Constitutional genetic variation at the human aromatase gene (Cyp19) and breast cancer risk

N Siegelmann-Danieli* and KH Buetow†

Division of Population Science, Fox Chase Cancer Center, Philadelphia, PA, USA

Summary The activity of the aromatase enzyme, which converts androgens into oestrogens and has a major role in regulating oestrogen levels in the breast, is thought to be a contributing factor in the development of breast cancer. We undertook this study to assess the role of constitutional genetic variation in the human aromatase gene (Cyp19) in the development of this disease. Our genotyping of 348 cases with breast cancer and 145 controls (all Caucasian women) for a published tetranucleotide repeat polymorphism at intron 4 of the Cyp19 gene revealed the presence of six common and two rare alleles. Contingency table analysis revealed a significant difference in allelic distribution between cases and controls ($\chi^2 = 13.52, P = 0.019$). The allele measuring 171 bp was over-represented in cases; of 14 individuals homozygous for this allele, 13 were cases. These individuals had a higher incidence of cancer in family members and an earlier age at diagnosis than other cases. In sequencing Cyp19’s coding exons and regulatory regions, we discovered a perfect association between a silent polymorphism (G→A at Val80) and the high-risk genotype. Our conclusion is that constitutional genetic variation at the Cyp19 locus is associated with the risk of developing breast cancer, with the 171-bp allele serving as the high-risk allele.

Keywords: breast cancer; aromatase gene; oestrogens; Cyp19

This work examined the role of constitutional genetic variation at a candidate locus, the human aromatase gene (Cyp19), in breast cancer development. Cyp19 encodes for a cytochrome P450 aromatase, an enzyme catalysing the conversion of androgens into oestrogens (Tan and Muto, 1986; Mendelson et al, 1990). Aromatase activity has been demonstrated in multiple tissues including normal and transformed breast tissues (James et al, 1987; Bulun and Simpson, 1994; Santen et al, 1994; Sasano et al, 1994). Several reports suggest a positive feedback between breast tumours and aromatase activity in neighbouring breast parenchyma (James et al, 1987; Sasano et al, 1994; Schmidt and Loffler, 1994; Purohit et al, 1995). This suggests that local oestrogen production within the breast, by the aromatase enzyme, might affect breast cancer development and progression.

The Cyp19 gene has been previously characterized and mapped to chromosome 15q21.1 (Chen et al, 1988). It is a single copy gene which spans over more than 70 kb, with the translated exons II–X spanning over only 30 kb (Harada, 1988; Harada et al, 1990; Toda et al, 1990). Multiple non-translated exons I are located at the 5′ region and control the gene expression in a tissue-specific manner and under complex hormonal regulation (Mahendroo et al, 1993; Means et al, 1991; Harada et al, 1993; Toda and Shizuta, 1993; Honda et al, 1994; Toda et al, 1995; Zhao et al, 1995) (Figure 1). These regulatory regions might also participate in the pathogenesis of malignant breast transformation because different exons I are found to control aromatase expression in normal breast (exon 1.4) and during malignant transformation (exon 1.3 and promoter II) (Agarwal et al, 1996; Utsumi et al, 1996; Zhou et al, 1996a, b).

The open reading frame is identical in all expressing tissues examined to date and consists of 1509 bp encoding a 503-amino-acid residue protein (Corbin et al, 1988; Harada, 1988; Means et al, 1989). Rare syndromes of complete aromatase deficiency have been described at the DNA sequence level (exon 6/intron 6 splicing mutation with an 87-bp insert described by Harada et al, 1992; a compound heterozygosity state for exon 10 mutations described by Conte et al, 1994). A common, high heterozygosity tetranucleotide simple tandem repeat polymorphism (STRP) in intron 4 has been previously described (Polymeropoulos et al, 1991). It is not known, however, whether genetic polymorphism at this locus is associated with specific phenotype.

In this study, association-based gene mapping methods were used to assess whether the Cyp19 locus plays a role in determining breast cancer risk. Patterns of constitutional genetic variation were measured at the Cyp19 intron for STRP (modified after Polymeropoulos et al, 1991) and contrasted in cases with breast cancer and healthy controls. To identify whether specific STRP alleles were in linkage disequilibrium with other mutations, Cyp19 coding and regulatory regions were sequenced in cases identified with the high-risk genotype and controls with the putative low-risk genotype.

**PATIENTS AND METHODS**

**Study population**

Participants were consecutive non-related Caucasian women, aged 27–79 years old, living in the greater Philadelphia region and attending Fox Chase Cancer Center (FCCC) breast cancer and cancer-screening clinics and FCCC network hospitals (cancer clinics and non-cancer-related clinics), or hospital employees.
from 1988 to 1994. Participation was voluntary, and participants were required to sign an institutionally-approved consent form. Cases were women diagnosed with breast cancer, controls were women with no history of cancer other than non-melanomatos skin cancer or cervical carcinoma in situ. The study group included 348 cases and 145 controls with a similar age distribution: mean age (at diagnosis for cases and at enrolment for controls) of 54.3 and 54.9 years old and median age of 53.5 and 55 years old in cases and controls respectively.

