Androgen Receptor (CAG)$_n$ and (GGC)$_n$ Polymorphisms and Breast Cancer Risk in a Population-Based Case-Control Study of Young Women[1]

Nicola M. Suter, Kathleen E. Malone,[2] Janet R. Daling, David R. Doody, and Elaine A. Ostrander

Divisions of Clinical Research and Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024 [N. M. S., E. A. O.]; Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024 [K. E. M., J. R. D., D. R. D.]; and School of Public, Health and Community Medicine, Department of Epidemiology, University of Washington, Seattle, Washington 98195 [K. E. M., J. R. D.]

Abstract

The androgen receptor (AR) is involved in the regulation of hormone-responsive genes and, as such, variation within the gene is hypothesized to play a role in breast cancer susceptibility. We therefore assessed the relationship between *AR* repeat variation and breast cancer in young women from the general population. Women diagnosed with breast cancer before age 45 years and age-matched controls, all participants in a population-based case-control study of breast cancer, were assessed for length variation in the (CAG)$_n$ and (GGC)$_n$ *AR* repeats within the *AR* gene. Results were generated from 524 cases and 461 controls. As per previous studies, (CAG)$_n$ repeat lengths of <22 were classified as short (S), and those of ≥22 were classified as long (L). For (GGC)$_n$ repeats, those < 17 were classified as short, and those ≥ 17 were classified as long. Women with a cumulative (CAG)$_n$ repeat size of ≥43 showed a modest increase in risk for breast cancer [odds ratio (OR), 1.3; 95% confidence interval (CI), 1.0–1.7]. Women with a (GGC)$_n$ long (L) allele and those with a ≥33 cumulative repeat size had a decreased risk of breast cancer (OR, 0.7; 95% CI, 0.5–0.9). Among women homozygous for the (CAG)$_n$ short (S) allele and those with any (GGC)$_n$ L allele, an increased risk of breast cancer in relation to ever use of oral contraceptives [OCs; OR = 1.9 (95% CI, 1.0–3.6) and OR = 1.7 (95% CI, 0.9–3.5), respectively] was observed. An increased risk for OC use, however, was not observed among women with the CAG L or GGC S allele. This study, one of the first to examine both (CAG)$_n$ and (GGC)$_n$ in a population-based study for its relation to breast cancer risk, suggests a reduced risk in young women with (GGC)$_n$ repeat lengths of ≥17. In addition, these data suggest that *AR* repeat length may be partly responsible for the increased risk for early-onset breast cancer in women who use OCs, although these findings need replication in other populations.

Introduction

Within the general population, breast cancer risk is modified by a large number of environmental and genetic factors. Such factors may affect susceptibility and/or progression of the disease and are likely to be important in women with either sporadic or inherited forms of breast cancer. Because hormonal factors play a role in the development and maturation of the breast, variation within genes involved in hormone production and regulation is hypothesized to be particularly important.

The androgen receptor protein is a member of the nuclear receptor subfamily of steroid receptors that functions as a transcription factor to regulate the transactivation of hormone-responsive genes and is thus of specific interest. The *AR*[3] gene contains two trinucleotide repeats within exon 1; a polyglutamine tract encoded by a (CAG)$_n$ repeat, and a polyglycine tract encoded by a $(GGT)_3GGG(GGT)_2(GGC)_n$ sequence, known as the (GGN)$_n$ repeat (1, 2). Most of the domain, inclusive of the repeats, is necessary for full wild-type activity of the gene (3), and both repeats have been considered as risk factors for cancer (4).

Initial cancer-related studies of *AR* repeat variation focused exclusively on association with prostate cancer risk. Studies suggest a 2–3-fold increase in risk, particularly of aggressive disease, in men with short *versus* long (CAG)$_n$ repeat alleles (5–7). Each additional (CAG)$_n$ repeat has been shown to correlate with a 3–14% decrease in prostate cancer risk (7, 8). The length of the polyglutamine chain has specifically been shown to correlate inversely with transcriptional activity (9–11). In addition, men with extreme (CAG)$_n$ repeat expansion, associated with spinal and bulbar muscular atrophy or Kennedy's disease, experience androgen insensitivity, presumably due to loss of transactivation (12, 13). Subsequently, several studies have linked (GGC)$_n$ repeat length with prostate cancer risk (7, 14–17).

