Journal of Human Genetics (2010) 55, 189–193

www.nature.com/jhg

SHORT COMMUNICATION

# Identification of CYP19A1 single-nucleotide polymorphisms and their haplotype distributions in a Korean population

Su-Jun Lee[1], Woo-Young Kim[1], Ji-Yeob Choi[1], Sang Seop Lee[1] and Jae-Gook Shin[1,2]

**Aromatase, encoded by the *CYP19A1* gene, is a key enzyme in the biosynthesis of estrogen. In an effort to screen for *CYP19A1* single-nucleotide polymorphisms (SNPs) in Koreans, the *CYP19A1* gene was directly sequenced in 50 normal subjects. A total of 19 variations were identified: four in exons, ten in introns, six in the 5′-untranslated region (UTR) and one in 3′-UTR. The distribution of *CYP19A1* $(TTTA)_n$ polymorphisms was such that the most frequent allele was $(TTTA)_7$ (66%), followed by $(TTTA)_{11}$ (30%), $(TTTA)_{12}$ (3%) and $(TTTA)_{13}$ (1%). The order of the frequency distribution of *CYP19A1* variations, other than that of the $(TTTA)_n$ variant, was IVS6-106T>G and IVS7-79A>G (57%); 1531C>T (56%); IVS5-16T>G and IVS6+36A>T (54%); −196A>C and −77G>A (49%); IVS2-59A>G and 240A>G (48%); −278C>T (31%); IVS4+27TCTI>D (29%); −144C>T and −588G>A (19%); 790C>T (16%); and other minor alleles with less than 5% frequency. Nineteen variations were used to characterize linkage disequilibrium (LD) structures at the *CYP19A1* locus, which resulted in three LD blocks. Eight tagging SNPs in *CYP19A1* were determined. Identification of *CYP19A1* SNPs with LD blocks and tagging SNPs creates an important resource for genotype–phenotype association studies for estrogen-related phenotypes.**
*Journal of Human Genetics* (2010) **55,** 189–193; doi:10.1038/jhg.2010.6; published online 26 February 2010

**Keywords:** aromatase; CYP19A1; linkage disequilibrium; SNPs; Tag SNPs

## INTRODUCTION

The cytochrome P450 19A1 (*CYP19A1*) gene encodes an aromatase that is responsible for the conversion of androgens to estrogens. *CYP19A1* is located on chromosome 15q21 and spans 123 kb, including nine protein coding exons and a large 5′-untranslated region of 93 kb with alternative tissue-specific promoters.[1] Several untranslated first exons are expressed under the control of a tissue-specific promoter and joined to the second exon through a common splice acceptor site.[2] The expression of aromatase in humans has been demonstrated in gonads, adipose, placenta, skin, brain and bone. Given that estrogens have an important role in cell proliferation,[3] many studies have suggested that variation in estrogen levels contributes to the risk of estrogen-dependent diseases, such as breast cancer,[4] endometrial cancer[5] and prostate cancer.[6] In addition to cancer association studies, *CYP19A1* has been implicated in the risk of obesity in Chinese women,[7] in the risk of essential hypertension in males,[8] as well as in bone mass and fracture risk.[9] One of the important features of aromatase is evidenced by the fact that selective aromatase inhibitors are increasingly used to treat postmenopausal women with estrogen-responsive breast cancer.[10] Therefore, *CYP19A1* coding for aromatase has been an attractive target for identification of common genetic polymorphisms that may account for different levels of estrogen in humans. However, no efforts to resequence the *CYP19A1* gene for a Korean population have been made. To determine the distribution of *CYP19A1* alleles and haplotypes, we conducted direct DNA sequencing of the *CYP19A1* gene in a Korean population for the first time. Analysis of the *CYP19A1* allele distribution and haplotype structure with linkage disequilibrium (LD) was performed.

## MATERIALS AND METHODS

### Study subjects

DNA samples from 50 unrelated, healthy Koreans were obtained from the DNA repository bank at INJE Pharmacogenomics Research Center for direct DNA sequencing (Inje University College of Medicine, Busan, Korea).[11] The Institutional Review Board (IRB) of Busan Paik Hospital (Busan, Korea) approved the research protocol for the use of human DNA from blood samples.