Table 1 Primers used to sequence Cyp19 coding regions and major non-translated exons I that are transcribed in adipose, ovarian and placenta tissues (see also Figure 1)

| Exon | Primer sets | Forward oligonucleotide | Reverse oligonucleotide | Annealing temperature (°C) | Size (bp) |
|------|-------------|-------------------------|-------------------------|----------------------------|-----------|
| Cis1 | Cis1        | CCAGGATCAAGGAGACCT       | TTCTAGGAGAAAGCGAAGAAG   | 56                         | 303       |
| Cis2 | Cis2        | GGATTTTTTGGCAAGGAGAG    | CCCAGCTCATCATCCCCAC     | 58                         | 278       |
| I.1  | Ex1.l       | AGAGAAGGAGAAGGAATCTCAGA | ATCTGATGGAGAAGCGAGATG   | 59                         | 310       |
| Ex1.la | AAGGACAGGGTTTGAAGGAGCT | GCAGGTGTTTTCCCTCTGCT   | 66                         | 321       |
| Ex1.lb | AGCCCTGCGCTCTTCTTTTT  | CTCCTCGCTCTTCTGACGCA    | 58                         | 514       |
| I.4  | Ex1.40      | TCTCTGAAAATGTCAGCTGCGA  | CGAGGGTGTCAGGTATTTCA    | 60                         | 155       |
| Ex1.41 | ATGATGCTATTGGAAGTTGGA  | AGCTGAGAACAGACAGAAGA    | 67                         | 347       |
| Ex1.42 | TCTCTGATCCCCAGGAACAG  | TGGCAAGCACAGACAAGCT     | 57                         | 334       |
| Ex1.43 | AAGAAGGCTAGGAGTACCTTGA | TTTTGTGTTCTCTGACTTGG   | 56                         | 420       |
| I.3  | Ex1.3       | ATGGGTCCTTCAAGGAGAAGA   | ATGGGTTGTCAGATTTCCAA    | 62                         | 480       |
| Ex2  | Ex2         | GCAGCTCAATTTACGCCCCTC   | ATGAGTGGCAAAACTACCAA    | 57                         | 246       |
| Ex3  | Ex3         | GTAGAAGCAGAAGAGAAGAGA   | GCAGTTGATAGGTTGAGGA     | 55                         | 374       |
| Ex4  | Ex4         | AGCTGCTCTCCTAGTCAAAGGT  | TTAGCAGTGGCAAGCTGTCG    | 65                         | 414       |
| Ex4a | Ex4a        | AGCTGCTCTCCTAGTCAAGAAT  | TACCTTTCTAAGAGGAAGTGA   | 58                         | 238       |
| Ex5  | Ex5         | TTGCTCTGTTGATGTGAGCA    | GTCTAGATGGTGAAGCTGTA    | 53                         | 242       |
| Ex6  | Ex6         | TCTGAGGCTACCTCTTGGAGGC  | AGAAAGTTACCTGAGAGGCC    | 56                         | 288       |
| Ex7  | Ex7         | CATGAGAATAATGCTTGTTGGC  | GGCTGTTTGGATTGGGAATT    | 56                         | 335       |
| Ex8  | Ex8         | AGTCTGGAGCATCTCCTTATT   | ATGATAGCATGGTGAAGAC     | 52                         | 324       |
| Ex9  | Ex9         | CACAGGTTAGGACGAACATAA   | GCTCTTACATTTTCTGCA      | 57                         | 257       |
| Ex10 | Ex10.1      | TGAATTACAAACAGAAGAGCTGAG | AACTTGGGCCTGCTGTTTCT    | 57                         | 538       |
| Ex10.2 | TCGTCTCTCGCTACCAAGC   | ACACTAAGCCAGGTTGGTGG   | 56                         | 285       |
| Ex10.3 | TGGTTGATTACAGCACCATA  | TCTCTGTTGTTGACCACATTA   | 56                         | 489       |
| Ex10.4 | AATGTAGGCTTGTAGTCTT   | TCTCTGTTGAGACAACATAT    | 59                         | 480       |
| Ex10.5 | TGAATCAGTTGATGTCAGT   | TTTTGGGAGTGGACACCTGA    | 53                         | 279       |

*Ex4a was designed to amplify exon 4 and its immediate flanking region, whereas primers Ex4 amplified also the intron 4 STRP region. The sequencing products in both cases and controls did not include the 35-bp region located at ~80 bp to ~45 bp upstream to exon 8 according to gene bank data. Primers Ex10.1 were used to amplify the coding region of exon 10, and Ex10.2 to 10.5 to amplify its 3' UTR up to the second polyadenylation site.