Recently, the role of *AR* repeat length and breast cancer risk has been evaluated in several sets of women. A particular focus has been the hypothesized association between *AR* repeat and mutations in the *BRCA1/2* genes. In a study of 165 cases and 139 controls, all of whom carried confirmed germ-line

[3] The abbreviations used are: AR, androgen receptor; OR, odds ratio; CI, confidence interval; OC, oral contraceptive.

mutations in the *BRCA1* gene, Rebbeck *et al.* (18) reported an increased risk of breast cancer among women with at least one long (repeat length ≥28) (CAG)$_n$ allele, compared with women with two short alleles (Hazard Ratio, 3.7; 95% CI, 1.6–8.4). Rebbeck *et al.* (18) also noted a correlation between earlier age at diagnosis and longer repeat length. However, a study of *BRCA1* and *BRCA2* mutation carriers of Ashkenazi Jewish origin, a subset of whom had breast cancer, demonstrated no association with breast cancer risk (19). When women from breast/ovarian cancer families were examined in a case-only analysis, Menin *et al.* (20) found no association with (CAG)$_n$ repeat length and age at diagnosis or tumor type.

When women from the general population are considered, the results are equally inconclusive. A study by Giguère *et al.* (21) of 255 incident cases and 461 controls from Quebec City reported that women with short CAG alleles (≤39 repeats totaled from both alleles) have a 50% reduction in risk compared with women for whom the sum of the repeats is ≥40 (OR, 0.5; 95% CI, 0.3–0.83). The association was largely confined to postmenopausal women and was not affected by family history (21). Haiman *et al.* (22) assessed (CAG)$_n$ in a larger number of cases 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 breast cancer risk among women with a first-degree family history who carried a (CAG)$_n$ repeat allele of ≥22, and they also presented results according to menopausal status. A study of 508 cases and 426 controls from Britain, however, showed no association (23). The study of Dunning *et al.* (23) is difficult to compare with those of Giguère *et al.* (21) and Haiman *et al.* (22). It included two cohorts of women; the first was a series of prospectively ascertained patients from a single hospital who were <71 years of age at diagnosis. A set of matched anonymous controls was also included with this set. The second series was a retrospectively ascertained group of patients identified through the East Anglian Cancer Registry as part of the Anglian Breast Cancer Study and comprised patients < 55 years of age together with a set of random controls. The study did not assess postmenopausal women separately, and family history information was not included, nor was *BRCA1/2* status reported.

To date, only one population-based study has examined the role of *AR* polymorphisms in young women, irrespective of *BRCA1* status. Spurdle *et al.* (24) reported no association between the risk of breast cancer and (CAG)$_n$ polymorphisms in a study of 368 Australian cases diagnosed before the age of 40 years and 284 controls. The study, however, did not consider the role of the (GGC)$_n$ repeat. The two studies to date that have addressed the role of (GGC)$_n$ repeat length and breast cancer risk, although largely negative, focused exclusively on women across a wide spectrum of age and/or women who are likely *BRCA1/2* carriers (19, 23). In the present study, therefore, we consider the role of polymorphisms with the (CAG)$_n$ and (GGC)$_n$ repeats in a population-based, case-control study of 524 cases diagnosed before age 45 years and 461 age-matched controls.

Materials and Methods

Study Population. A characterization of the study population has been reported previously and is summarized only briefly (25). Cases were identified through the Cancer Surveillance System of Western Washington, a population-based cancer registry and a participant in the National Cancer Institute Surveillance, Epidemiology, and End Results Program (SEER). The study identified all incident first primary breast cancer cases diagnosed before age 45 years from May 1, 1990 to December 31, 1992 in women of all races and ethnic backgrounds who were residents of King, Pierce, and Snohomish counties at the time of diagnosis. Controls were identified through random digit dialing conducted in the same three-county metropolitan area and were frequency-matched to cases on 5-year age group and reference year (26). Information on potential risk factors for breast cancer, including family history, was obtained through a structured in-person interview. Reference date for the interview, a date beyond which exposure information was not considered, was the month and year of diagnosis for cases and a randomly assigned date for controls.

