Androgen receptor (CAG)n polymorphisms and breast cancer risk in a Han Chinese population

J. Dang[1,2], L. Peng[1,2], H.J. Zhong[1,2] and Z.H. Huo[1,2]

[1]Key Laboratory of Fertility Preservation and Maintenance,
Ningxia Medical University, Ministry of Education, Yinchuan, China
[2]Department of Medical Genetics and Cell Biology, Ningxia Medical University,
Yinchuan, China

Corresponding author: Z.H. Huo
E-mail: huozhh@163.com

ABSTRACT. The androgen receptor (AR) is involved in the differentiation and growth of breast cancer. Genetic markers in the *AR* gene have a plausible role in modulating the risk of breast cancer. In this study, we studied the association of breast cancer and the trinucleotide repeat polymorphism (CAG)n in exon 1 of the *AR* gene in 202 patients with breast cancer and 183 healthy controls from our hospital (Yinchuan, China). Repeat lengths were determined by fluorescent DNA fragment analysis using the ABI GeneScan software and DNA sequencing. We detected 17 short tandem repeat alleles in exon 1 in the Han population of Ningxia Province, China. The CAG repeat number ranged from 14 to 31 and the frequency ranged from 0.339 to 24.460%. Generally, (CAG)n repeat lengths <22 were classified as short (S), and those >22 were classified as long (L). No association was found between breast cancer and the S/L (CAG) variants. However, the frequency of the (CAG)25 repeats in the breast cancer group was significantly higher than that

in the control group (P = 0.033, odds ratio = 1.790, 95% confidence interval = 1.044-3.069). These findings indicate a role for *AR* gene (CAG)n variations in breast cancer and might be informative for future genetic or biological studies on breast cancer, although these findings need replication in other populations.

Key words: Androgen receptor; Breast cancer; Short tandem repeat

INTRODUCTION

Breast cancer (BC) is the most frequently diagnosed cancer worldwide (Jemal et al., 2011). In China, BC has surpassed cervical cancer and become the leading cause of cancer-related deaths among women (Jemal et al., 2011). The development of BC is a multifactorial and complex process influenced by multiple genetic variants and environmental factors (Nathanson et al., 2001). It is estimated that approximately one third of the total risk of BC might be attributed to heritable factors such as a polymorphism in the androgen receptor (AR) (Lichtenstein et al., 2000; Haiman et al., 2002).

AR, a member of the nuclear receptor superfamily, is known to be involved in a complex network of signaling pathways that collectively regulate cell proliferation (Liao and Dickson, 2002; Yeh et al., 2003). Recently, AR has been used as a useful marker in the classification of BC subtype (Lehmann et al., 2011; Guedj et al., 2012). In BC, AR is more widely expressed than are estrogen receptor alpha (ER) and progesterone receptor (PR). In the Nurses' Health Study, which involved 2171 women with invasive breast cancer, 77% tissue samples showed positive staining for AR by immunohistochemistry (Lichtenstein et al., 2000). The significance of AR in human BC is further emphasized by a recent finding, which reported that AR could be targeted therapy in estrogen receptor negative breast tumors (Ni et al., 2011). It has been reported that loss of AR expression is associated with early onset, high nuclear grade, and negative expression of ER, PR, and human epidermal growth factor receptor 2 (HER2) in breast tumors (Agoff et al., 2003; Gonzalez-Angulo et al., 2009). However, the mechanisms underlying the contribution of loss of AR expression to breast carcinogenesis remain unclear.

AR is also a family member of the steroid hormone receptors, and functions as a classic ligand-activated intracellular transcription factor as well. The AR protein possesses four distinct functional domains, including a regulatory domain at the amino terminus (AF-1 site), a DNA-binding domain composed of two zinc fingers, a hinge region with a nuclear localization signal, and a C-terminal ligand-binding domain (AF-2 site) (Friedlander and Ryan, 2012). The DNA-binding domain of the AR binds to androgen response elements in both the promoter and enhancer regions of target genes and creates an active transcription complex upon recruitment of co-regulatory proteins (Friedlander and Ryan, 2012).

