Shorter CAG Repeat in the AR Gene is Associated with Atypical Hyperplasia and Breast Carcinoma

FRANCINE BLUMENTAL DE ABREU1*, LEANDRO PIROLO JR 2*, RENATA DE AZEVEDO CANEVARI3, FABIOLA ENCINAS ROSA1, FRANCISCO ALVES MORAES NETO4, JOSÉ ROBERTO FIGARO CALDEIRA5, CLAUDIA APARECIDA RAINHO1 and SILVIA REGINA ROGATTO3

1Department of Genetics, Institute of Biosciences, São Paulo State University, UNESP, Botucatu, SP; 2Department of Biology, Londrina State University, UEL, Londrina, PR; 3Department of Urology, Faculty of Medicine, São Paulo State University, UNESP, Botucatu, SP; 4Department of Pathology, Amaral Carvalho Hospital, Jau, SP; 5Department of Senology, Amaral Carvalho Hospital, Jau, SP, Brazil

Abstract. Background: Previous reports into the role of [CAG]n repeat lengths in the androgen receptor (AR) gene indicate that these may play an important part in the development and progression of breast cancer, however, knowledge regarding benign breast lesions is limited. Patients and Methods: PCR-based GeneScan analysis was used to investigate the [CAG]n repeat length at exon 1 of the AR gene in 59 benign breast lesions (27 fibroadenomas, 18 atypical hyperplasias, and 14 hyperplasias without atypia) and 54 ductal breast carcinomas. Seventy-two cancer-free women were used as a control group. In addition, [CAG]n repeats were evaluated for the presence of loss of heterozygosity (LOH) and microsatellite instability (MSI) in a subset of these samples (27 fibroadenomas, 14 hyperplasias without atypia and 22 breast carcinomas). Results: Shorter [CAG]n repeat lengths were strongly correlated with atypical hyperplasias (p=0.0209) and carcinomas (p<0.0001). LOH was found in 1/12 and 4/20 informative cases of hyperplasias without atypia and breast carcinomas, respectively. Three patients with breast carcinoma who had previously presented atypical hyperplasia showed a reduction in the [CAG]n repeat length in their carcinomas. Conclusion: Short [CAG]n repeat length (≤20) polymorphisms are strongly associated with breast carcinomas and atypical hyperplasias. Although non-significant, a subgroup of patients with breast carcinoma and genotype SS showed an association with parameters of worse outcome.

Increased incidence and awareness of breast cancer have led to increased screening for malignancy. However, the majority of biopsies reveal no malignancy. Some are normal and many are diagnosed as various types of benign breast disorders. Epidemiological studies demonstrate that the risk of developing breast cancer varies according to the histological category of benign breast disorders (1, 2).

In this context, it is not known whether proliferative breast lesions are simply markers of increased risk, or could themselves be actual precursors of malignancies (3-9). More recently, Hartmann et al. (2) studied 9,087 women and reported that atypical hyperplasia had a relative risk of 4.24, proliferative disease without atypia had a relative risk of 1.88, and non-proliferative lesions had a relative risk of 1.27. Other reports were considered to provide weaker evidence (10-13). Some of these benign lesions, including usual ductal hyperplasia and atypical ductal hyperplasia, are assumed to represent actual precursors of malignancy, even though these lesions are histologically considered to be benign (14).

Studies aimed at identifying common microsatellite alterations (length variation, allele loss, or microsatellite instability) in genes that may represent low penetrance cancer susceptibility alleles are becoming increasingly common. Although the penetrance of such genetic variants may be low, the fact that they are very common means they may account for a large proportion of cancers. In the context of hormone-sensitive diseases, such as breast lesions, genes involved in hormone interactions are very attractive candidates (15).

The androgen receptor (AR) is a transcription factor mediating the action of androgens (16). The AR gene is
composed of eight exons and is mapped to Xq11-12. The androgen receptor contains an amino-terminal region that is variable in length, a central cysteine-rich DNA-binding domain and a carboxyl-terminal ligand-binding domain (17, 18). The large exon 1 has a polymorphic [CAG]_n microsatellite and encodes the transactivation domain that mediates target gene transcriptional activation (19). The length of the [CAG]_n repeats varies among individuals and this polymorphism is believed to be related to AR transcriptional activity. Alleles of longer repeat lengths have been associated with decreased efficacy in inducing target gene transcription (19, 20).

Genetic variation in AR repeat length has also been evaluated in relation to breast cancer risk. However, these investigations have yielded somewhat confusing results. Four studies observed that women with a long [CAG]_n repeat AR allele showed increased breast cancer risk (21-24), although the evidence is not entirely consistent (25-30). In addition, some reports showed that shorter length repeats are associated with more aggressive forms of breast cancer (31, 32). The presence of one or two long alleles [CAG]_n has been associated with a slight increase in the risk of breast cancer in young women (30). In benign breast lesions, Kasami et al. (33) studied 48 cases of fibroadenomas and 24 cases of ductal carcinoma in situ (DCIS). The authors showed that the [CAG]_n repeat length in DCIS were longer than in fibroadenomas. Rosenberg et al. (34) evaluated 15 atypical hyperplastic lesions from 12 subjects and reported LOH on the AR locus in one of the cases.