Interviews were completed on 642 cases (84.0%) and 608 controls (90.5% telephone screener rate × 81.5% interview response rate = 73.8% overall response rate). Blood collection occurred in two phases. As part of the initial case-control study, funding was provided to attempt early blood draw (preceding adjuvant therapy) on at least 50% of the cases (and a smaller proportion of controls) for the purpose of investigating endogenous hormone levels. Blood collection from the remaining women was completed as part of a subsequent genetic-epidemiology study. Altogether, blood was collected from 540 cases (84.1%) and 476 controls (78.3%). Of these, 337 (62.4%) case samples and 176 (37.0%) control samples were obtained at the time of the original study, and 203 (37.6%) case samples and 300 (63.0%) control samples were obtained as part of the subsequent study (which was begun approximately 2 years later). DNA was unavailable for this study for 16 cases and 14 controls due to low DNA yields.

Analysis for both (CAG)$_n$ and (GGC)$_n$ repeats was performed on samples from 524 cases (81.6%) and 462 controls (76.0%). Complete results were generated for 524 cases and 461 controls for the (CAG)$_n$ repeat and 515 cases and 443 controls for the (GGC)$_n$ repeat. There was one control sample for which (GGC)$_n$ repeat results were not ascertained due to PCR failure, one control sample in which only a single allele was determined for the (CAG)$_n$ repeat, and 9 case and 18 control samples for which only one allele of the (GGC)$_n$ repeat could be determined. For these subjects, variables designating CAG or GGC repeat genotype were coded as unknown when totally unclassifiable but coded into specific categories when possible [*i.e.*, a subject with one undetermined (GGC)$_n$ allele and one long (GGC)$_n$ would be classified as unknown for the three-level variable that stratifies genotypes into short-short, short-long, and long-long but could be classified as having at least one long allele in the two-level variable that compares those with the short-short genotype *versus* those with the long-long/short-long genotype]. Statistical analyses were performed both including and excluding these women.

Molecular Genetics. DNA samples were extracted from frozen buffy coats using standard protocols and sent blinded to the laboratory for genotyping, with each plate containing a mixture of cases and controls, unknown to the laboratory technician. The (CAG)$_n$ and (GGC)$_n$ repeats of the *AR* were amplified using nested primers as described previously (4). To assess (CAG)$_n$ repeat size, 25 ng of germ-line DNA was amplified in a 10-μl volume for 30 cycles using the following conditions: initial denaturation at 96°C for 60 s; followed by 30 s each at 94°C, 60°C, and 72°C. The product was diluted 1:100, 5 pmol of IR$_{770}$dATP were added (Boehringer Mannheim), and a second round of PCR amplification was carried out using an annealing temperature of 62°C for 30 cycles. The (GGN)$_n$ repeat was amplified as described previously (4).

Products were resolved on denaturing polyacrylamide gels

using a Li-Cor Model 4200 automated infrared DNA sequencer (Li-Cor, Lincoln, NE). To assess PCR product size, every fourth lane (17 lanes in total on each gel) was loaded with a size standard encompassing 15 PCR products ranging in size from 50 to 425 bp. Genotypes were determined with commercially available SAGA software (Li-Cor). Sizing is accurate to within 0.2 bp. In addition, for both CAG and GGC, individuals with previously determined, known-sized alleles were included once on every gel. For CAG, this was an individual with 26- and 15-bp alleles. For GGC, this was an individual with 16- and 12-bp alleles. Because of difficulties in reproducibly amplifying both alleles for the GGC samples, all putatively homozygote (GGC) samples were run twice to verify allele status. Samples were scored as homozygotes only if a single dark band was visible on both gels. We did occasionally observe a fainter band in a sample lane that, based on size, could have been a poorly amplifying secondary allele. In such cases, to be conservative, only the one allele was scored, and the second allele was scored as unknown. This scenario involved only a small fraction of samples (9 cases and 18 controls).