The *AR* gene is located on chromosome Xq11-12. The *AR* gene promoter region contains a polymorphic CAG trinucleotide repeat that encodes an uninterrupted polyglutamine tract in the N-terminal domain, ranging from 6 to 39 repeats in healthy individuals (Buchanan et al., 2004). Shorter CAG repeat length is reported to be associated with increased risk of prostate cancer (Nelson and Witte, 2002), ovarian cancer (Santarosa et al., 2002), BC (Haiman et al., 2002; Suter et al., 2003), and endometrial cancer (Sasaki et al., 2003; Rodriguez et al., 2006). Further studies should be performed to replicate these findings and to evaluate the effects of this *AR* variation on the risk of BC.

We hypothesized that CAG repeats in *AR* might contribute significantly to the predisposition to develop BC. In this study, we investigated this variant in a Chinese population to verify the putative association between AR polymorphism and BC.

MATERIAL AND METHODS

Subjects

This case-control study enrolled 202 unrelated female patients with BC (52.36 ± 4.67 years, means ± standard deviation). The patients were hospitalized in the Department of General Surgery, the First Hospital Affiliated to the Ningxia Medical University (Ningxia, China), between May 2009 and July 2012. The diagnosis of BC in patients was confirmed by histological examination of biopsies or resected tissues. Patients with cervical cancer, ovarian cancer, or other malignancies were excluded. One hundred and eighty-three women (mean age of 54.62 ± 6.55 years) were recruited as healthy controls in the Medical Examination Center of our hospital. Exclusion criteria consisted of taking other prescribed medications that could affect the central nervous system; history of seizures, hematological diseases, or severe liver or kidney damage; smoking; hypertension; or previous use of oral contraceptives. All participants were from a non-genetically related Han Chinese population in the Ningxia Hui Autonomous Region (China). The study was performed according to the Guidelines of the Medical Ethics Committee of our hospital (Yinchuan, China). Written informed consent was obtained from all participants.

Determination of *AR* CAG repeat polymorphism length

Peripheral blood samples were collected from the enrolled subjects in tubes coated with EDTA. Genomic DNA was extracted from blood leukocytes using the EZNA™ Blood DNA Midi Kit (Omega Bio-Tek, Norcross, GA, USA) according to the manufacturer protocol. The sequences of the primers for amplification of the (CAG)n are as follows: a 6-carboxyfluorescein (FAM)-labeled upstream primer: 5'-GTG CGC GAA GTG ATC CAG A-3', and anti-sense: 5'-GTT CCT CAT CCA GGA CCA GGT A-3' (Figure 1). The reaction volume for polymerase chain reaction (PCR) amplification was 12 μL and consisted of the following reagents: 30-100 ng genomic DNA, 2X PCR buffer, 15 mM MgCL$_2$, 5 pM primer, and 0.5 U Golden Taq DNA Polymerase (Applied Biosystems (ABI), Foster City, CA, USA). PCR amplification cycle parameters were set as follows: predenaturation at 95°C for 5 min, denaturation at 95°C for 45 s, annealing at 60.5°C for 1 min, and extension at 72°C for 45 s, for a total of 32 cycles, with a final extension at 72°C for 15 min, followed by a hold at 4°C. All the PCR products were purified and sequenced with an ABI Big Dye Terminator (version 3.1) cycle sequencing ready reaction kit and an ABI PRISM 3100 sequencer (Applied Biosystems) in accordance with the manufacturer instructions. The PCR products were analyzed and genotypes determined using an ABI 3730 sequencer equipped with the GeneScan software (Applied Biosystems).

