Identification of an Aromatase Haplotype That Is Associated with Gene Expression and Postmenopausal Osteoporosis

José A. Riancho, Carmen Valero, Angel Naranjo, Dieter J. Morales, Carolina Sañudo, and Maria T. Zarrabeitia

Departments of Internal Medicine (J.A.R., C.V., C.S.) and Surgery (A.N., D.J.M.), Hospital U. M. Valdecilla, and Unit of Legal Medicine (M.T.Z.), University of Cantabria, 39008 Santander, Spain

Context: Osteoporosis has a significant genetic component. The aromatase-dependent conversion of androgenic precursors is the main source of estrogens in postmenopausal women.

Objective: The objective of the investigation was to study the relationship of a set of single nucleotide polymorphisms (SNPs) of the aromatase gene with osteoporosis and determine their functional influence on gene transcription.

Design, Participants, and Methods: This was a case-control study including 135 women with vertebral fractures due to postmenopausal osteoporosis and 312 controls. Alleles at four SNPs situated between exons 1.2 and 3 were determined by Taqman assays. Total aromatase RNA and differential allelic-specific expression were studied by RT-real time PCR in adipose tissue samples taken from 50 individuals.

Results: The SNPs studied were in strong linkage disequilibrium. A common haplotype, present in about half of the population, was identified as being associated with an increased risk of fractures (odds ratio 1.8, 95% confidence interval 1.2–2.8, \( P = 0.006 \)). There was evidence of differential allelic expression. In heterozygous individuals, transcripts bearing T alleles at rs700518 SNP (which were included in the risk haplotype) were less abundant than those with the alternative C alleles (\( P < 0.001 \)). Total aromatase expression was four times lower in fat samples from individuals who were homozygotes for the unfavorable alleles than in the opposite homozygotes (\( P = 0.007 \)).

Conclusions: A common haplotype of aromatase associated with gene expression is also associated with the risk of osteoporotic vertebral fractures in postmenopausal women. These data are in line with the hypothesis that the aromatase-dependent synthesis of estrogens plays an important role in bone homeostasis in postmenopausal women. (J Clin Endocrinol Metab 92: 660–665, 2007)

Estrogens play a critical role in bone homeostasis, and the decline in estrogen availability after the menopause is a central factor in the development of postmenopausal osteoporosis (1). However, genetic factors are known to have an important influence on bone mass and osteoporosis risk (2–4). The main source of estrogens in postmenopausal women is the aromatization of androgenic precursors, a reaction catalyzed by aromatase, the product of the CYP19A1 gene (5). In humans the gene is expressed in not only the gonads but also other tissues, including the adipose tissue and bone. The influence of aromatase-derived estrogens on the skeleton is revealed by the reduction of bone mineral density (BMD) in postmenopausal women treated with aromatase inhibitors (6).

The CYP19A1 gene is located on chromosome 15q21 and spans about 120 kb. It has a complex structure, with a 30-kb coding region and a long 5′ region, spanning about 90 kb. This 5′ region contains regulatory sequences and several variants of the first exon, whose transcription is driven by different tissue-specific promoters. These alternative forms of the first exon are transcribed and, after RNA splicing, joined onto a common splice acceptor site in exon 2, immediately upstream of the coding region (7). Thus, the mature mRNA contains a variable nontranslated first exon and a common set of exons 2–10 coding for the amino acid sequence of aromatase. Several investigators have reported an association between some common polymorphisms of the aromatase gene and BMD (8, 9). In a previous study, we reported that polymorphisms situated between exon I.2 and the coding region showed a closer association to BMD than those situated in regions further 5′ upstream (10). Therefore, in this study we genotyped other polymorphisms in that region to identify combination haplotypes associated with osteoporotic fractures and explored their functional relevance by determining their influence on gene transcription.

Subjects and Methods

Study subjects

We studied two groups of postmenopausal women aged over 50 yr. The fracture group included 135 women with postmenopausal osteoporosis and nontraumatic vertebral fractures confirmed by lateral x-rays of the spine showing a decrease in vertebral height of at least 20%. The control group included 312 control women recruited by voice and writ-
Aromatase gene expression in adipose tissue