Information on tumour characteristics at time of diagnosis was abstracted retrospectively through review of the computerized medical charts. This review was blind with respect to the outcome of any genetic test used for this study. The following data were obtained: staging status according to the TNM system (American Joint Committee on Cancer, 1992); histopathological type and grade; fibrocystic changes and ductal carcinoma in situ (DCIS); oestrogen and progesterone receptor status; bilateral breast cancer; and clinical or gross-pathological multifocal tumours. These data were available for about 50–80% of participants. In 88%, staging information was based on pathological finding at time of diagnosis; in only 12% was it clinically based. The majority of cases had invasive breast tumours (96%), with 84% diagnosed with T1 and 2 (tumour up to 5 cm in greatest dimension) and 12% with T3 and 4 (tumours more than 5 cm in greatest dimension or with dermal/chest wall invasion). Axillary nodal metastases were noted in 38% of cases; in 74%, the histology type was ductal carcinoma and in 8% it was lobular carcinoma. The rest had either both or adenocarcinoma not otherwise specified. Histology grade was well to moderate in 54% and steroid receptors were positive in 63% of cases. The cases in our population had a relatively low incidence of DCIS in comparison with 1991 National Institute of Health Surveillance, Epidemiology, and End Results (SEER) data (4% vs 12.4% respectively). In other respects, the cases were comparable to women newly diagnosed with invasive breast cancer in the United States during the study period (Ries et al, 1994).

Information on family history of cancer was available for 262 cases. A positive family history of breast cancer in a first-degree family member was noted in 60 families (23%). Family history consistent with hereditary breast or breast and ovarian syndromes.
Chi-squared analysis revealed the allele distribution to be significantly different between cases and controls at $P = 0.019$. $\chi^2$ 5df = 13.52.

### RESULTS

#### Genetic mapping of the Cyp19 locus

Genotyping was first conducted on the CEPH reference panel. The CYP19 locus was observed to be in Hardy–Weinberg equilibrium and transmitted in a Mendelian fashion. To confirm that the locus under investigation was indeed the CYP19 locus previously described in the literature, the STRP was genetically mapped using the CEPH genotype resource (Buetow, 1996). Significant pairwise linkage was observed between the CYP19 STRP and six markers localized to human chromosome 15 (D15S220, D15S117, D15S125, D15S131, D15S114 and D15S175). Complete linkage (no recombination) was observed between CYP19 and D15S220 (lod score = 16.26). Multipoint linkage analysis localized CYP19 with lod 3 support to the interval defined by loci D15S172 and D15S117 in the Cooperative Human Linkage Center’s (CHLC) version 4.0 recombination-minimization map (http://WWW.CHLC.ORG). The maximum likelihood location for the locus was in the 3.5-cM interval defined by D15S648–D15S117, and is in agreement with prior mapping to 15q21.1 (Chen et al., 1988).

#### Cyp19 allele’s distribution in cases and controls

Cases and controls analysed by autoradiography showed six common alleles with sizes ranging from 168 bp to 191 bp (allele 1, 168 bp; allele 2, 171 bp; allele 3, 175 bp; allele 4, 183 bp; allele 5, 187 bp; allele 6, 191 bp). Two rare alleles of sizes 179 bp and 195 bp were also identified in two cases and three controls. Table 2 summarizes the frequencies of the six common alleles in the study participants (692 alleles for cases and 284 alleles for controls). The 171-bp allele (allele 2) was over-represented in cases (odds ratio 1.47, confidence interval 0.993–2.17), whereas the 191-bp allele (allele 6) was over-represented in controls. Contingency table analysis revealed the allelic distribution to be significantly different between cases and controls ($\chi^2$ 7df = 16.07, $P = 0.019$). Inclusion of the two rare alleles in the analysis did not significantly affect the results ($\chi^2$ 7df = 16.07, $P = 0.024$).

To rigorously determine absolute allele size in base pairs, we genotyped samples from 78 individuals using the ABI 373 automated electrophoresis apparatus and software. As expected, the results for the six common alleles were 100% concordant with those obtained by autoradiography. Genotyping results by both methods are shown in Figure 2 (autoradiography results from the CEPH reference panel A, and genotyping Cyp19 locus on the ABI 373 in B).

#### Cyp19 genotype’s distribution in cases and controls: identifying a high-risk genotype

Genotype analysis revealed that homozygotes for the 171-bp allele were 5.4 times more likely to be in the cases group, with a
frequency of 3.78% and 0.70% in cases and controls respectively. Moreover, of 14 study participants homozygous for the 171-bp allele, 13 were cases. All 13 cases had unilateral invasive breast tumours, with comparable tumour characteristics to those of other cases (data not shown). They had a median age at diagnosis 5.5 years younger than other cases (mean and median age at diagnosis were 50.9 and 48 years in cases homozygous for the 171-bp allele, 54.2 and 53.5 years in total case group respectively), but this difference was not statistically significant. The only control individual homozygous for the 171-bp allele was 46 years of age at time of inclusion.

