the gene: the three polymorphisms generate only three common haplotypes. We can see no association of breast cancer risk with any markers in these haplotypes. However, in a very recent paper, Siegelmann-Danieli and Buetow (21) report two polymorphisms 5′ of exon I1. It is not certain if these are in functional domains, but from the limited data presented, they do not appear to be in strong linkage disequilibrium with the TTATA repeat. This indicates that linkage disequilibrium does not extend across the entire gene and so from our present study we cannot exclude variations at the 5′-end of the gene from altering breast cancer risk. So far, there have been no polymorphisms identified in the various forms of exon I or their promoters; if there were, these would be worth further study, particularly if they affect the exons or promoters used in mammairy adipose tissue or ovaries (25,26).

We were unable to replicate the results of Kristensen et al. (19) for the TTATA12 allele in our sample or those of Siegelmann-Danieli and Buetow (21) for TTATA allele 2 or Haiman for the TTATA10 allele (20). Furthermore, there was no difference in the allele distribution or genotype distribution between cases and controls for any of the polymorphisms studied.

Only the TTATA10 allele showed a significant association in the meta-analysis, but it is very rare, with a carrier frequency of 1.2% (allele frequency 0.6%) and the functional significance is not apparent. In terms of public health, this allele would have no great effect; the population attributable risk of breast cancer due to this allele is calculated to be 1.57% (95% CI 0.43–3.65%). However, this result may point to the TTATA repeat having a function in itself, as Kristensen et al. (19) has suggested.

Although we were unable to find an association of these CYP19 gene polymorphisms with breast cancer risk in this unselected British population of breast cancer cases under the age of 65, it remains plausible that common variants in the CYP19 gene may exist which have an effect on the expression or activity of Cyp19 in certain tissues. These could, in turn, alter levels of oestrogen metabolites and affect breast cancer risk in particular subsets of females. More detailed epidemiological studies are required to test these possibilities.

Acknowledgements

We would like to thank Victoria Basham, Joanna Dearden, Nicola Foster, Jane Gregory, Patricia Harrington, Carol Houghton, Julian Lipscombe, Simon McBride, Carole Pye, Karen Redmond, Paul Russell and the EPIC management team (Sheila Bingham, Nick Day, Kay–Tee Khaw and Suzie Oakes) for access to and preparation of the patient and control DNA samples. We would also like to thank Susan Ramus for valuable discussions. This work was funded by the Cancer Research Campaign (CRC). B.A.J.P. is a CRC Gibb Fellow. F.D. is a Hitchings-Eliion Fellow of the Burroughs-Wellcome Fund.

References


Received July 1, 1999; accepted September 23, 1999