### Sequencing and identification of SNPs

Genomic DNA was prepared from peripheral whole blood using the QiAamp blood kit (Valencia, CA, USA). Primers for PCR amplification were identical to those used previously.[12] All exons, exon–intron boundaries and relevant promoter regions were PCR amplified. PCR contains primers at 0.2 $\mu mol\,l^{-1}$ each, 2 U of *r-Taq* polymerase (Takara Bio, Shiga, Japan), and each dNTP at 0.2 $mmol\,l^{-1}$ per reaction. PCR cycle parameters include initial denaturation (94 °C, 5 min, 1 cycle), 35 cycles of a denaturation/annealing/extension (94 °C,

[1]Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine, Inje University, Busan, South Korea and [2]Department of Clinical Pharmacology, Inje University College of Medicine, Inje University, Busan, South Korea
Correspondence: Profesor S-J Lee, Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine, Inje University, 633-165, Gaegum2-dong, Busanjin-gu, Busan 614-714, South Korea.
E-mail: 2sujun@inje.ac.kr
Received 4 November 2009; revised 4 January 2010; accepted 7 January 2010; published online 26 February 2010

30s/60 °C, 30s/72 °C, 30 s) and a final extension at 72 °C for 10 min. PCR was performed on a GeneAmp PCR 9700 (Applied Biosystems, Foster City,CA, USA). Amplified products were purified using a PCR purification kit (NucleoGen, Ansan, Korea), and sequencing was performed using an ABI Prism 3700XL Genetic Analyzer (Applied Biosystems). A software package, PC Gene (Oxford Molecular, Campbell, CA, USA), was used to identify variants with single-nucleotide substitutions in heterozygous or homozygous mutations.

### Linkage disequilibrium and haplotype analysis

Hardy–Weinberg equilibrium and LD were analyzed by SNPAlyze software (version 4.1; Dynacom, Yokohama, Japan). $|D'|$ and rho square ($r^2$) values were used to access pairwise LD between single-nucleotide polymorphisms (SNPs), as described previously.[11,13]

### Tag SNP selection

Nineteen variations identified by direct DNA sequencing were applied to select tagging SNPs. The program Tagger was used to select tagging SNPs, which combines the simplicity of pairwise methods with the potential efficiency of multimarker approaches (http://www.broad.mit.edu/mpg/tagger/). The efficiency of tagged SNPs depends on the LD between itself and the tag SNP, as measured by the pairwise correlation coefficient ($r_p^2$). We selected a set of tag SNPs in which all known common variants (minor allele frequency $>0.05$) had an estimated $r_p^2>0.8$ with at least one tag SNP. Some SNPs are weakly correlated with other single SNPs but may be efficiently tagged by a haplotype defined by multiple SNPs, reducing the total number of tag SNPs needed. Therefore, we attempted to define the correlation coefficient between each SNP and a haplotype of tagging SNPs as $r_s^2>0.8$.

## RESULT AND DISCUSSION

Genetic polymorphisms in *CYP19A1* may change the function of the aromatase enzyme, resulting in altered levels of estrogen biosynthesis. Thus, its polymorphism may be implicated in the risk of estrogen-related diseases. Although *CYP19A1* is a key enzyme in the process of estrogen synthesis, no resequencing efforts have yet been made in people of Korean descent. In this study, direct DNA sequencing analysis of the *CYP19A1* gene in 50 Koreans revealed a total of 19 variations. A summary of their frequencies identified is provided in Table 1. The locations of SNPs in relation to the genomic structure of *CYP19A1* are shown in Figure 1. $\chi^2$-tests were used to compare observed variants with expected variants in the study population. No deviation from Hardy–Weinberg equilibrium was observed for the SNPs identified ($P>0.05$). Frequencies of functionally important and relatively well-studied *CYP19A1* variants in different ethnic groups were summarized in Table 2. Several studies have been conducted to address the association between *CYP19A1* $(TTTA)_n$ polymorphisms and breast cancer risk as a genetic marker.[14–15] We observed that the most frequent *CYP19A1* (TTTA) allele was $(TTTA)_7$ (66%), followed by $(TTTA)_{11}$ (30%), $(TTTA)_{12}$ (3%) and $(TTTA)_{13}$ (1%) (Table 3). Kim *et al.*[16] reported that the frequency of the low type, which means that the TTTA copy number was under 10, was 55.5% and that of the high type (TTTA copy number with over 10) was 44.5% in 102 Korean control subjects. The similar report in a Korean population[17] is described in Table 3. Although $(TTTA)_8$ was not detected, this allele was the third most frequent allele in Caucasian and African-American populations, occurring at ~12.5 and 2.5%, respectively.[12] In the Korean population, the frequency of $(TTTA)_{12}$ was 3%; however, this allele was detected in 10% of Japanese subjects,[14] indicating that the allele frequency in one population cannot be assumed to be equally applicable in a similarly defined population. The promoter SNP −278C>T, known as a very rare allele in Caucasian and African-American populations, had an occurrence rate of 31% in our study,