Family data were available for 13 of these individuals and are summarized in Table 3. In ten individuals, there was a first-degree family relative with a solid cancer or leukaemia. In six cases, there was a family history of breast cancer (in families 6, 9 and 10 it is suggestive of putative hereditary breast or breast/ovarian cancers). Although individuals homozygous for the 171-bp allele had a higher prevalence of a positive history of breast cancer in a first-degree family relative compared with other cases, this difference was not statistically significant ($P = 0.12$).

Two additional genotypes were disproportionally distributed among cases and controls. Heterozygotes for the 171-/187-bp alleles were 1.85 times more likely to be in the cases group (frequency of 13% and 7% in cases and controls respectively), and heterozygotes for the 187-/191-bp alleles were five times more likely to be controls (frequency of 1.12% and 5.63% in cases and controls respectively). No participants were homozygous for the 191-bp allele, which was observed to be over-represented in controls (Table 2). Other genotypes were either rare (frequency of less than 1.5%) or occurred with a similar frequency among cases and controls.

We concluded that the 171-bp allele represents a high-risk allele. Individuals homozygous for this allele are considered to carry a high-risk genotype. Heterozygotes for the 187-/191-bp alleles are considered to carry a putative low-risk genotype.

No significant association was observed between the occurrence of one or more copies of the 171-bp allele and the following cases’ characteristics: menopausal status (using age 50 as a cut-off point); invasive vs non-invasive tumours; tumour size (T-stage 1 and 2 vs 3 and 4); lymph node involvement; multifocal or bilateral tumours; histology type and grade; background histology changes and receptor status.

### Sequencing results

To identify whether specific STRP allele variants were in linkage disequilibrium with other mutations at the $Cyp19$ gene, sequencing results were contrasted for cases with the high-risk STRP genotype (homozygous for the 171-bp allele) and controls with the low-risk STRP genotype (heterozygotes for the 187-/191-bp alleles). Sequencing efforts were targeted at the coding regions of $Cyp19$ (including the 3’ UTR of exon 10 up to the second polyadenylation site), and at major non-translated regulatory exons I for adipose, ovarian and placenta tissues (see Figure 1). At least five individuals in each group were sequenced. Sequence variants confirmed by the reverse sequence were extended to additional individuals (up to 12 cases and the one control) homozygous for the 171-bp allele (DNA for one high-risk case was not available for sequencing reactions).

Table 4 summarizes the identified $Cyp19$ sequence variants and their frequencies among cases with the high-risk genotype and controls with the low-risk genotype (the 12 cases and one control carrying the high-risk genotype are reported together, as they displayed the same sequence variants). Of identified sequence variants, two occurred in coding exons of the $Cyp19$ gene, and were detected only among individuals carrying the high-risk genotype. The first, a silent variation at exon 3 ($G\rightarrow A$, Val80), occurred in its homozygous state in all individuals carrying the high-risk

**Figure 2** Genotyping for the $Cyp19$ locus. (A) $Cyp19$ genotyping by the autoradiography method: CEPH family no. 1354. (B) Genotyping $Cyp19$ locus on ABI 373 sequencer.
Table 3  Family history of cancer in 14 Caucasian women homozygous for the 171-bp allele

| Individual's number | Age (years) | First degree                                        | Second or more degree                              |
|---------------------|-------------|----------------------------------------------------|----------------------------------------------------|
| Cases               |             |                                                    |                                                    |
| 1                   | 52          | Mother, breast cancer at 51                        | Father, lung cancer at 62                          |
| 2                   | 51          |                                                    | M uncle, lung cancer                                |
| 3                   | 58          | Brother, lung cancer at 55                         | P uncle, lung cancer                                |
| 4                   | 47          | Mother, breast cancer at 30                        | M grandmother's brother, prostate cancer at 86     |
| 5                   | 60          | Brother, leukaemia at 59                           | M grandmother, leukaemia at 60                     |
| 6                   | 59          | Sister, ovarian cancer at 49                       | M grandmother's brother, cancer at 89             |
| 7                   | 72          | Daughter, breast cancer at 49                       |                                                    |
| 8                   | 45          | (Not known)                                        | P aunt, breast cancer at 70                        |
| 9                   | 48          |                                                    | P cousin, breast cancer at 28                       |
| 10                  | 47          | Mother, breast cancer                              | P uncle, colon cancer                               |
| 11                  | 36          | Father, lung cancer                                | P uncle, CNS cancer                                 |
| 12                  | 47          | Mother, uterine cancer                             | P uncle, spine cancer                               |
| 13                  | 40          |                                                    | M great aunt, breast cancer                         |
| Control             | 46          | Brother, testicular cancer                         |                                                    |

*In seven individuals, data were available from self-administered questionnaires, and in six individuals by reviewing the medical chart with confirmation by a phone call. In individual no. 8 data were not available. M, maternal; P, paternal.
Cyp19 and breast cancer risk

British Journal of Cancer (1999) 79(3/4), 456–463
© Cancer Research Campaign 1999