Statistical analysis

All data were analyzed by the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium was evaluated by chi-square tests. Differences between allele

and genotype frequencies in the patients and controls were evaluated by the Fisher exact test or the Pearson chi-square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of BC. Subject age was treated as a covariant for binary logistic regression. P values were calculated based on codominant, dominant for the rare allele, heterosis, and recessive for the rare allele models of inheritance. Models of multiple-logistic regression were used to test the independence of individual allelic effects.
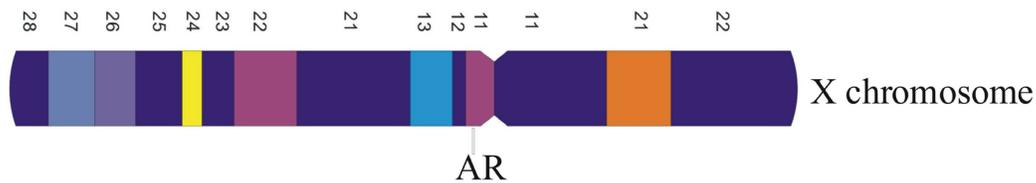
T C T A C C C T C G G C C G C C G T C C A A G A C C T A C C G A G G A G C T

T T C C A G A A T C T G T T C C A G A G C G T G C G C G A A G T G A T C C A

G A A C C C G G G C C C C A G G C A C C C A G A G G C C G C G A G C G C A

G C A C C T C C C G G C G C C A G T T T G C T G C T G C T G C A G C A G C A

G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A

G C A G C A G C A G C A G C A G C A G C A G C A A G A G A C T A G C C C C

A G G C A G C A G C A G C A G C A G C A G G G T G A G G A T G G T T C T C

C C C A A G C C C A T C G T A G A G G C C C C A C A G G C T A C C T G G T

C C T G G A T G A G G A A C A G C A A C C T T C A C A G C C G C A G T C G

G C C C T G G A G T G C C A C C C C G A G A G A

Figure 1. *AR* gene CAG repeat sequence on the X chromosome (red: primers, blue: repeat unit). AR, androgen receptor.

RESULTS

The (CAG)n repeat distribution did not deviate from Hardy-Weinberg equilibrium for controls. The mean number of (CAG)n repeats did not differ significantly between the patient versus control groups (17.08 and 16.06, P = 0.5). The repeat numbers ranged from 14 to 31 in the samples in this study (Figure 2) and the frequency ranged from 0.339 to 24.460%. The dominant allele was (GT)22. The allele distribution histogram of patients with BC was shifted to the right, with longer alleles over-represented in BC (Figure 3).

In previous studies, (CAG)n repeat lengths of <22 were classified as short (S), and those of 22 and larger were classified as long (L). No association was found between BC and the S/L (CAG) variants (Table 1). When tested for single allelic association of alleles, the Pearson chi-square test gave the following P values: 0.952 for (CAG)14, 0.512 for (CAG)15, 0.917 for (CAG)16, 0.597 for (CAG)17, 0.493 for (CAG)18, 0.498 for (CAG)19, 0.946 for (CAG)20, 0.781 for (CAG)21, 0.913 for (CAG)22, 0.858 for (CAG)23, 0.606 for (CAG)24, 0.033 for (CAG)25, 0.646 for (CAG)26, 0.554 for (CAG)27, 0.646 for (CAG)28, 0.952 for

(CAG)29, and 0.528 for (CAG)31 (Table 2). The frequency of the (CAG)25 repeats in the BC group was significantly higher than that in the control group (P = 0.033, OR = 1.790, 95%CI = 1.044-3.069).
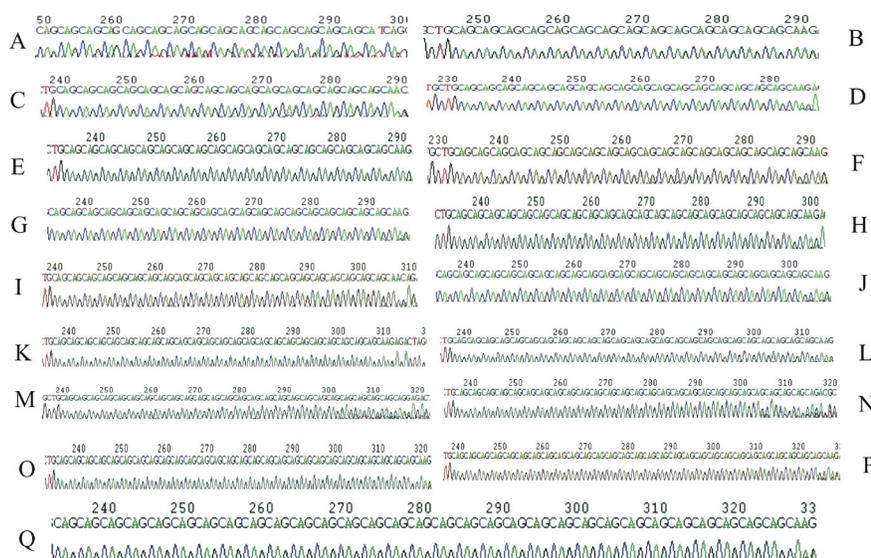