**Table 1 Distribution of *CYP19A1* single-nucleotide polymorphisms in a Korean population**

| | | | | Subject (n) | | | |
|---|---|---|---|---|---|---|---|
| | Position[a] | Amino-acid substitution | Reference | wt/wt | wt/mt | mt/mt | Frequency (%) |
| 5-UTR exon 1.1 | −588G>A | | rs7176005 | 33 | 15 | 2 | 19 |
| 5-UTR exon 1.1 | −278C>T | | rs28892002 | 24 | 21 | 5 | 31 |
| 5-UTR exon 1.1 | −144C>T | | rs6493497 | 33 | 15 | 2 | 19 |
| 5-UTR exon 1.4 | −339G>C[b] | | | 49 | 1 | 0 | 1 |
| 5-UTR exon 1.6 | −196A>C | | rs10459592 | 12 | 27 | 11 | 49 |
| 5-UTR exon 1.6 | −77G>A | | rs4775936 | 12 | 27 | 11 | 49 |
| Exon 2 | 115T>C | W39R | rs2236722 | 46 | 4 | 0 | 4 |
| Intron 2 | IVS2-59 A>G | | Rs3759811 | 13 | 26 | 11 | 48 |
| Exon 3 | 240A>G | V80V | rs700518 | 13 | 26 | 11 | 48 |
| Intron 4 | IVS4+27TCT I>D | | rs11575899 | 26 | 19 | 5 | 29 |
| Intron 4 | IVS4+77 (TTTA)7 | | rs57921193 | | | | |
| | *n*=7 | | | | 10 | 28 | 66 |
| | *n*=11 | | | | 14 | 8 | 30 |
| | *n*=12 | | | | 3 | 0 | 3 |
| | *n*=13 | | | | 1 | 0 | 1 |
| Exon 5 | 501C>A | S167S | rs35900050 | 49 | 1 | 0 | 1 |
| Intron 5 | IVS5-16 T>G | | rs4324076 | 12 | 22 | 16 | 54 |
| Intron 6 | IVS6+36 A>T | | rs1143704 | 12 | 22 | 16 | 54 |
| Intron 6 | IVS6-106 T>G | | rs2304463 | 10 | 23 | 17 | 57 |
| Exon 7 | 790C>T | R264C | rs700519 | 36 | 12 | 2 | 16 |
| Intron 7 | IVS7+26 C>T | | rs2289105 | 49 | 1 | 0 | 1 |
| Intron 7 | IVS7-79 A>G | | rs2289105 | 10 | 23 | 17 | 57 |
| 3-UTR | 1531 C>T | | rs10046 | 10 | 24 | 16 | 56 |

[a]The reference sequence used was GenBank accession no. NC_000015. Position is indicated in respect to the start codon ATG *CYP19A1* gene; the A in ATG is +1 and the next base toward to 5′ is −1.
[b]New variant allele indentified.