Table 4

| Sequence variants in coding exons | Cases with the high-risk genotype (5–13 tested) | Controls with the low-risk genotype (five tested) |
|----------------------------------|-----------------------------------------------|-----------------------------------------------|
| Exon 3 G→A at Val80               | (13 tested) AA, 1.0                           | (Five tested) GG, 1.0                           |
| Exon 7 C→T in codon 264 (Arg→Cys264) | (Nine tested) CC, 0.9 / CT, 0.1 | (Five tested) CC, 1.0                           |

| Variants in non-coding regions | Cases with the high-risk genotype (5–13 tested) | Controls with the low-risk genotype (five tested) |
|--------------------------------|-----------------------------------------------|-----------------------------------------------|
| G→A at – 483 bp to exon 1.1 start | (Five tested) GG, 0.6 / GA, 0.2 / AA, 0.2 | (Five tested) GG, 0.8 / GA, 0.2 |
| C→T at – 41 bp to exon 1.1 start | (12 tested) CC, 0.66 / CT, 0.17 / TT, 0.17 | (Five tested) CC, 0.8 / CT, 0.2 |
| Intron 4 TTTA repeats at intron 4 | (Five tested) Short repeat | (Five tested) Long repeat |
| Intron 5 G→T at – 16 bp to exon 6 start | (12 tested) TT, 0.6 / GT, 0.4 | (Five tested) GT, 0.2 / GG, 0.8 |
| Intron 6 T→A at + 36 bp to intron 6 start | (12 tested) AA, 0.6 / TA, 0.4 | (Five tested) TA, 0.2 / TT, 0.8 |
| Intron 7 C→T at + 26 bp to intron 7 start | (Nine tested) CC, 0.1 / CT, 0.2 / TT, 0.7 | (Five tested) CC, 1.0 |
| 3′ UTR T→C at + 19 bp to intron 10 start | (11 tested) CC, 0.6 / TC, 0.4 | (Five tested) TC, 0.2 / TT, 0.8 |
| 3′ UTR G→T at + 162 bp to intron 10 start | (11 tested) GG, 0.2 / GT, 0.35 / TT, 0.45 | (Five tested) GG, 0.8 / GT, 0.2 |
| 3′ UTR T→C at + 1294 to intron 10 start | (Five tested) TT, 1.0 | (Five tested) TT, 0.6 / TC, 0.4 |

*The 12 cases and one control carrying the high-risk genotype are reported together because they displayed the same sequence variants. *Additionally, one CC homozygote and three CT heterozygotes were observed among high-risk cases for which sequencing results could be obtained in only one direction. *The sequencing results for exon 4 and its 3′ flanking region indicated that allele 2 would produce a 171-bp product and allele 5 a 187-bp product. *Two additional CC homozygotes and two CT heterozygotes were observed among high-risk cases for which sequencing results could be obtained in only one direction.

STRP genotype and in none of the low-risk controls (Figure 3). The second, a heterozygous state for C→T in exon 7 (Arg→Cys264), appeared in one of nine high-risk cases sequenced in both directions, and in three of four additional high-risk cases with sequencing results obtained in only one direction (frequency of 0.1–0.3 among high-risk cases). All other identified variants were in non-translated regions of the Cyp19 gene. Of note, in concordance with the results of the automated electrophoresis genotyping analysis, the sequencing results for exon 4 and its 3′ flanking region indicated that allele 2 would produce a 171-bp product and allele 5 a 187-bp product.

The 168-bp and 187-bp alleles were the most common alleles in the study population. While the latter (together with the low-risk 191-bp allele) was found to be associated with G in Val80, we elected to check this region in individuals carrying the 168-bp STRP allele. By sequencing ten individuals homozygous for the 168-bp allele (five cases and five controls), we found all to be homozygous for G in codon Val80.

**DISCUSSION**

Our study examined the association of constitutional genetic variation in the human aromatase gene (Cyp19) with breast cancer occurrence among Caucasian women in the greater Philadelphia region. Our cases population displayed typical features of non-selected cases with invasive breast cancer diagnosed in the United States during the study period, both in tumour characteristics and family history (see Patients and methods section). By typing close to 1000 chromosomes, we were able to identify eight alleles in the study population, sizes ranging from 168 bp to 195 bp [Polymeropoulos et al (1991) identified five alleles sized 154–178 bp, on a typing of only 46 chromosomes]. To facilitate future replicative studies, accurate allele sizes were determined by typing 78 samples using florescence-based, automated electrophoresis, and confirmed the allele’s transmission in co-dominant segregation in 12 CEPH families.

Contingency table analysis revealed a statistically significant difference in allelic distribution between cases and controls. Specifically, the 171-bp allele was over-represented in cases, and the 191-bp allele was more common in controls. All but one of 14 individuals homozygous for the 171-bp allele were cases. Though not formally significant, these homozygous cases were diagnosed at a younger mean and median age than the total cases’ group, and displayed a remarkably high incidence of breast and other cancers in their family histories. Preliminary data in African–American women show over-representation of individuals homozygous for...
the 171-bp allele in cases with breast cancer (5 out 13 cases; one out of eight controls). These results provide additional support for the suggestion that constitutional genetic variation at or near the Cyp19 locus is associated with the risk of developing breast cancer, with the 171-bp allele serving as the high-risk allele or at least its surrogate.