Five μl of a 1:50 dilution of the initial PCR product were used as the template for the second round of amplification, following the addition of 2.5 pmol of $IR_{700}dATP$. Products were resolved on denaturing polyacrylamide gels using a Li-Cor Model 4200 automated infrared DNA sequencer (Li-Cor). Genotypes were determined with commercially available SAGA software (Li-Cor). As described previously, $(CAG)_n$ repeat alleles of <22 were classified as short, and those 22 and larger were classified as long (4, 24). $(GGC)_n$ repeats of <17 units within the $(GGN)_n$ repeat were classified as short, whereas those 17 and larger were classified as long (4).

We assessed repeatability of genotyping for both cases and controls by blind duplication of a set of a set of six samples that the laboratory technician was unaware of. For $(CAG)_n$, there was perfect agreement in all genotyping calls. For $(GGC)_n$, there was perfect agreement for five of six samples. For the sixth sample, we were able to call only one of the two alleles, and there was agreement on that result in both assays. The second allele was very faint on one assay and missing on the other and hence was simply called as an X or unknown. Thus, for genotyped alleles called (representing 12 for CAG and 11 for GGC), we noted perfect agreement.

BRCA1 and BRCA2 Analysis. A subset of the samples analyzed in this study had been analyzed previously for germ-line mutations in the *BRCA1* and *BRCA2* genes (27, 28). Cases screened for mutations in *BRCA1/2* ($n = 146$) were targeted based on having been diagnosed before 35 years and/or having a first-degree family history of breast cancer. Ten cases were found to have germ-line mutations that were considered likely to be disease associated (27, 28). In addition, 240 controls with a first-degree family history or who were between the ages of 40 and 44 years at reference date were tested for *BRCA1*, and 37 controls with a first-degree family history were tested for *BRCA2*. No disease-associated mutations were found in controls.

Statistical Analysis. The *t* test was used to compare the mean lengths of alleles in cases *versus* controls and to compare the mean ages at diagnosis in cases according to genotype. The $(CAG)_n$ and $(GGC)_n$ repeats were assessed as categorical variables, based on both the mean number of repeats and previously published results (4, 24).

To assess the relationship between the categories of repeats and the risk of breast cancer (and also to assess the association between risk factors such as OC use and the risk of

breast cancer according to genotype), we used logistic regression to obtain ORs as estimates of the relative risk and 95% CIs (29). All analyses were completed using Stata statistical software.

Because reference age and year were matching variables for the frequency matching used in the original study, all risk estimates presented are age- and reference year-adjusted (both as continuous variables). Both established and suspected risk factors for breast cancer were evaluated for their potential confounding influence on the analyses of *AR* genotype in relation to breast cancer risk and on the analyses of risk factors within genotype. Potential confounders were evaluated by single variable addition to the models as well as simultaneous consideration of all potential confounders, and a change of 10% or more in an OR was set as the criterion for confounding. The following variables were evaluated as potential confounders: race; history of a live birth; age at first live birth; OC use (ever use, duration of use, recency of use); age at menarche; menopausal status; lifetime weekly average alcohol consumption; family history of breast cancer; quartile of body mass index; and history of ever having a mammogram.