Gene expression was analyzed by real-time RT-PCR. Subcutaneous fat samples were obtained during abdominal surgical procedures, snap frozen in liquid nitrogen, and stored at −70 °C. Total RNA was isolated by using a commercial kit (lipid minikit, QIAGEN), following the manufacturer’s instructions, after homogenizing samples (100 mg) with a ball mill (Sartorius, Goettingen, Germany). Aliquots of RNA (1–4 µg) were reverse transcribed with the Superscript III kit (Invitrogen, Paisley, UK), using random hexamers as primers. In negative control reactions, were reverse transcribed with the Superscript III kit (Invitrogen, Paisley, UK). Genotyping was performed using allele-specific Taqman probes labeled with VIC and FAM. Primers and probes were designed to check for consistency of results. DNA analysis was performed blindly and samples had to be retyped. Random samples were analyzed twice to check for consistency of results. The existence of potential binding sites within the transcript region (Taqman gene expression assays, Applied Biosystems). The cycle threshold (Ct) was determined. This represents the cycle at which a significant increase in fluorescence is first detected and is inversely related to the amount of target cDNA in the starting material. The results were normalized to a TATA box binding protein (TBP) expression analyzed in the same reaction plate to take into account possible between-sample differences in RNA amount and quality as well as variability in the efficiency of RT. Control samples of reverse-transcribed reference human RNA (Stratagene, La Jolla, CA) were also included. The relative RNA levels were calculated using the formula 2^ΔCt, where ΔCt is the difference between the control cDNA Ct and the sample cDNA Ct when the target gene (aromatase) is amplifiable, and ΔΔCt is the difference between control cDNA Ct and the sample cDNA Ct when the control gene (TBP) is amplifiable.

Differential allelic expression

Differential allelic expression was studied in a series of fat samples from individuals who were heterozygous for the rs700518 polymorphism (situated in the coding region of the gene) by a modification of the RT-coupled 5’ nuclease assay-based procedure (13, 14). RNA was extracted from fat samples as described above and residual DNA was removed by treating samples with DNase. After RT as described, the expression level of each allele was studied by real-time PCR with universal PCR master mix and the primers and allele-specific probes used for genotyping. Amplification of each allele was determined by monitoring the fluorescence of FAM and VIC-labeled probes. The Ct of each allele was determined. When the expression from one allele is dominant, amplification of the transcript from that allele will reach the threshold at an earlier cycle, thus having a smaller Ct. The difference in Ct (ΔCt) is a measure of differential expression between the alleles. In general, the expression of one allele to the other is in a ratio 1:2^ΔCt. Although cycle thresholds are directly related to the amount of specific cDNA present in the sample, the allele-specific Ct may also be influenced by other factors, such as the different fluorescence signal efficiency of FAM and VIC labels. Therefore, a standard curve was constructed with serial dilutions of two homozygous genomic DNA mixed in different proportions. The regression curve fitted an exponential function (r^2 = 0.98) that was used to calculate the expression ratio from each allele from the observed ΔCt with the formula: allele ratio = 3.592 × 10^0.237ΔCt.

The ratio of allele-specific cDNA from reverse-transcribed RNA isolated from adipose tissue samples was compared with the ratio of allele-specific genomic DNA isolated from the blood cells of heterozygous individuals (which is expected to equal 1).

Statistical analyses

Hardy-Weinberg equilibrium was tested with HWSIM software (http://kru.ch.med.yale.edu/hwsim/hwsim.doc). Linkage disequilibrium between different loci was estimated by the normalized distances and correlation coefficients, estimated with Haploview (www.broad.mit.edu/mpg/haplovie). Haploview reconstruction was performed with two different software packages, Phase v2 (15) (http://www.stat.washington.edu/stephens/software.html) and Chaplin (16) (http://www.genetics.emory.edu/labs/epstein/software/), and haplotype frequencies were computed. Very similar results were obtained with both programs. Differences in the global distribution of haplotypes between patients and controls were first assessed with a PAC-likelihood permutation test implemented in Phase. Then the association of frequent (>5%) individual haplotypes with vertebral fractures was estimated with the Wald test, and different models were compared with the Akaike information criterion (AIC), a combined index of model prediction ability and simplicity, by using Chaplin software. Finally, haplotype-associated risk of fracture was estimated by the odds ratios derived from logistic regression analysis with SPSS software (SPSS Inc., Chicago, IL). The non-parametric Jonckheere-Terpstra test for ordered categories was used to analyze the differences in aromatase gene expression between subjects with different genotypes.
Results

Haplotypes and fractures

The characteristics of the groups are shown in Table 1. Control women were slightly younger, taller, and more likely to exercise than women with fractures. Allele frequencies are shown in Table 2. There was no evidence for departure from Hardy-Weinberg equilibrium proportions.