Our DNA sequencing efforts focused on the coding regions of the CYP19 locus and the major non-translated exons I controlling its expression in adipose, ovarian and placental tissues. It included the non-translated exons I, which control aromatase expression in normal breast adipocytes (exon I4) and during malignant breast transformation (exon I3 and promoter II). By analysing sequences obtained for individuals with the high-risk genotypes (homozygous for the 171-bp allele) and controls with the putative low-risk genotypes (heterozygous for the 187/-191-bp alleles), we identified several common nucleotide variants. Of particular interest is the silent variation in exon 3 (G→A, Val80), which was observed to be in complete association with the high-risk genotype and was not observed in 15 sampled individuals carrying the common alleles (168-bp and 187-bp alleles) or the low-risk allele (191 bp allele). Further studies should assess its occurrence in a larger population of cases and controls, and test whether this variant can affect aromatase activity by modifying Cyp19RNA splicing or stability. This variant was previously observed by Sourdaine et al (1994) in genomic DNA from two breast tumours selected for insensitivity to the aromatase inhibitor 4-hydroxyandrostenedione (4-OHA), but not in two other breast tumours showing slightly lower basal aromatase activity and sensitivity to inhibition by 4-OHA. Sourdaine et al (1994) also observed the T→C variant at +19 to intron 10 start in the two resistant tumours (which in our work was more common among controls carrying the low-risk genotype), and a heterozygous state for the exon 7 C→T (Arg→Cys264) in only one of them. A fifth breast tumour evaluated by Sourdaine et al (1994), with no detectable aromatase activity, did not show any of the sequence variants detected in the tumours resistant to 4-OHA. Sourdaine et al (1994) identified no other coding region sequence variants in any of the breast tumours; the intron 4 STRP region and the non-translated regulatory exons I were not evaluated in their work. Taken together with our work, it is possible to theorize that the G→A variant in codon Val80, which occurred in complete association with the high-risk STRP genotype, is related to the phenotype of slightly elevated basal aromatase activity with resistance to inhibition by 4-OHA. However, only direct analysis addressing the effects of G→A (Val80) variation on aromatase expression in the breast can prove it, especially considering that Sourdaine et al (1994) selected the breast tumours in their work by their response to inhibition by 4-OHA and not their baseline aromatase activity.

Another possibility is that the intron 4 STRP itself produces an intrinsic effect on aromatase expression. Triplet repeat variations in non-coding regions of the FMRI gene and MK-PK gene are known to result in fragile X syndrome and myotonic dystrophy respectively (Fu et al, 1991; Verkerk et al, 1991; Brook et al, 1992; Mahadevan et al, 1992; Kunst and Warren, 1994). Constitutional genetic variation in variable tandem repeats located some 1000 bp downstream from the coding region of the H-RAS1 proto-oncogene have been shown to be associated with cancer risk in case–control studies (Krontiris et al, 1985, 1993). Functional studies have also suggested that the variable tandem repeats themselves might interact with transcription factors affecting expression of associated genes, or might alter RNA stability (Green and Krontiris, 1993; Trepicchio and Krontiris, 1993; Kennedy et al, 1995). Finally, it is possible that the high-risk STRP allele is associated with a novel breast-specific exon I not yet described. Recently, Stratakis et al (1996) reported a family with a transmitted phenotype of increased aromatase activity, which was observed to be associated with a specific STRP variant and a novel non-translated exon I. This finding implies the existence of an as-yet- unidentified variation in the 5′ flanking sequence, which controls aromatase expression and is associated with the high-risk STRP variant.

In BALB/cD2 mice, integration of the mouse mammary tumour virus to the In5 locus (located at the 3′ UTR of mouse Cyp19 gene) resulted in up-regulation of the aromatase gene; it was associated with the development of preneoplastic malignant tumours (Durgam and Tekmal, 1994; Tekmal and Durgam, 1995). In female Sprague–Dawley rats, treatment with the aromatase inhibitor fadrozole hydrochloride prevented the appearance of spontaneous benign and malignant mammary tumours (Gunson et al, 1995). Understanding the biological significance of the human high-risk allele may lead to prevention of breast cancer in high-risk individuals identified by Cyp19 genotyping.

In conclusion, our study shows a significant difference in constitutional genetic variation at intron 4 of the human Cyp19 gene between cases and controls (Caucasian women). We identified a high-risk allele at this locus, although its biological significance has yet to be determined. Confirmatory studies are necessary to conclude that Cyp19 genotyping can conclusively identify women at an increased risk for developing breast cancer. Ultimately, it is our hope that an understanding of the biological significance of the high-risk allele will bring us closer to a medical prevention of breast cancer, in which high-risk individuals identified with Cyp19 genotyping are offered specific treatment affecting aromatase activity.