Results

Study Population Characteristics. Complete data were collected for 524 cases and 461 controls for the $(CAG)_n$ polymorphism and 515 cases and 443 controls for the $(GGC)_n$ polymorphism. In the whole study, the racial distribution of the tested women was as follows: 89% Caucasian; 4% African American; 5% Asian/Pacific Islander; 1% American Indian/Aleutian; 1% Hispanic; and 0.4% other. We observed some differences between subsets of individuals tested and those for whom samples were not available. Tested cases tended to be older than untested cases ($P = 0.03$). No such differences were observed in controls. In addition, untested cases were more likely to have advanced-stage disease ($P = 0.001$). Specifically, 30.2% of tested cases *versus* 42.1% of untested cases had regional disease, and 1.4% and 7.0% had distant disease, whereas 51.4% and 40.4% had local disease. As expected, therefore, tested cases (89.1%) were more likely to be alive as of the last follow-up date (March 2000) than untested cases (55.9%; $P = 0.001$). We observe no difference in cases or controls between those tested and those not tested with regard to family history of breast cancer.

Tested cases and controls did not differ significantly with regard to reference age or race (white *versus* non-white; Table 1). However, cases were more likely to have reported a family history of breast cancer than controls, particularly a first-degree family history, which was reported by 16.8% of cases but only 6.9% of controls.

AR Repeat Characteristics. The mean number of $(CAG)_n$ repeats for cases *versus* controls (22.0 and 21.9) did not differ significantly ($P = 0.5$). Allele sizes ranged from 8 to 35 among cases and from 8 to 38 among controls, and allele distribution was approximately normal for both cases and controls. Allele distributions for both the $(CAG)_n$ and $(GGC)_n$ repeats by case-control status for short and long alleles are shown in Fig. 1. The distribution and its peaks are similar to those seen in other studies (18–20, 22). The mean number of $(GGC)_n$ repeats within the $(GGN)_n$ repeat was 15.4 in cases and 15.6 in controls, respectively ($P = 0.13$), and ranged from 3 to 18. For $(GGC)_n$, the most common allele was 16 repeats, which was observed in 70.5% of cases and 69.2% of controls. As per previous studies, $(CAG)_n$ repeat lengths of <22 were classified as short (S), and those of 22 and larger were classified as long

Table 1 Characteristics of cases and controls

Characteristic	Cases (n = 524)		Controls (n = 462)	
	N	%	N	%
Age at reference (yrs)				
<35	69	13.2	79	17.1
35+	455	86.8	383	82.9
Race				
White	465	88.7	413	89.4
Non-white	59	11.3	49	10.6
Family history				
None	321	62.1	340	75.4
First degree	87	16.8	31	6.9
Second degree	109	21.1	80	17.7
Unknown	7		11	
Age at menarche (yrs)				
8–12	277	52.9	235	51.0
13–15	234	44.7	211	45.8
16+	13	2.5	15	3.3
Ever used OCs				
No	91	17.4	102	22.1
Yes	433	82.6	360	77.9
History of a live birth				
No	144	27.5	92	19.9
Yes	380	72.5	370	80.1
No. of live births				
None	144	27.6	92	20.0
1	93	17.8	89	19.3
2	194	37.2	168	36.4
3+	91	17.4	112	24.3
Age of first live birth				
None	144	27.6	92	19.9
<20 yrs	65	12.4	88	19.0
20–29 yrs	225	43.0	224	48.5
30+ yrs	89	17.0	58	12.6
Menopausal status				
Premenopausal	461	88.3	395	85.9
Postmenopausal	61	11.7	65	14.1
Unknown	2			
Body mass index				
15.7 to <21.2	127	24.9	118	26.0
21.2 to <23.6	157	30.8	113	24.9
23.6 to <27.0	117	22.9	111	24.5
27.0–57.2	109	21.4	111	24.5
Missing	14			

(L). For $(GGC)_n$, repeats of <17 were classified as short, and those of 17 or larger were classified as long (7, 23, 24).

The $(CAG)_n$ repeat distribution did not deviate from Hardy-Weinberg equilibrium for either controls ($P = 0.09$) or cases ($P = 0.6$), using the Markov Chain algorithm of Guo and Thompson (31). The $(GGC)_n$ allele distribution was not in Hardy-Weinberg equilibrium for either controls ($P = <0.0001$) or cases ($P = 0.0001$). Exclusion of non-white women did not alter these results. Overall, we observed fewer S/L (GGC) genotypes than expected in both cases and controls. We speculate that the number of both SS and LL genotypes was greater than expected for both cases and controls, due to a slight tendency for the PCR assay to amplify only one allele in a small subset of heterozygotes.