Despite the high frequency of benign breast disorders, few reports have described the AR locus. In the present study, MSI and LOH were both analyzed and evaluated for an association between the AR [CAG]_n repeat length and the relative risk for breast cancer and benign breast lesions, in comparison to a control group of typical Brazilian females with multiracial ethnicity and additional risk factors.

**Patients and Methods**

*Samples and DNA extraction.* The [CAG]_n repeat length was evaluated in 113 women with breast disease, including 27 fibroadenomas, 18 hyperplasias with atypia, 14 hyperplasias without atypia and 54 ductal breast carcinomas from Amaral Carvalho Hospital, Jau (SP, Brazil). The median age was 49.5 years (range 17-91 years). Three patients with atypial hyperplasia developed breast cancer after a one year interval. The breast carcinoma patients had undergone segmental resection or mastectomy and none of the patients had received radiotherapy or chemotherapy prior to surgery. Immediately after surgery, the samples were bisected and half of the tumor was carefully processed for histological examination in order to confirm the previous diagnosis. A manual microdissection was performed on frozen tissue to ensure the presence of at least 80-90% of tumor cells. This section was immediately snap frozen in liquid nitrogen and stored at −80°C. The breast carcinomas were submitted to histopathological classification according to the WHO Classification (35), clinical staging was determined by the UICC TNM classification (36) and the malignancy of infiltrating carcinomas was scored according to the Scarff-Bloom and Richardson grading system (37). All the patients were advised of the procedures and provided written informed consent. The Brazilian Ethics Committee CONEP approved this study. Follow-up was maintained from January 1996 to December 2005, with consultations every 6 months. A control group of 72 hospital-based individuals with a median age of 41.8 years (range 26-87 years) was randomly selected from the general population without cancer (in particular, breast and uterus neoplasia, evaluated by clinical and ultrasonography tests) and negative cancer familial history.

Genomic DNA from breast tissues and corresponding blood leukocytes was prepared by standard SDS/proteinase K digestion followed by phenol and chloroform extraction and ethanol precipitation. DNA was stored at −20°C until amplification by polymerase chain reaction (PCR).

**PCR-based GeneScan analysis.** Genomic DNA was amplified with fluorescent labeled-primers, according to Bharaj et al. (38). The fluorescent PCR products were analyzed on 5% polyacrylamide-urea (Long Ranger Singel Packs, BMA, Rockland, ME, USA) denaturing gels using an ABI Prism 377 DNA automated sequencer (Applied Biosystems, Foster City, CA, USA). Digital images of fluorescent gel data were acquired using Data Collection software and analyzed using GeneScan software (both from Applied Biosystems). Each fluorescent peak was quantified by its size (in base pairs), peak height and area. The results were imported into the Genotyper software (Applied Biosystems) for further analysis.

The [CAG]_n repeat length of AR was determined by analyzing the size of a PCR product containing the polymorphic microsatellite. The number of [CAG]_n repeats was calculated by subtracting the number of nucleotides in the invariant part of the PCR fragment from the total PCR product length and dividing the resulting number by three (39). Since the Brazilian population is known to be highly miscegenated, the cut-off [CAG]_n repeat length data reported by Ribeiro et al. (40) was used. The authors genotyped 200 individuals from two cities of São Paulo State and found mean [CAG]_n repeat length of 20.65. In the present study, the control group from the same State showed a mean of 20 [CAG]_n repeats, so this number of repeats was selected to establish the short and long alleles.

To assess the LOH, an imbalanced factor was defined, using as parameters both the peak areas and heights of the two alleles in the normal and tumor samples (41). A ratio of ≤0.60 was considered to be indicative of LOH. Allelic interpretations were made based on concordant inter-assay results in three replicates.

**Statistical analysis.** The data were analyzed to determine whether short or long androgen receptor [CAG]_n alleles were associated with altered risk of breast cancer or benign breast lesions. All the alleles were divided into two groups: those with fewer than 20 repeats were defined as short (S), whereas those with 20 or more were defined as long (L). The participants were analyzed by the Fisher Exact Test using a cut-off value [CAG]_n ≤18 versus >18 for allele 1 and [CAG]_n ≤22 versus >22 for allele 2 and the samples were evaluated as a dichotomous variable of the short mean (S) and long mean (L) of the two alleles (number of [CAG]_n repeats in
allele 1 plus the number of [CAG]_n repeats in allele 2 divided by 2) to compare cases and controls. The subjects were separated into three categories: both short alleles (SS), both long alleles (LL), and one short and the other long allele (SL). To evaluate the relationship between the three categories of alleles in breast group versus control patients and to compare clinical and pathological characteristics, the Chi-square test was used. The mean [CAG]_n repeats in allele 1 was compared between categories of clinical and pathological variables with the use of analysis of variance (ANOVA). A level of significance of 5% was considered for all statistical tests.