Figure 2. DNA sequencing results of PCR products showing *AR* gene (CAG)n repeat alleles [repeat number (n): A:14, B:15, C:16, D:17, E:18, F:19, G:20, H:21, I:22, J:23, K:24, L:25, M:26, N:27, O:28, P:29, Q:31]. PCR, polymerase chain reaction; AR, androgen receptor.
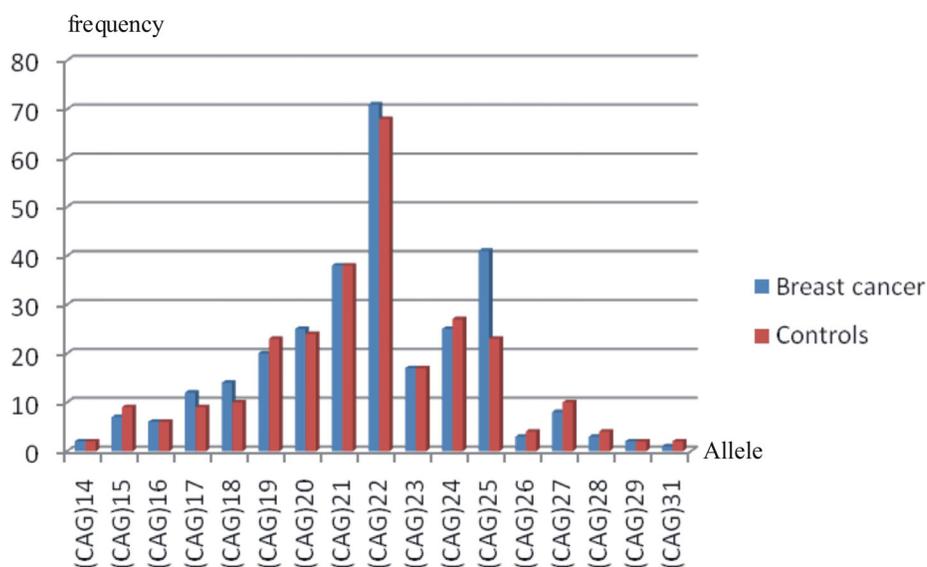

Figure 3. Allele frequency distribution of *AR* gene (CAG)n repeats in patients with breast cancer and controls. Allele size is reported as the number of CAG repeats.

Table 1. Association of *AR* gene (CAG)n repeats with breast cancer risk.

AR	Controls (N = 183)		Breast cancer (N = 202)		P value	OR (95%CI)
	N	%	N	%		
CAG length						0.781
S/S	37	20.2	37	18.3	0.636	0.885 (0.553-1.469)
S/L	76	41.5	81	40.1	0.775	0.842 (0.627-1.415)
L/L	70	38.3	84	41.6	0.505	1.149 (0.764-1.730)
CAG allele						
S	150	41.0	155	38.4	0.458	0.896 (0.671-1.197)
L	216	59.0	249	61.6		

Short (S) ≤22 repeats; long (L) ≥22 repeats. AR = androgen receptor; OR = odds ratio; CI = confidence interval.