AR Repeats and Breast Cancer Risk. In all analyses, $(CAG)_n$ and $(GGC)_n$ alleles were assessed in terms of S and L genotypes and considered as categorical variables. Compared with women with the CAG S/S genotype, there was a suggestive but modest increase in the risk of breast cancer for women with the S/L (OR, 1.3; 95% CI, 0.9–1.8) and L/L genotypes (OR, 1.2; 95%

CI, 0.8–1.7; Table 2). Similar results were observed for the risk of breast cancer in relation to having one or two L alleles [the S/L and L/L genotypes (OR, 1.3; 95% CI, 0.9–1.7)], compared with women with a S/S genotype. Women with a total $(CAG)_n$ repeat number (the sum of both alleles) of 43 or more *versus* 42 or less were at a similarly increased but not statistically significant risk (OR, 1.3; 95% CI, 1.0–1.7).

A 30% and 50% reduction in breast cancer risk was observed, respectively, among women with the S/L and L/L $(GGC)_n$ genotypes, compared with women with the S/S genotype [OR = 0.7 (95% CI, 0.5–1.0) and OR = 0.5 (95% CI, 0.3–0.9)]. Women with at least one L allele (S/L and L/L genotypes) had a 30% reduction in the risk of breast cancer compared with those with the S/S genotype (OR, 0.7; 95% CI, 0.5–0.9). Risk was similarly decreased for women with a cumulative $(GGC)_n$ allele of 33 or larger compared with those with 32 or fewer total $(GGC)_n$ repeats (OR, 0.7; 95% CI, 0.5–0.9).

The combined effects of the $(CAG)_n$ and $(GGC)_n$ repeats were also considered. We observed no obvious joint effects but note that our modest sample size would have limited our ability to detect any but the strongest of effects.

The small number of non-Caucasian women in the study (59 cases and 49 controls) precluded analysis of effect modification by race. However, assessment of *AR* genotype with risk of breast cancer among Caucasian women alone gave virtually identical results compared with those of the study overall (data not shown). The small number of women under the age of 35 years (69 cases and 79 controls) precluded analysis of age-specific effects. Whereas the associations with the $(GGC)_n$ repeat in women 35 years of age and older was similar to effects observed in the study overall, the magnitude of the associations with the $(CAG)_n$ repeat in the women age 35 years and older was slightly decreased compared with those seen in the study as a whole (data not shown). None of the potential confounders examined substantively altered the associations between *AR* genotypes and the risk of breast cancer.

Risk Factor Analysis. A selection of breast cancer risk factors known to pertain to hormonal pathways was assessed for variation in effect by genotype. Factors for which we observed no discernible variation in association with the risk of breast cancer according to $(CAG)_n$ or $(GGC)_n$ genotype included age at menarche, history of a birth, age at first live birth, number of births, and body mass index (data not shown). An evaluation of these relationships by levels of the cumulative number of $(CAG)_n$ or $(GGC)_n$ repeats was similarly unremarkable.

There was some variation in the association between OC use and breast cancer risk by genotype for both $(CAG)_n$ and $(GGC)_n$ repeats (Tables 3 and 4). Specifically, among women who were $(CAG)_n$ S/S homozygotes and among women who were either heterozygotes or homozygotes for a $(GGC)_n$ L allele, an increased risk for breast cancer in relation to ever having used OCs was observed [OR = 1.9 (95% CI, 1.0–3.6) and OR = 1.7 (95% CI 0.9–3.5), respectively]. Neither ever use nor specific features such as duration, recency, and age of first use appeared to be related to breast cancer risk in women with an L allele in $(CAG)_n$ or with the $(GGC)_n$ S/S genotype.