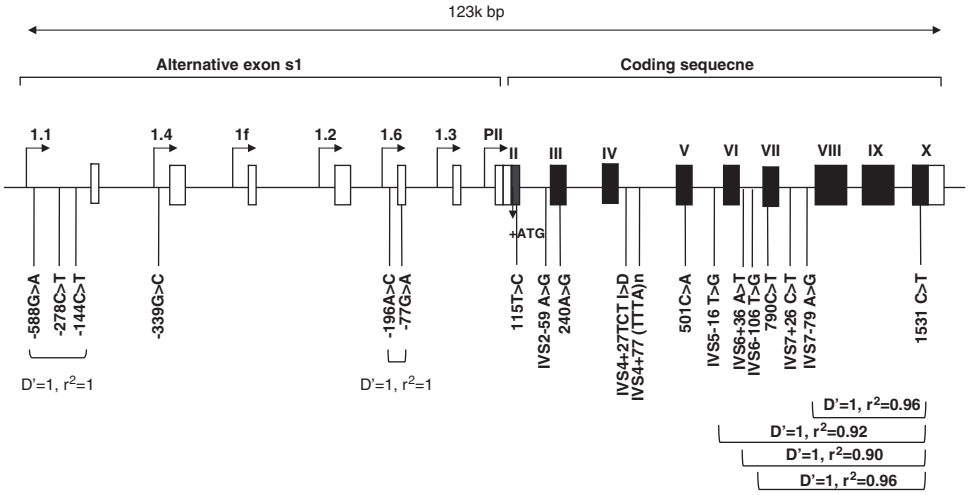


**Figure 1** Genomic organization of *CYP19A1* and the location of SNPs. Nine coding exons are indicated by black boxes. The first noncoding exons and 3′ and 5′-untranslated regions are indicated by white boxes. The values of *D*′ and $r^2$ between SNPs are described. Position numbers refer to the +1 ATG start codon of GenBank accession number NC_000015.

**Table 2 Distribution of functionally important CYP19A1 variants in different populations**

| | | *Allele frequency* | | | |
|---|---|---|---|---|---|
| *Variants* | *Subject (n)* | *Asian* | *Black* | *White* | *Reference* |
| **2* (W39R) | 200[a] | 0.10 | ND | ND | Nativelle-Serpentini *et al.*[22] |
| | 199[a] | 0.10 | ND | ND | Miyoshi *et al.*[23] |
| | 50[b] | 0.04 | ND | ND | Present study |
| | 70[c] | 0.03 | 0.00 | 0.00 | Haiman *et al.*[4] |
| | 60[d] | 0.07 | 0.00 | 0.00 | Ma *et al.*[12] |
| **3* (T201M) | 732[a] | ND | ND | 0.03 | Olson *et al.*[24] |
| | 50[b] | 0.00 | ND | ND | Present study |
| | 60[d] | 0.00 | 0.05 | 0.05 | Ma *et al.*[12] |
| **4* (R264C) | 272[a] | 0.13 | ND | ND | Tao *et al.*[21] |
| | 345[a] | 0.18 | ND | ND | Cai *et al.*[25] |
| | 112[a] | 0.12 | ND | ND | Song *et al.*[26] |
| | 50[b] | 0.16 | ND | ND | Present study |
| | 70[c] | 0.27 | 0.16 | 0.03 | Haiman *et al.*[4] |
| | 60[d] | 0.12 | 0.23 | 0.03 | Ma *et al.*[12] |
| **5* (M364T) | 50[b] | 0.00 | ND | ND | Present study |
| | 60[d] | 0.01 | 0.00 | 0.00 | Ma *et al.*[12] |
| 1531C>T | 345[a] | 0.44 | ND | ND | Cai *et al.*[25] |
| (rs10046) | 390[a] | 0.53 | ND | ND | Zhang *et al.*[20] |
| | 50[b] | 0.56 | ND | ND | Present study |
| | 633[b] | 0.42 | ND | ND | Matsumoto *et al.*[27] |
| | 70[c] | 0.40 | 0.24 | 0.56 | Haiman *et al.*[4] |
| | 60[d] | 0.54 | 0.19 | 0.56 | Ma *et al.*[12] |
| | 225[e] | 0.45 | ND | ND | Shimodaira *et al.*[8] |