The four SNPs studied were in strong linkage disequilibrium, with normalized distance values between 0.89 (rs700518-rs1062033 pair) and 0.97 (rs4775936-rs1062033 pair) (all P < 0.001). The corresponding correlation coefficient varied between 0.66 and 0.87. Thus, only 15 of the 81 possible haplotypes were observed, and the majority of subjects had one of two discordant haplotypes that together comprised 89% of the population. The haplotype distribution in the fracture and control groups is shown in Fig. 1. There was a statistically significant difference in the overall haplotype frequency distribution in both groups (P = 0.01). Regarding the two most common haplotypes, which had opposite sequences, haplotype 1 (TCGC) was overrepresented in fracture patients: 55.5% vs. 47.2% in controls. On the contrary, haplotype 2 (CTAG) was less frequent in fracture patients than controls: 34.5% vs. 39.2%. Screening of the association of individual haplotypes with fracture status using different genetic models revealed that only haplotype 1 was significantly associated with fractures, best fitting a recessive model (beta = 1.65, AIC = 1638, P = 0.016 in a recessive model; beta = 1.34, AIC = 1639, P = 0.041 in a multiplicative model). According to their AIC, joint models including several haplotypes did not show a better performance than single haplotype models. The presence of two copies of haplotype 1 was significantly associated with osteoporotic fractures, with an unadjusted odds ratio of 1.8 (95% confidence interval 1.2–2.8; P = 0.006; Table 3), in comparison with women with one or no copies of the haplotype. Similar results were found in multiple logistic regression models including possible confounding factors such as age and weight (odds ratio 1.8; confidence interval 1.2–2.9; P = 0.010).

Polymorphisms and aromatase gene expression

Expression of the aromatase gene was studied in 50 individuals who were taken to surgery due to abdominal problems (usually gastrointestinal cancer, cholelithiasis, or hernia). RNA levels were similar in both sexes (P = 0.60) and were not related to the neoplastic or benign nature of the underlying disease (P = 0.89). Haplotype 1, which was shown to be associated with osteoporotic fractures, included a C allele at the rs1062033 locus. Individuals bearing such C alleles showed lower levels of aromatase gene expression than those with the alternative G allele, as revealed by the smaller levels of RNA shown in Fig. 2 (P = 0.007). Aromatase RNA levels were about four times higher in samples from individuals who were homozygotes for the G allele than in homozygotes for the C allele. Similar differences were observed when subjects were classified according to rs700518 alleles (P = 0.012).

Haplotype 1 included a T allele at rs700518 locus, which is located in the common coding region of the aromatase gene. The analysis of the allele-specific expression in a group of unselected heterozygous individuals (n = 10) for this locus revealed that T alleles were underexpressed in comparison with C alleles. Thus, the average ratio of C-containing transcripts to T-containing transcripts was 1.97 ± 0.47 (ranging from 1.39 to 2.67; Fig. 3), which was significantly different from 1 and from the value of 1.01 ± 0.16 found in samples of genomic DNA (P < 0.001).

Discussion

This study shows that a common haplotype of the aromatase gene, present in about half of the population, is associated with osteoporosis. The haplotype appears to behave as recessive; thus, the group of postmenopausal women bearing two copies of the haplotype, roughly one fourth of the population, appeared to have an increased risk of osteopo-

---

**TABLE 1.** Characteristics of control and fracture women

<table>
<thead>
<tr>
<th></th>
<th>Control (n = 312)</th>
<th>Fracture (n = 135)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>67 ± 8</td>
<td>70 ± 7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>155 ± 6</td>
<td>153 ± 6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>66 ± 10</td>
<td>64 ± 10</td>
<td>0.06</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>27.4 ± 14.1</td>
<td>27.3 ± 3.8</td>
<td>ns</td>
</tr>
<tr>
<td>Age at menarche (yr)</td>
<td>14 ± 2</td>
<td>14 ± 2</td>
<td>ns</td>
</tr>
<tr>
<td>Age at menopause (yr)</td>
<td>50 ± 5</td>
<td>49 ± 5</td>
<td>ns</td>
</tr>
<tr>
<td>Duration of menopause (yr)</td>
<td>17 ± 10</td>
<td>21 ± 9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Calcium intake (mg/d)</td>
<td>688 ± 378</td>
<td>632 ± 357</td>
<td>0.15</td>
</tr>
<tr>
<td>Smokers (%)</td>
<td>12.0</td>
<td>10.7</td>
<td>ns</td>
</tr>
<tr>
<td>Alcohol (%)</td>
<td>10.1</td>
<td>3.3</td>
<td>0.04</td>
</tr>
<tr>
<td>Weight-bearing exercise (%)</td>
<td>32.6</td>
<td>18.4</td>
<td>0.01</td>
</tr>
</tbody>
</table>

ns, Not significant.