Table 2. Statistics of the *AR* (CAG)n repeat in patients with breast cancer and controls.

(CAG)n repeat	Breast cancer		Controls		P	OR (95%CI)
	N	%	N	%		
(CAG)14	2	0.678	2	0.719	0.952	0.942 (0.132-6.733)
(CAG)15	7	2.373	9	3.237	0.512	0.717 (0.263-1.951)
(CAG)16	6	2.034	6	2.158	0.917	0.941 (0.300-2.954)
(CAG)17	12	4.068	9	3.237	0.597	1.267 (0.526-3.056)
(CAG)18	14	4.746	10	3.597	0.493	1.335 (0.583-3.058)
(CAG)19	20	6.780	23	8.273	0.498	0.806 (0.432-1.503)
(CAG)20	25	8.475	24	8.633	0.946	0.980 (0.546-1.760)
(CAG)21	38	12.881	38	13.669	0.781	0.934 (0.576-1.513)
(CAG)22	71	24.068	68	24.460	0.913	0.979 (0.668-1.434)
(CAG)23	17	5.763	17	6.115	0.858	0.939 (0.469-1.878)
(CAG)24	25	8.475	27	9.712	0.606	0.861 (0.487-1.523)
(CAG)25	41	13.898	23	8.273	0.033	1.790 (1.044-3.069)
(CAG)26	3	1.017	4	1.439	0.646	0.704 (0.156-3.173)
(CAG)27	8	2.712	10	3.597	0.554	0.747 (0.291-1.921)
(CAG)28	3	1.017	4	1.439	0.646	0.704 (0.156-3.173)
(CAG)29	2	0.678	2	0.719	0.952	0.942 (0.132-6.733)
(CAG)31	1	0.339	2	0.719	0.528	0.469 (0.042-5.206)

AR = androgen receptor; OR = odds ratio; CI = confidence interval.

DISCUSSION

In previous studies, ER and PR expression was demonstrated to be associated with good prognosis of women with BC (Samaan et al., 1981; Carey et al., 2006). Recently, the roles of androgens and AR in predicting prognosis of BC and as therapeutic targets in BC have been examined (Schippinger et al., 2006; Hu et al., 2011). The present study examined the association of the *AR* gene with human BC.

AR is located on the X chromosome (Xp11-12). In breast epithelial cells of women, only one of two *AR* alleles is expressed, whereas the other is inactivated. Thus, the cells are only affected by a single "active" *AR* allele. BC is clonal, and each cancer should have a single allele consistently activated. Shan et al. (2000) demonstrated that the active allele is lost in all BC. The *AR* gene contains a highly polymorphic trinucleotide repeat (CAG)n in the first exon that encodes a glutamine tract. It has been hypothesized that the shorter the length of this tract, the greater the affinity of androgens to the AR and the greater the resulting androgenic effects (Kazemi-Esfarjani et al., 1995; Tut et al., 1997). Although the length of this tract varies between individuals, it is typically in the range of 9-32 repeat units. Involvement of the AR in

breast tumorigenesis has been suggested by the existence of inactivating germline mutations in the AR hormone-binding domain in male patients with breast cancer (Wooster et al., 1992; Lobaccaro et al., 1993), and by splice variants that disrupt the transactivation domain in female breast tumors and tumor cell lines (Zhu et al., 1997).

Terry et al. (2005) reported that women with two long AR alleles (≥22 CAG repeats) had an increased risk of ovarian cancer compared with those with two short alleles (<22 CAG repeats). However, the results from similar analyses have varied depending on the population and sample size (Sasaki et al., 2003; Rodriguez et al., 2006). In this study, CAG repeat numbers ranged from 14 to 31 in the samples, and the frequency ranged from 0.339 to 24.460%. The dominant allele was (GT)22. No association was found between BC and the S/L (CAG) variants. This result is consistent with a previous study investigating the association between *AR* CAG repeats and BC. In a study of 368 Australian patients diagnosed before the age of 40 years and 284 controls (Spurdle et al., 1999), no association between the risk of BC and (CAG)n polymorphisms was found. When women from families with breast/ovarian cancer were examined in a case-only analysis, no association was found between (CAG)n repeat length and age at diagnosis or tumor type (Menin et al., 2001). A study of 508 patients and 426 controls from Britain showed no association (Dunning et al., 1999). Overall, we also did not observe any significant association between the CAG repeat polymorphism and BC risk.