Among women with the $(CAG)_n$ S/S genotype, risk was elevated across all three duration of OC use categories [<5 years of OC use, OR = 1.8 (95% CI, 0.9–3.6); 5 to <10 years of OC use, OR = 1.9 (95% CI, 0.8–4.4); 10+ years of OC use, OR = 2.1 (95% CI, 0.9–5.0); trend test $P = 0.66$; Table 3], and among the women carrying at least one $(GGC)_n$ L allele, risks were increased in relation to OC durations of less than 5 years
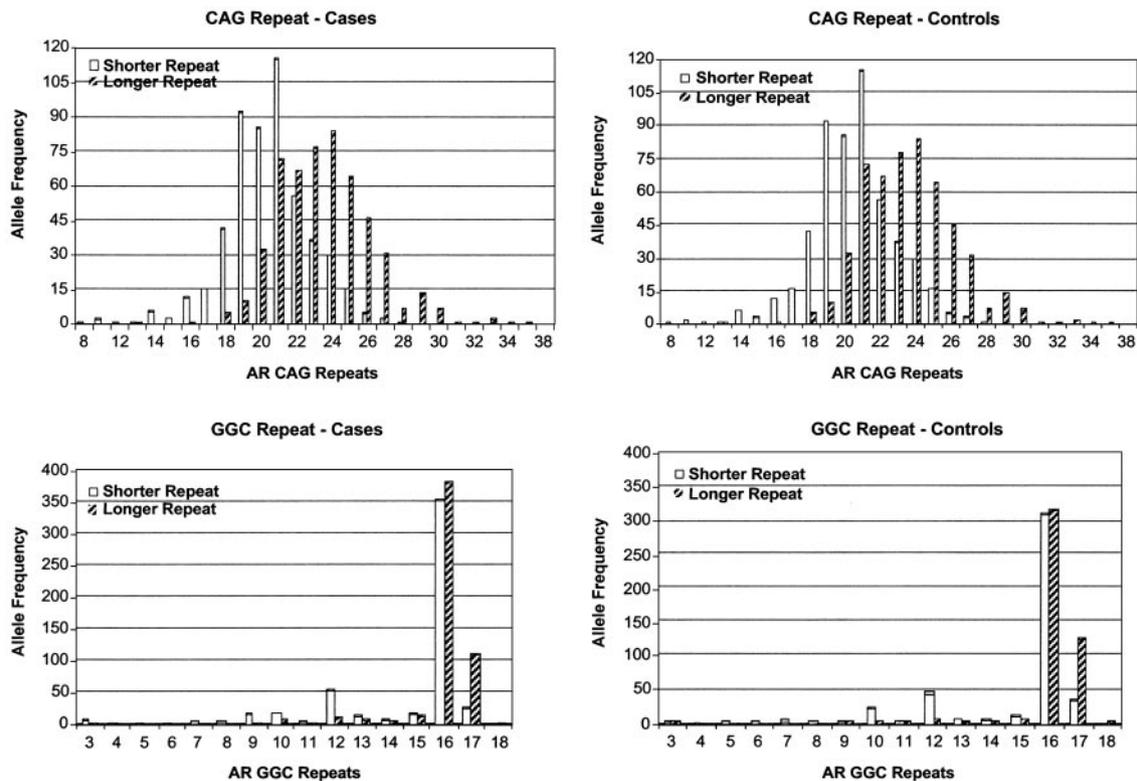
Fig. 1. CAG and GGC allele (repeat length) frequencies for breast cancer cases and controls. These figures exclude 1 control for CAG, 19 controls for GGN, and 9 cases for GGN because we did not have complete information on both allele lengths.

(OR, 1.8; 95% CI, 0.9–4.0) and 5 to <10 years (OR, 2.2; 95% CI, 1.0–5.1), but not in relation to OC use of 10 years or more (OR, 1.0; 95% CI, 0.4–2.5; Table 4). Among both groups, there were no obvious trends of increasing risk in relation to increased duration of OC use, and the magnitudes of the effects for specific duration categories, although increased, were quite similar to the magnitude of the effects observed in relation to "ever use."