### Results

The data of 113 patients for [CAG]_n polymorphism analysis is given in Table I. The genotypes SS for [CAG]_n repeat length distribution were statistically significant for hyperplasia with atypia and carcinoma versus controls ($p=0.0209$ and $p<0.0001$, respectively). The mean value of the two alleles with a cut-off value of 20 repeats did not differ between cases and controls (Table II).

In a second comparison, allele 1 was defined as short ($\leq 18$) and long ($>18$), and allele 2 was short ($\leq 22$) and long ($>22$). The atypical hyperplasia group showed a lower incidence of long alleles when allele 1 was analyzed. When allele 2 was considered, long alleles were present at low percentages or absent in carcinomas and atypical hyperplasias, respectively (Table II).

In the breast carcinoma group, the relationship between clinical and pathological parameters was evaluated regarding the genotypes and the allele 1 mean. There were no significant correlations between the genotypes and lymph node status, histological grade, tumor size, and immunohistochemical markers usually investigated in breast carcinomas (ER, PGR, HER-2, and Ki-67) (data not shown). Although not significant, in breast cancer patients with negative PGR status, a higher frequency of the SS genotype (66%) than the LL and SL genotypes (43% and 45%, respectively) was observed. In addition, in cases with histological grade III, a higher frequency of the SS genotype (48%) than the LL and SL genotypes (29% and 27%, respectively) was observed. In these same samples, the mean for allele 1 was lower (17 repeats) in tumors with histological grade III than in grades I and II (in both, 19 repeats).

The mean follow-up of the carcinoma patients was 46.1 months±14.2 months. During these intervals two patients relapsed, and six presented metastasis, two bone and four lung spreads. Five patients with metastasis presented the SS genotype.

DNA from matched pairs of breast lesions and normal tissues of 64 patients (27 fibroadenomas, 14 hyperplasias

### Table I. Mean [CAG]_n repeat length, AR genotypes, and polymorphism repeat range among different patients and cancer-free women.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Mean [CAG]_n repeat / SD</th>
<th>Mean allele 1</th>
<th>Mean allele 2</th>
<th>Genotypes</th>
<th>P-value</th>
<th>[CAG]_n repeat range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibroadenoma</td>
<td>20.8 / 2.7</td>
<td>19.1</td>
<td>21.4</td>
<td>SS</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Hyperplasia</td>
<td>20.5 / 2.3</td>
<td>19.2</td>
<td>21.7</td>
<td>SS</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Atypical hyperplasia</td>
<td>17.9 / 2.8</td>
<td>17.0</td>
<td>19.1</td>
<td>SS</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Breast carcinoma</td>
<td>19.2 / 2.8</td>
<td>17.8</td>
<td>20.5</td>
<td>SS</td>
<td>36</td>
<td>11</td>
</tr>
<tr>
<td>Control</td>
<td>20.2 / 3.0</td>
<td>18.4</td>
<td>22</td>
<td>SS</td>
<td>20</td>
<td>39</td>
</tr>
</tbody>
</table>

SD: Standard Deviation; SS: homozygous to short alleles; SL: heterozygous; LL: homozygous to long alleles. *$p<0.05$.

### Table II. Allelic distribution and average of [CAG]_n repeats of AR gene in benign breast lesions, breast carcinomas and cancer-free women.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Allele 1 P-value</th>
<th>Allele 2 P-value</th>
<th>Allele mean P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤18</td>
<td>&gt;18</td>
<td>≤22</td>
</tr>
<tr>
<td>Fibroadenoma</td>
<td>12</td>
<td>15</td>
<td>0.4990</td>
</tr>
<tr>
<td>Hyperplasia</td>
<td>5</td>
<td>9</td>
<td>0.2509</td>
</tr>
<tr>
<td>Atypical hyperplasia</td>
<td>15</td>
<td>3</td>
<td>0.0311*</td>
</tr>
<tr>
<td>Breast carcinoma</td>
<td>34</td>
<td>20</td>
<td>0.3648</td>
</tr>
<tr>
<td>Control</td>
<td>39</td>
<td>33</td>
<td>45</td>
</tr>
</tbody>
</table>

*Statistical significance $p<0.05$. 
without atypia and 22 breast carcinomas) were analyzed for MSI and LOH regarding AR [CAG]n polymorphism. Neither somatic in-frame contraction nor expansion was observed for the [CAG]n repeats in relation to normal DNA from the same patient. One out of 12 informative hyperplasias and four out of 20 informative breast carcinomas showed LOH on the AR locus (Figure 1A-C). Three cases lost the longer allele and two lost the shorter allele. Three patients with atypical hyperplasia subsequently developed a breast carcinoma. All of the breast tumors showed a decrease in [CAG]n repeats of one allele when compared to the hyperplasia samples from the same patient (Figure 1D-F).