ACKNOWLEDGEMENTS

This work is supported by an ASCO Young Investigator Award sponsored by Zeneca, a clinical investigator training grant CA01728 and a pilot project on a cancer genetics supplement to USPHS Institutional core grant CA06927 from the National Institutes of Health (to Dr Siegelmann-Danieli), and by an appropriation from the Commonwealth of Pennsylvania.

The authors acknowledge Katherine A, McGlynn PhD for kindly reviewing the manuscript, Michael N, Edmonson for analytical assistance and John Q Zhang and for technical assistance.

REFERENCES

Agarwal VR, Balun SE, Leitch M, Rohrich R and Simpson ER (1996) Use of alternative promoters to express the aromatase cytochrome P450 (CYP19) gene in breast adipose tissues of cancer-free and breast cancer patients. J Clin Endocrinol Metab 81: 3843–3849
American Joint Committee on Cancer (1992) Manual for staging of cancer, 4th edn. In Breast, pp. 149–154. JB Lipincott: Philadelphia
Berrino F, Mati P, Michelini A, Gianfranco B, Krogh V, Sciasio R, Pisani P, Panico S and Secreto G (1996) Serum sex hormone levels after menopause and subsequent breast cancer. J Natl Cancer Inst 88: 291–296
Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, Sohn R, Zemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shellbourne P, Buxton J, Jones C, Javonen V, Johnson K, Harper PS, Shaw DJ and Housman DE (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68: 799–808
Buetow KH (1996) Genetic mapping. In Current Protocols in Human Genetics. Dracopo1 N, Haines JL, Korf BR, Morton CC, Seidman CE, Seidman JG, Moir DT and Smith D (eds.), section 1, pp. 1.0–1–10.8. John Wiley & Sons: Brooklyn, NY

Bulun SE and Simpson ER (1994) Regulation of aromatase expression in human tissues. Breast Cancer Res Treat 30: 19–29

Chen S, Besman MJ, Sparkes RS, Zollman S, Kisak I, Mohandas T, Hall PF and Shively JE (1988) Human aromatase: cDNA cloning, Southern blot analysis, and assignment of the gene to chromosome 15. DNA 7: 27–38

Conte FA, Grumbach MM, Ito Y, Fisher CR and Simpson ER (1994) A syndrome of female pseudohermaphroditism, hypergonadotropic hypogonadism, and multicystic ovaries associated with missense mutations in the gene encoding aromatase (P450arom). J Clin Endocrinol Metab 78: 1287–1292

Corbin CJ, Graham-Lorenze S, McPhaul M, Mason JI, Mendelson CR and Simpson ER (1988) Isolation of a full-length cDNA insert encoding human aromatase system cytochrome P-450 and its expression in nonsteriodogenic cells. Proc Natl Acad Sci USA 85: 8948–8952

Durgam VR and Tekmal RR (1994) The nature and expression of int-5, a novel MMTV integration locus gene in carcinogen-induced mammary tumors. Cancer Lett 79: 179–186

Fu Y-H, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJMH, Harada N, Yamada K, Saito K, Kibe N, Dohmae S and Takagi Y (1990) Structural characterization of the human estrogen synthetase (aromatase) gene. Biochim Biophys Acta 1011: 725–732

Harada N (1988) Cloning of a complete cDNA encoding human aromatase: immunocommunication and sequence analysis. Biochim Biophys Acta 1011: 725–732

Harada N, Yamada K, Saito K, Kibe N, Dohmae S and Takagi Y (1990) Structural characterization of the human estrogen synthetase (aromatase) gene. Biochem Biophys Acta 1011: 365–372

Harada N, Ogawa H, Shozu M, Yamada K, Suhara K, Nishida E and Takagi Y (1992) Biochemical and molecular genetic analyses on placental aromatase (P450Arom) deficiency. J Biol Chem 267: 4781–4785

Harada N, Utsumi T and Takagi Y (1993) Tissue-specific expression of the human aromatase cytochrome P-450 gene by alternative use of multiple exons 1 and promoters, and switching of tissue-specific exons 1 in carcinogenesis. Proc Natl Acad Sci USA 90: 11312–11316

Honda S, Harada N and Takagi Y (1994) Novel exon 1 of the aromatase specific for aromatase transcripts in human brain. Biochim Biophys Acta 1191: 1153–1160

James VH, McNeill JM, Lai LC, Newton CJ, Gihlick MW and Reed MJ (1987) Aromatase activity in normal breast and breast tumor tissues: in vivo and in vitro studies. Steroids 50: 269–279

Kennedy GC, German MS and Rutter WJ (1995) The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. Nature Genet 9: 293–298

Krontiris TG, D’imartino NA, Colb M and Parkinson D (1985) Unique allelic restriction fragments of the human Ha-ras locus in leukocytes and tumor DNAs of cancer patients. Nature 313: 369–374