Abbreviation: ND, not determined.
*CYP19A1* polymorphism studies have been carried out not only in unrelated healthy subjects but also in various phenotypic conditions. This table provides the representative of control data from the various case studies in ethnically different populations. Details for subject information are described in the corresponding reference. *CYP19A1*2* and *CYP19A1*4* exhibit higher frequencies in Asian than the others, whereas *CYP19A1*3* shows higher occurrence in White and Black populations. The variant located in the 3-UTR, a 1531C>T, occurs in the similar frequency in Asian and White; however, Black population exhibited lower frequencies compared with the other ethnic groups.
[a]Frequency obtained from control subjects where only female subjects were studied in estrogen-related association studies.
[b]Frequency obtained from unrelated healthy subjects, including male and female subjects.
[c]Frequency obtained from the female subjects without a history of cancer in Multiethnic Cohort (MEC) studies.
[d]Frequency was from the Coriell Cell Repository, which had been anonymized by the National Institute of General Medical Sciences.
[e]Frequency from both male and female control subjects in estrogen-related association studies.

which is similar to that observed in other Asian subjects.[12] The −278C>T allele may represent one of the Asian dominant alleles in racial comparisons. A 790C>T (Arg264Cys), reported at a lower frequency in Caucasians, was found at 16% frequency in our study. The 790C>T variant (Arg264Cys) has been reported to have a positive association with increased breast cancer risk in Asians.[18]

**Table 3 Allele frequencies of the *CYP19A1* $(TTTA)_n$ polymorphism among controls in different populations**

| | *Asian* | | | | | *White* | | | *Black* | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Repeat number* | *Korean* | *Korean* | *Chinese* | *Japanese* | *Japanese* | *CA* | *CA* | *CA* | *AA* | *AA* |
| 7[a] | 0.66 | 0.52 | 0.53 | 0.69 | 0.53 | 0.48 | 0.50 | 0.47 | 0.83 | 0.76 |
| 8 | 0.00 | Freq[b] | 0.01 | 0.00 | 0.00 | 0.13 | 0.13 | 0.12 | 0.03 | 0.07 |
| 10 | 0.00 | Freq[b] | 0.02 | 0.002 | 0.02 | 0.01 | 0.03 | 0.02 | 0.01 | 0.01 |
| 11 | 0.30 | 0.36 | 0.35 | 0.20 | 0.42 | 0.34 | 0.32 | 0.35 | 0.13 | 0.10 |
| 12 | 0.03 | 0.10 | 0.09 | 0.10 | 0.02 | 0.003 | 0.02 | 0.05 | 0.00 | 0.05 |
| 13 | 0.01 | Freq[b] | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 |
| Total alleles | 100 | 376 | 120 | 376 | 82 | 120 | 132 | 284 | 120 | 207 |
| References | Present | 17 | 12 | 23 | 14 | 12 | 28 | 29 | 12 | 28 |

Abbreviations: AA, African American; CA, Caucasian.
This table provides the representative of control data from the various case studies in ethnically different populations. Details for subject information are described in the corresponding reference.
[a]The most frequent allele $(TTTA)_7$ includes both the $(TTTA)_7$ and $(TTTA)_7$ with additional 3-bp deletion.
[b]Combined frequency of $(TTTA)_8$, $(TTTA)_{10}$ and $(TTTA)_{13}$ is 0.0195.