---

**TABLE 2.** Allele frequencies in fracture patients and controls

<table>
<thead>
<tr>
<th>SNP</th>
<th>Allele</th>
<th>All</th>
<th>Control</th>
<th>Fracture</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs700518</td>
<td>C</td>
<td>0.459</td>
<td>0.477</td>
<td>0.418</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.541</td>
<td>0.523</td>
<td>0.582</td>
<td></td>
</tr>
<tr>
<td>rs4775936</td>
<td>C</td>
<td>0.575</td>
<td>0.555</td>
<td>0.621</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.425</td>
<td>0.445</td>
<td>0.379</td>
<td>0.07</td>
</tr>
<tr>
<td>rs767199</td>
<td>A</td>
<td>0.466</td>
<td>0.489</td>
<td>0.415</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.534</td>
<td>0.511</td>
<td>0.585</td>
<td>0.04</td>
</tr>
<tr>
<td>rs1062033</td>
<td>C</td>
<td>0.588</td>
<td>0.572</td>
<td>0.625</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.412</td>
<td>0.428</td>
<td>0.375</td>
<td>0.13</td>
</tr>
</tbody>
</table>

---

Fig. 1. Haplotype distribution in fracture patients and controls (only haplotypes with a frequency above 1% are shown).
The critical role of estrogens in bone homeostasis is widely recognized (1), and one of its consequences is the loss of bone occurring after the menopause, when the gonadal function ceases. However, despite the negative results of some studies (18), most investigators have found a positive relationship between free estradiol and BMD in postmenopausal women as well as an increased fracture risk in women with the lowest serum levels of estradiol (19–23). Therefore, available evidence indicates that residual estrogen activity continues to play a role in bone metabolism after the menopause.

The aromatization of androgenic precursors is the main source of estrogens in men and in women after the menopause. A variety of human tissues have been shown to express the aromatase gene. Apart from the gonads, aromatase activity and/or RNA transcripts have been detected in muscle, fat, bone, nervous tissue, and several tumors (24). The adipose tissue is usually regarded as a major source of estrogens after the menopause. In fact, serum estradiol levels correlate with fat mass in postmenopausal women, and there is an inverse relationship between body weight and osteoporosis risk (25, 26). Therefore, although multiple humoral and physical mechanisms may be involved in the association between body mass and bone mass, it seems likely that the aromatase-derived production of estradiol in the adipose tissue helps to ameliorate the tendency to lose bone after the menopause. This does not exclude a role of locally produced estradiol at the bone level. Bone cells express a number of enzymes involved in steroid metabolism, including aromatase (27–29). Therefore, estradiol may also have a local effect on the bone tissue, paracrine or even autocrine, as shown in other tissues (5, 30).

There is wide evidence that aromatase activity influences bone homeostasis. Mice not expressing aromatase have a decreased bone mass (31, 32). In humans, a few patients with loss-of-function mutations of the aromatase gene resulting in profound skeletal abnormalities have been described (33). The pharmacological inhibition of aromatase is also associated with a decrease in BMD and an increased risk of fractures (6, 34).

Given the complex nature of osteoporosis, apart from cases of rare mutations, it seems unlikely for a single gene to play a quantitatively large role in determining bone mass. Indeed, in most association studies candidate genes only explain a small fraction of the total BMD variance, usually about 1–3%. However, the overall importance of genetic factors is beyond doubt. It has been suggested that heredity may explain between 40 and 80% of BMD variance and about 30% of fracture risk (2, 3, 35). Our data and those of the previously published studies mentioned above point toward aromatase as a likely candidate gene contributing to determine, in part, osteoporosis risk.

Genetic association studies may be biased by a number of factors, including population stratification. Although difficult to exclude completely, we tried to limit this potential bias by studying cases and controls from a limited geographical area. More importantly, we showed that the polymorphisms defining the osteoporosis-associated haplotype are associated with differences in aromatase gene expression. Thus, although the SNPs we studied do not induce changes in the amino acid sequence of aromatase, they are associated with functional differences at the genomic level.

Allele-specific expression differences can be identified by comparing the relative levels of exonic SNP alleles in RNA samples from heterozygous subjects. Within an individual RNA sample, alleles are exposed to the same cellular environment, and their differential expression must be due to

**TABLE 3.** Distribution of haplotype 1 in fracture patients and controls

<table>
<thead>
<tr>
<th>Group</th>
<th>All</th>
<th>Homozygotes</th>
<th>Heterozygotes</th>
<th>Noncarriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fracture</td>
<td>135 (100)</td>
<td>48 (35.6)</td>
<td>57 (42.2)</td>
<td>30 (22.2)</td>
</tr>
<tr>
<td>Control</td>
<td>312 (100)</td>
<td>72 (23.1)</td>
<td>162 (51.9)</td>
<td>78 (25.0)</td>
</tr>
</tbody>
</table>

Data represent number of patients (percentage).