Discussion

Exposure to endogenous and exogenous hormones is known to influence breast cancer risk. The signs of these hormones are manifest in the breast via hormone receptors. Among them, the androgen receptor binds specifically to androgens and mediates androgen action by activating the transcription of androgen-regulated genes. Previous reports have shown that the length of the polymorphic region of [CAG]n repeats in exon 1 is inversely related to the ability of the androgen receptor to transactivate other genes (42).

In the present study, patients with breast disorders and cancer-free women were analyzed regarding the [CAG]n repeat length of the AR gene. The comparison between cases and controls showed that there was a significant decrease in the [CAG]n repeat length in atypical hyperplasia and carcinoma groups when the alleles were evaluated together or alone. The length of [CAG]n repeats in both alleles (genotypes) or in the mean of allele 1 was not statistically correlated with tumor size, lymph node involvement and Ki-67 status in breast cancer. Our results agree with Yu et al. (31) who found no correlation between [CAG]n repeats and the clinical or pathological features studied. However, the length of [CAG]n repeats in either one or both alleles was inversely correlated with tumor histological grade. The authors detected that tumors with higher grades tended to have shorter [CAG]n repeats and suggested that longer repeats may occur more frequently in less aggressive cancers. In the current study, despite no significant association between clinical and pathological data and [CAG]n repeats, a high percentage of histological grade III and shorter allele 1 were observed in SS genotype cases. Although non significant, an association between negative PGR status and SS genotype was also detected in the present study. PGR is a predictive marker for endocrine therapy and the negative protein expression of this gene has been associated with a worse prognosis in breast cancer (43).

Interestingly, during the period of this study, six patients with breast carcinomas developed metastasis: five presented SS genotype and three of these cases showed negative PGR status. Although the SS genotype was shown to be a marker of the risk of breast cancer development, in a subgroup of the same, it could also be an indicator of worse outcome. In the literature, short [CAG]n repeats have been associated with aggressive forms of the disease, including high histological grade and positive lymph node status (31, 32). To our knowledge, this is the first study showing an association between [CAG]n repeats and metastasis. Considering that the current investigation was based on only a small number of tumors, principally those that developed metastasis, and a short term follow-up, further study will be required to evaluate this association.

In this study, six cases of breast cancer showed first degree family history of breast carcinoma. Two of them were heterozygous for [CAG]n genotype and four showed SS genotypes. Contrarily, Haiman et al. reported that longer AR repeat alleles may be involved in modifying family history-associated breast cancer risk (22). Loss of heterozygosity was detected in one case of typical hyperplasia (8%), four breast carcinomas (20%), and was absent in fibroadenomas. Rosenberg et al. (34) evaluated 15 atypical hyperplastic lesions from 12 patients and observed LOH on the AR locus in one case. In another study, Franco et al. (14) studied 13 microsatellite markers, including the AR gene, in 32 fibroadenomas coexisting with breast cancer in the same breast and 26 other fibroadenomas, and found no LOH at AR locus.

Microsatellite instability was observed as a somatic event in three patients with atypical hyperplasia that subsequently presented as breast cancer. The breast carcinoma samples showed decreased length in one allele when compared to hyperplasia samples from the same patient. These data suggest that LOH and MSI at the AR locus are not a common event in fibroadenomas, but may play a role in the progression of atypical hyperplasias to more aggressive phenotypes.

The current study shows a strong association between the presence of short alleles [CAG]n repeats in atypical hyperplasias and breast carcinomas compared to cancer-free individuals and confirmed a low frequency of MSI and LOH in the AR gene in benign breast diseases.

Acknowledgements

The authors would like to thank the Conselho Nacional de Pesquisa (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - Brazil for their support of this study.

References

Figure 1. AR gene polymorphism analysis using the PCR-based GeneScan in which allele loss (LOH) or microsatellite instability (MSI) was detected. The size (in base pairs) of the amplified fragment, peak height, and peak area are shown below each peak, respectively. A-C: Allelotypes of normal and tumor samples (n and t, respectively) of breast carcinoma (A and B) and hyperplasia without atypia (C). The arrows indicate the LOH. D-F: MSI was detected in breast carcinoma (c) compared to atypical hyperplasia (h) samples from the same patient (D, E and F, respectively).


Received October 19, 2006
Revised February 6, 2007
Accepted February 7, 2007