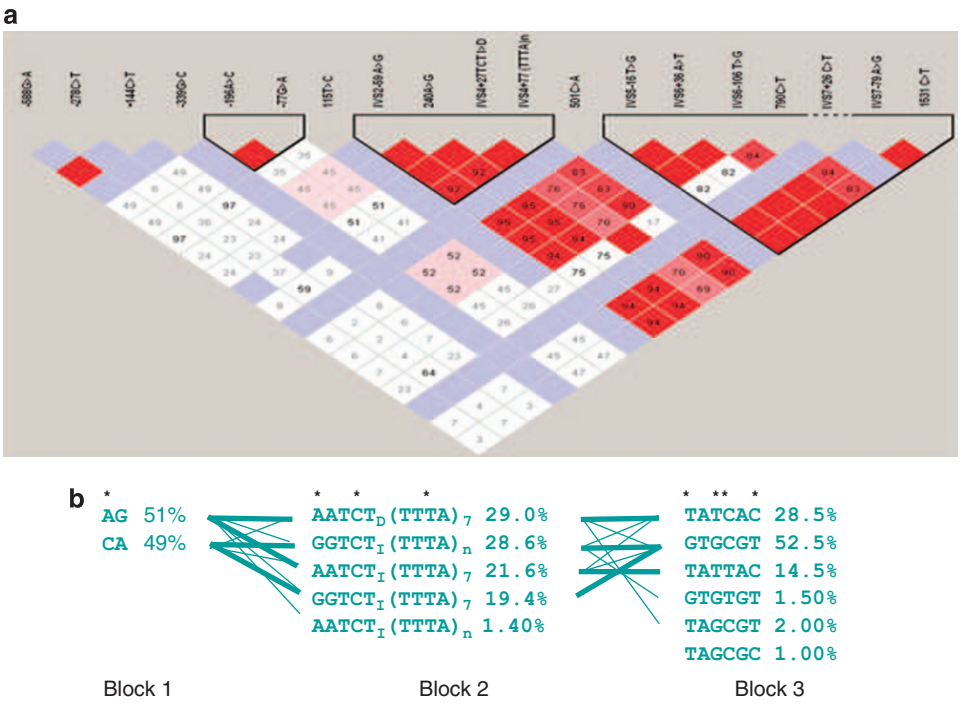


**Figure 2** Linkage disequilibrium (LD) map of *CYP19A1* single-nucleotide polymorphisms obtained from normal healthy subjects ($n$=50). The 19 *CYP19A1* variations identified were included in the LD analysis using the statistics |*D*′| and $r^2$ values.[13] **(a)** Haploview of *CYP19A1* SNPs along with their locations in the *CYP2C9* gene. Red depicts a significant linkage between the pair of SNPs. Numbers inside the square refer to the *D*′ value multiplied by 100. **(b)** *CYP19A1* SNPs and their occurrences in common haplotype structures in three blocks. Eight tag SNPs are marked by star symbols. The frequency of each haplotype is shown at the edge. Thick lines between haplotypes indicate the most common crossings between haplotypes and thinner lines signify less common crossings.

However, the results of another study did not support this hypothesis, and instead suggested that this allele is associated with a lowered risk of breast cancer.[4] The 1531C>T (rs10046) allele has been studied extensively, which includes its potential gender-specific association with essential hypertension,[8] increased level of estrogen,[19] breast cancer risk in Japanese populations[20] and endometrial cancer risk in Chinese women.[21] This variant, observed at 56% frequency in this study and located in haplotype block 3, exhibited a haplotype structure with IVS7-79A>G, IVS6-106 T>G, IVS6+36A>T and IVS5-16T>G (*D*′=1, $r^2$=0.92–0.96) in block 3 (Figures 2a and b). A set of tag SNPs, which will be useful for association mapping studies, were determined (Figure 2b). It is suggested that eight tag SNPs would be needed to track all important haplotype blocks in the *CYP19A1* gene in Koreans. Nineteen variations were used to characterize LD structures at the *CYP19A1* locus, resulting in three LD blocks (Figure 2a). The LD blocks identified were compared with the same SNPs reported in the HapMap database. Two SNPs (−77G>A and −196A>C), located in the 1.6 promoter, are in strong LD in Chinese and Japanese populations, as well as in the Korean group examined here (data not shown). These two SNPs were not linked with 1531C>T (rs10046) in Koreans (*D*′=0.48, 0.48; $r^2$=0.17, 0.17). However, a relatively strong LD was observed in Japanese

($D'$=1, 0.94; $r^2$=0.65, 0.64) and Chinese ($D'$=0.87, 0.87; $r^2$=0.5, 0.5) populations between these two loci, as well as in 1531C>T (rs10046). It may be of interest to determine whether these mutations in the 1.6 promoter can cause population or environmental differences in the expression of the aromatase protein in the given populations.

In summary, we resequenced the *CYP19A1* gene for the first time in a Korean population and identified 19 variations. Frequencies, haplotypes, LD structures and tagging SNPs were determined. Genetic information with respect to the degree of LD in Koreans might be useful for future genotype–phenotype association studies, especially in terms of breast cancer and hormone-related diseases.