Krontiris TG, Delvin B, Karp D, Robert NJ and Risch N (1993) Association between risk of cancer and mutations in the HRAS1 minisatellite locus. N Engl J Med 329: 517–523

Kont KB and Warren ST (1994) Cryptic and polar variation of the fragile X repeat (FMR-1) Tissue-specific and –5 regulatory activity of the human CYP19 (aromatase) gene. Biochim Biophys Acta 1215: 13–116

Means GD, Kilgore MW, Mahendroo MS, Mendelson CR and Simpson ER (1991) Tissue-specific promoters regulate aromatase cytochrome P450 gene expression in human ovary and fetal tissues. Mol Endocrinol 5: 2005–2013

Mendelson CR, Means GD, Mahendroo ML, Corbin CJ, Steinkamp MP, Graham-Lorenze S and Simpson ER (1990) Use of molecular probes to study regulation of aromatase cytochrome P-450. Biol Reprod 42: 1–10

Miller SA, Dykes DD and Homeister JW (1998) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215

Mohrmannopoulus MH, Xiao H, Rath DS and Merril CR (1991) Tetranucleotide repeat polymorphism at the human aromatase cytochrome P-450 gene (CYP19). Nucleic Acids Res 19: 195

Purohit A, Ghilchik MW, Duncan L, Wang DY, Singh A, Walker MM and Reed MJ (1995) Aromatase activity and interleukin-6 production by normal and malignant breast tissues. J Clin Endocrinol Metab 80: 3052–3058

Ries LAG, Miller BA, Hankey BF, Kosary CL, Harris A and Edwards BK (1994) 1973–1991 Tables and graphs, National Cancer Institute. In SEER Cancer Statistics Review, p. 128. Publication no. 94-2789. NIH: Bethesda, MD

Sasano H, Nagura H, Harada N, Goukon Y and Kimura M (1994) Immunolocalization of aromatase and other steroidogenic enzymes in human breast disorders. Hum Pathol 25: 530–535

Schmidt M and Loffler G (1994) The human breast cancer cell line MDA-MB231 produces an aromatase stimulating activity. Eur J Cell Biol 63: 96–101

Soukup J, Parker MG, Telford J and Miller WR (1994) Analysis of the waromatase cytochrome P450 gene in human breast tumors. J Mol Endocrinol 13: 331–337

Stratakis CA, Vottero A, Brodie A, DeArkine D, Lu O, Mitsiades CS, Meada T, Yamamoto Y, Sagara Y, Ikeda H and Shizuta Y (1996) Biochemical and molecular genetics of the syndrome of increased aromatase activity: segregation with a marker from within the P450arom gene and evidence for an aberrant alternative splicing of its 5’-end mRNA. Am J Hum Genet 59: A43

Tan L and Nuto M (1986) Purification and reconstitution properties of human placental aromatase. A cytochrome P450-type monoxygenase. Eur J Biochem 156: 243–250

Tekmal RR and Durgam VR (1995) The overexpression of int-5/Aromatase, a novel MMTV integration locus gene, is responsible for D2 mammary tumor cell proliferation. Cancer Lett 88: 147–155

Toda K and Shizuta Y (1993) Molecular cloning of a cDNA showing alternative splicing of the 5'-untranslated sequence of mRNA for human aromatase P-450. Eur J Biochem 213: 383–389

Toda K, Terashima M, Kawamoto T, Sumimoto H, Yokoyama Y, Kuribayashi I, Mitsuhashi Y, Yue W, Flor AW, Garrica A, Mitsisidas CS and Chronos GP (1990) Structural and functional characterization of human aromatase P-450 gene. Eur J Biochem 193: 559–565

Toda K, Yang L-X and Shizuta Y (1995) Transcriptional regulation of the human aromatase cytochrome P450 gene expression in human placental cells. J Steroid Biochem Mol Biol 53: 181–190

Trepcichio WL and Krontiris TG (1993) IGH minisatellite suppression of USF-binding-site- and Em-mediated transcriptional activation of the adenovirus major late promoter. Nucleic Acids Res 21: 977–985

Verkerk AJMH, Pieretti M, Stulciffe JS, Fu Y-H, Kilh DP, Pizzati A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, van Ommen BJ, Blondon LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey T, Nelson DL, Oostra BA and Warren ST (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65: 905–914

Utsumi T, Harada N, Maruta M and Takagi Y (1996) Presence of alternatively spliced transcripts of aromatase gene in human breast cancer. J Clin Endocrinol Metab 81: 2344–2349

Zhao Y, Mendelson CR and Simpson ER (1995) Characterization of the sequences of the human CYP19 (aromatase) gene that mediate regulation by glucocorticoids in adipose stromal cells and fetal hepatocytes. Mol Endocrinol 9: 340–349

Zhou D, Clarke P, Wang J and Chen S (1996a) Aromatase gene expression and its exon usage in human breast tumors. J Biol Chem 271: 15194–15202