The most intriguing finding of the present study is the association of the long *AR* alleles with BC. The frequency of the (CAG)25 repeats in the BC group was significantly higher than that in the control group. Haiman et al. (2002) assessed (CAG)n in a larger number of patients and controls, the vast majority of whom were postmenopausal, and found little evidence of an association in the aggregate. However, they did find an increased BC risk in the women with a first-degree family history who carried a (CAG)n allele of >22 repeats; they also presented results according to menopausal status. In their study, women carrying at least one longer allele [(CAG)n > 22] had a 3-fold increased risk of BC compared to those with two shorter alleles. Among African-American women with a first-degree family history of BC, a significant increase in BC risk was associated with carrying one or two *AR* long alleles. Additionally, Suter et al. (2003) found a modest increase in premenopausal BC risk associated with a repeat length of >22 units. There therefore appears to be a continuous gradient of risk associated with alleles of different sizes. Although our P was significant when a cut-point of 25 CAG units was used, this reflects the power of the study. In summary, our results further support that longer CAG repeats might increase the risk of BC, at least among some groups of women.

In conclusion, the long alleles of the CAG repeat polymorphism in AR were found. These finding encourage future efforts in searching for functional polymorphisms within and close to the *AR* gene using a systemic approach in a larger sample.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Agoff SN, Swanson PE, Linden H, Hawes SE, et al. (2003). Androgen receptor expression in estrogen receptor-negative breast cancer. Immunohistochemical, clinical, and prognostic associations. *Am. J. Clin. Pathol.* 120: 725-731.

Buchanan G, Yang M, Cheong A, Harris JM, et al. (2004). Structural and functional consequences of glutamine tract variation in the androgen receptor. *Hum. Mol. Genet.* 13: 1677-1692.

Carey LA, Perou CM, Livasy CA, Dressler LG, et al. (2006). Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 295: 2492-2502.

Dunning AM, McBride S, Gregory J, Durocher F, et al. (1999). No association between androgen or vitamin D receptor gene polymorphisms and risk of breast cancer. *Carcinogenesis* 20: 2131-2135.

Friedlander TW and Ryan CJ (2012). Targeting the androgen receptor. *Urol. Clin. North Am.* 39: 453-464.

Gonzalez-Angulo AM, Stemke-Hale K, Palla SL, Carey M, et al. (2009). Androgen receptor levels and association with *PIK3CA* mutations and prognosis in breast cancer. *Clin. Cancer Res.* 15: 2472-2478.

Guedj M, Marisa L, de Reynies A, Orsetti B, et al. (2012). A refined molecular taxonomy of breast cancer. *Oncogene* 31: 1196-1206.

Haiman CA, Brown M, Hankinson SE, Spiegelman D, et al. (2002). The androgen receptor CAG repeat polymorphism and risk of breast cancer in the Nurses' Health Study. *Cancer Res.* 62: 1045-1049.

Hu R, Dawood S, Holmes MD, Collins LC, et al. (2011). Androgen receptor expression and breast cancer survival in postmenopausal women. *Clin. Cancer Res.* 17: 1867-1874.

Jemal A, Bray F, Center MM, Ferlay J, et al. (2011). Global cancer statistics. *CA Cancer J. Clin.* 61: 69-90.

Kazemi-Esfarjani P, Trifiro MA and Pinsky L (1995). Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n-expanded neuronopathies. *Hum. Mol. Genet.* 4: 523-527.

Lehmann BD, Bauer JA, Chen X, Sanders ME, et al. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* 121: 2750-2767.