## ACKNOWLEDGEMENTS

This study was supported by a Korea Science and Engineering Foundation (KOSEF) grant funded by the Ministry of Education, Science and Engineering (MOEST; no. R13-2007-023-00000-0) and by a grant from the Korea Health 21 R&D Project, Ministry for Health, Welfare and Family Affairs (A030001), Republic of Korea. This work was also supported by a grant from Inje University, 2006.

1 Bulun, S. E., Sebastian, S., Takayama, K., Suzuki, T., Sasano, H. & Shozu, M. The human CYP19 (aromatase P450) gene: update on physiologic roles and genomic organization of promoters. *J. Steroid Biochem. Mol. Biol.* **86,** 219–224 (2003).
2 Means, G. D., Kilgore, M. W., Mahendroo, M. S., Mendelson, C. R. & Simpson, E. R. Tissue-specific promoters regulate aromatase cytochrome P450 gene expression in human ovary and fetal tissues. *Mol. Endocrinol.* **5,** 2005–2013 (1991).
3 Henderson, B. E., Ross, R. K., Pike, M. C. & Casagrande, J. T. Endogenous hormones as a major factor in human cancer. *Cancer Res.* **42,** 3232–3239 (1982).
4 Haiman, C. A., Stram, D. O., Pike, M. C., Kolonel, L. N., Burtt, N. P., Altshuler, D. *et al.* A comprehensive haplotype analysis of CYP19 and breast cancer risk: the multiethnic cohort. *Hum. Mol. Genet.* **12,** 2679–2692 (2003).
5 Setiawan, V. W., Doherty, J. A., Shu, X. O., Akbari, M. R., Chen, C., De Vivo, I. *et al.* Two estrogen-related variants in CYP19A1 and endometrial cancer risk: a pooled analysis in the Epidemiology of Endometrial Cancer Consortium. *Cancer Epidemiol. Biomarkers Prev.* **18,** 242–247 (2009).
6 Modugno, F., Weissfeld, J. L., Trump, D. L., Zmuda, J. M., Shea, P., Cauley, J. A. *et al.* Allelic variants of aromatase and the androgen and estrogen receptors: toward a multigenic model of prostate cancer risk. *Clin. Cancer Res.* **7,** 3092–3096 (2001).
7 Long, J. R., Shu, X. O., Cai, Q., Wen, W., Kataoka, N., Gao, Y. T. *et al.* CYP19A1 genetic polymorphisms may be associated with obesity-related phenotypes in Chinese women. *Int. J. Obes (Lond).* **31,** 418–423 (2007).
8 Shimodaira, M., Nakayama, T., Sato, N., Saito, K., Morita, A., Sato, I. *et al.* Association study of aromatase gene (CYP19A1) in essential hypertension. *Int. J. Med. Sci.* **5,** 29–35 (2008).
9 Masi, L., Becherini, L., Gennari, L., Amedei, A., Colli, E., Falchetti, A. *et al.* Polymorphism of the aromatase gene in postmenopausal Italian women: distribution and correlation with bone mass and fracture risk. *J. Clin. Endocrinol. Metab.* **86,** 2263–2269 (2001).
10 Santen, R. J., Brodie, H., Simpson, E. R., Siiteri, P. K., Brodie, A. History of aromatase: saga of an important biological mediator and therapeutic target. *Endocr. Rev.* **30,** 343–375 (2009).
11 Lee, S. S., Jeong, H. E., Yi, J. M., Jung, H. J., Jang, J. E., Kim, E. Y. *et al.* Identification and functional assessment of BCRP polymorphisms in a Korean population. *Drug Metab. Dispos.* **35,** 623–632 (2007).
12 Ma, C. X., Adjei, A. A., Salavaggione, O. E., Coronel, J., Pelleymounter, L., Wang, L. *et al.* Human aromatase: gene resequencing and functional genomics. *Cancer Res.* **65,** 11071–11082 (2005).
13 Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B. *et al.* The structure of haplotype blocks in the human genome. *Science.* **296,** 2225–2229 (2002).