Fig. 2. Aromatase gene expression in adipose tissue samples, according to genotype (alleles at the rs1062033 locus). Results represent the amount of aromatase RNA, corrected for the amount of TBP RNA, in arbitrary units (mean and SE).

Fig. 3. Allele-specific expression. Aromatase gene transcripts bearing a C allele or T allele at locus rs700518 were measured by real-time RT-PCR with allele-specific probes. The ratio of C to T alleles is shown in the figure. The results obtained when the same PCR procedure was applied to CT heterozygous genomic DNA samples are also shown for comparison purposes.
regulatory polymorphisms situated within the same chromosomal region. When allelic expression differences are strongly influenced by transregulatory sequences, one exonic SNP allele will be expressed at a higher level in some heterozygous individuals, whereas the other allele will be preferentially expressed in different individuals. However, when allelic expression differences arise from a cis-regulatory polymorphism in strong linkage disequilibrium with a coding SNP, the same exonic SNP will tend to be expressed at a higher level in all heterozygous individuals (36). This phenomenon has been reported to occur in about one fifth to one third of human genes (37, 38). In the present study, we found that T alleles at the exonic locus rs700518, present in the osteoporosis-associated haplotype, are underexpressed in comparison with C alleles. This was a consistent finding, and the average allele C to allele T ratio was about 2, similar to the ratio found in studies of other genes showing differential allelic expression (13, 14, 38).

The identification of functional variations is central to the elucidation of the genetic component of complex diseases. Most functional studies of potential regulatory SNPs are based on in vitro techniques such as transient transfection experiments with allele-specific constructs. Although informative, the results of those experiments may sometimes be difficult to interpret, especially in large genes with a complex structure, due to the limited size of the cloned regions and the fact that they are studied outside their normal chromosomal environment. In this context, direct assessment of the relative abundance of allelic transcripts may be preferable to investigating allele-specific expression differences in the normal chromosomal context (38). This methodology allowed us to identify some allelic variants of the aromatase gene with functional consequences. The polymorphisms were associated with differences in gene expression, as demonstrated by the marked differences in the levels of SNP-specific alleles in samples from heterozygous individuals and the differences in the amount of total aromatase RNA between individuals with different genotypes. Interestingly enough, those polymorphisms associated with a lower level of gene expression were also associated with an increased risk of vertebral fractures.

However, despite this association we cannot establish which SNP is actually responsible for the differences in gene transcription. Searching the TRANSFAC database of transcription factor binding sites by using an adipocyte-specific profile revealed potential binding sites for transcription factors such as specificity protein-1 (Sp1) and CCAAT/enhancer-binding protein (C/EBP) in the vicinity of the loci rs1062033 and rs767199. Those factors have been shown to modulate aromatase gene expression in some tissues (39, 40). However, it is unclear to us whether those potential binding sites are actually functional.

In fact, the active polymorphism could be one of those studied here or another unidentified polymorphism in linkage disequilibrium with them. We did not use a map-based or tagging approach but a sequence-based approach to select the SNPs to study. An a posteriori search of the HapMap database suggests that CYP19A1 gene may have seven haplotype blocks, with rs1062033 sitting in block 3 and the three other SNPs in block 2. Further studies analyzing a larger number of SNPs distributed throughout the gene are needed to clearly identify which one of those polymorphisms, or others in linkage disequilibrium with them, is actually responsible for the differences in aromatase gene expression.

In conclusion, we have identified a set of SNPs in the aromatase gene that are associated with differences in gene expression and are in strong linkage disequilibrium, resulting in a common haplotype that is associated with the risk of osteoporotic fractures. These data give further support to the hypothesis that aromatase-derived estrogens play a role in bone homeostasis in postmenopausal women. Pending confirmation in other population groups, they also raise the possibility of using these gene variants as potential tools to identify women at higher risk of osteoporosis who could become a special target population for preventive strategies.

Acknowledgments

Received July 26, 2006. Accepted November 10, 2006.

Address all correspondence and requests for reprints to: José A. Riancho, Department of Internal Medicine, Hospital U. M. Valdecilla, University of Cantabria, 39008 Santander, Spain. E-mail: riancho@unican.es.

This work was supported by Fondo de Investigaciones Sanitarias (FIS 04/28).

Disclosure: The authors have no conflicts of interest.

References


JCEM is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.