Liao DJ and Dickson RB (2002). Roles of androgens in the development, growth, and carcinogenesis of the mammary gland. *J. Steroid Biochem. Mol. Biol.* 80: 175-189.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, et al. (2000). Environmental and heritable factors in the causation of cancer - analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* 343: 78-85.

Lobaccaro JM, Lumbroso S, Belon C, Galtier-Dereure F, et al. (1993). Male breast cancer and the androgen receptor gene. *Nat. Genet.* 5: 109-110.

Menin C, Banna GL, De Salvo G, Lazzarotto V, et al. (2001). Lack of association between androgen receptor CAG polymorphism and familial breast/ovarian cancer. *Cancer Lett.* 168: 31-36.

Nathanson KL, Wooster R and Weber BL (2001). Breast cancer genetics: what we know and what we need. *Nat. Med.* 7: 552-556.

Nelson KA and Witte JS (2002). Androgen receptor CAG repeats and prostate cancer. *Am. J. Epidemiol.* 155: 883-890.

Ni M, Chen Y, Lim E, Wimberly H, et al. (2011). Targeting androgen receptor in estrogen receptor-negative breast cancer. *Cancer Cell* 20: 119-131.

Rodriguez G, Bilbao C, Ramirez R, Falcón O, et al. (2006). Alleles with short CAG and GGN repeats in the androgen receptor gene are associated with benign endometrial cancer. *Int. J. Cancer* 118: 1420-1425.

Samaan NA, Buzdar AU, Aldinger KA, Schultz PN, et al. (1981). Estrogen receptor: a prognostic factor in breast cancer. *Cancer* 47: 554-560.

Santarosa M, Bidoli E, Gallo A, Steffan A, et al. (2002). Polymorphic CAG repeat length within the androgen receptor gene: identification of a subgroup of patients with increased risk of ovarian cancer. *Oncol. Rep.* 9: 639-644.

Sasaki M, Sakuragi N and Dahiya R (2003). The CAG repeats in exon 1 of the androgen receptor gene are significantly longer in endometrial cancer patients. *Biochem. Biophys. Res. Commun.* 305: 1105-1108.

Schippinger W, Regitnig P, Dandachi N, Wernecke KD, et al. (2006). Evaluation of the prognostic significance of androgen receptor expression in metastatic breast cancer. *Virchows Arch.* 449: 24-30.

Shan L, Yang Q, Nakamura M, Nakamura Y, et al. (2000). Active allele loss of the androgen receptor gene contributes to loss of androgen receptor expression in female breast cancers. *Biochem. Biophys. Res. Commun.* 275: 488-492.

Spurdle AB, Dite GS, Chen X, Mayne CJ, et al. (1999). Androgen receptor exon 1 CAG repeat length and breast cancer in women before age forty years. *J. Natl. Cancer Inst.* 91: 961-966.

Suter NM, Malone KE, Daling JR, Doody DR, et al. (2003). Androgen receptor (CAG)n and (GGC)n polymorphisms and breast cancer risk in a population-based case-control study of young women. *Cancer Epidemiol. Biomarkers Prev.* 12: 127-135.

Terry KL, De Vivo I, Titus-Ernstoff L, Shih MC, et al. (2005). Androgen receptor cytosine, adenine, guanine repeats, and haplotypes in relation to ovarian cancer risk. *Cancer Res.* 65: 5974-5981.

Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L, et al. (1997). Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. *J. Clin. Endocrinol. Metab.* 82: 3777-3782.

Wooster R, Mangion J, Eeles R, Smith S, et al. (1992). A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenstein syndrome. *Nat. Genet.* 2: 132-134.

Yeh S, Hu YC, Wang PH, Xie C, et al. (2003). Abnormal mammary gland development and growth retardation in female mice and MCF7 breast cancer cells lacking androgen receptor. *J. Exp. Med.* 198: 1899-1908.

Zhu X, Daffada AA, Chan CM and Dowsett M (1997). Identification of an exon 3 deletion splice variant androgen receptor mRNA in human breast cancer. *Int. J. Cancer* 72: 574-580.