14 Miyoshi, Y. & Noguchi, S. Polymorphisms of estrogen synthesizing and metabolizing genes and breast cancer risk in Japanese women. *Biomed. Pharmacother.* **57,** 471–481 (2003).
15 Mitrunen, K. & Hirvonen, A. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. *Mutat. Res.* **544,** 9–41 (2003).
16 Kim, J. Y., Lee, C. S., Kim, H. O., Jo, Y. H., Lee, J., Jung, M. H. *et al.* The association between genetic polymorphisms in CYP19 and breast cancer risk in Korean women. *Oncol. Rep.* **22,** 487–492 (2009).
17 Hur, S. E., Lee, S., Lee, J. Y., Moon, H. S., Kim, H. L. & Chung, H. W. Polymorphisms and haplotypes of the gene encoding the estrogen-metabolizing CYP19 gene in Korean women: no association with advanced-stage endometriosis. *J. Hum. Genet.* **52,** 703–711 (2007).
18 Lee, K. M., Abel, J., Ko, Y., Harth, V., Park, W. Y., Seo, J. S. *et al.* Genetic polymorphisms of cytochrome P450 19 and 1B1, alcohol use, and breast cancer risk in Korean women. *Br. J. Cancer.* **88,** 675–678 (2003).
19 Haiman, C. A., Dossus, L., Setiawan, V. W., Stram, D. O., Dunning, A. M., Thomas, G. *et al.* Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res.* **67,** 1893–1897 (2007).
20 Zhang, L., Gu, L., Qian, B., Hao, X., Zhang, W., Wei, Q. *et al.* Association of genetic polymorphisms of ER-alpha and the estradiol-synthesizing enzyme genes CYP17 and CYP19 with breast cancer risk in Chinese women. *Breast Cancer Res. Treat.* **114,** 327–338 (2009).
21 Tao, M. H., Cai, Q., Zhang, Z. F., Xu, W. H., Kataoka, N., Wen, W. *et al.* Polymorphisms in the CYP19A1 (aromatase) gene and endometrial cancer risk in Chinese women. *Cancer Epidemiol. Biomarkers Prev.* **16,** 943–949 (2007).
22 Nativelle-Serpentini, C., Lambard, S., Seralini, G. E. & Sourdaine, P. Aromatase and breast cancer: W39R, an inactive protein. *Eur. J. Endocrinol.* **146,** 583–589 (2002).
23 Miyoshi, Y., Iwao, K., Ikeda, N., Egawa, C. & Noguchi, S. Breast cancer risk associated with polymorphism in CYP19 in Japanese women. *Int. J. Cancer.* **89,** 325–328 (2000).
24 Olson, J. E., Ingle, J. N., Ma, C. X., Pelleymounter, L. L., Schaid, D. J., Pankratz, V. S., *et al.* A comprehensive examination of CYP19 variation and risk of breast cancer using two haplotype-tagging approaches. *Breast Cancer Res. Treat.* **102,** 237–247 (2007).
25 Cai, H., Shu, X. O., Egan, K. M., Cai, Q., Long, J. R., Gao, Y. T. *et al.* Association of genetic polymorphisms in CYP19A1 and blood levels of sex hormones among postmenopausal Chinese women. *Pharmacogenet. Genomics.* **18,** 657–664 (2008).
26 Song, C. G., Hu, Z., Yuan, W. T., Di, G. H., Shen, Z. Z., Huang, W. *et al.* Effect of R264C polymorphism in CYP19A1 gene on BRCA1/2-negative hereditary breast cancer from Shanghai population of China. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* **23,** 181–183 (2006).
27 Matsumoto, Y., Suzuki, A., Shibuya, N., Sadahiro, R., Kamata, M., Goto, K. *et al.* Effect of the cytochrome P450 19 (aromatase) gene polymorphism on personality traits in healthy subjects. *Behav. Brain Res.* **205,** 234–237 (2009).
28 Probst-Hensch, N. M., Ingles, S. A., Diep, A. T., Haile, R. W., Stanczyk, F. Z., Kolonel, L. N. *et al.* Aromatase and breast cancer susceptibility. *Endocr. Relat. Cancer.* **6,** 165–173 (1999).
29 Siegelmann-Danieli, N. & Buetow, K. H. Constitutional genetic variation at the human aromatase gene (Cyp19) and breast cancer risk. *Br. J. Cancer.* **79,** 456–463 (1999).