Among women with the $(CAG)_n$ S/S genotype, a >2-fold increase in risk was observed among those who last used OCs within 10 years before reference date; risk decreased to a 1.6-fold excess among those whose most recent use was 10 years or more before the reference date. Among women with one or two $(GGC)_n$ L alleles, risk estimates varied by recency but not in any apparent dose-response pattern.

In women with the S/S genotype for $(CAG)_n$, estimates of the risk of breast cancer in relation to varying ages of first use ranged from 50% to 100% increases in risk. These effects were broadly similar to the increased risk that was observed for ever use of OCs within this group. Risk increased slightly from 1.5 for OC use beginning after age 21 years (95% CI, 0.6–3.7) to a 2-fold increase in risk for first use at ages 18–21 years and for use that began before age 18 years, but there was no statistically significant trend. Among women with at least one L $(GGC)_n$ repeat, breast cancer risk was increased for those whose first OC use occurred between ages 18 and 21 years (OR, 2.4; 95% CI, 1.1–5.2), but not for those whose first use began before age 18 years or for those whose first use was after age 21 years.

We assessed whether family history modified the effects of the various genotypes on the risk of breast cancer. Although risk estimates fluctuated to some extent by family history strata,

these differences were neither consistent, substantive, nor statistically significant. The results were also unaffected by *BRCA1* and *BRCA2* status. Associations did not change when women with likely disease-associated *BRCA1* and *BRCA2* germ-line mutations were (separately and together) excluded from the analysis (data not shown). We note, however, that only a subset of the study population was screened for *BRCA1* (39%) and *BRCA2* (19%) mutation status.

Discussion

Little is known about the role of the $(CAG)_n$ and $(GGN)_n$ repeats within the *AR* gene and breast cancer risk. The results reported here suggest that the presence of one or two long $(CAG)_n$ alleles or a cumulative $(CAG)_n$ repeat size exceeding 42 may be associated with a slight increase in the risk of breast cancer in young women and that the presence of one or two long $(GGC)_n$ alleles may be associated with a substantive reduction in the risk of breast cancer in young women. With regard to $(CAG)_n$, the observed risk estimates were all exceedingly modest in magnitude (20–30% increases in risk), and only the OR for the combined length of both alleles (a 30% increase in risk) was statistically significant.

The magnitude of the increased risk we observed with the $(CAG)_n$ genotype was broadly similar to the findings of a population-based study of breast cancer in Australian women under 40 years of age (24). In a study of 368 cases and 284 controls, Spurdle *et al.* (24) observed an OR of 1.4 for the relationship between breast cancer risk and carrying a long $(CAG)_n$ allele (defined as ≥22 repeats), but the CI included 1.0, indicating that the result was not statistically significant (OR,

Table 2   Association of CAG and GGC with breast cancer risk

AR	Cases (n = 524)		Controls (n = 462)		OR[a]	95% CI
	N	%	N	%		
CAG length[b]						
S/S	121	23.1	122	26.5	1.0	Reference
S/L	255	48.7	206	44.7	1.3	0.9–1.8
L/L	148	28.2	133	28.9	1.2	0.8–1.7
CAG length[b]						
S/S	121	23.1	122	26.5	1.0	Reference
Any L	403	76.9	339	73.5	1.3	0.9–1.7
CAG-cumulative of two alleles						
≤42	183	34.9	185	40.1	1.0	Reference
≥43	341	65.1	276	59.9	1.3	1.0–1.7
GGC length[c]						
S/S	405	78.6	319	72.0	1.0	Reference
S/L	84	16.3	90	20.3	0.7	0.5–1.0
L/L	26	5.0	34	7.7	0.5	0.3–0.9
GGC length[c]						
S/S	405	78.5	319	71.7	1.0	Reference
Any L	111	21.5	126	28.3	0.7	0.5–0.9
GGC-cumulative of two alleles						
≤32	430	83.5	345	77.9	1.0	Reference
≥33	85	16.5	98	22.1	0.7	0.5–0.9

[a] Adjusted for age at reference and reference year.
[b] Short (S) = <22 repeats, long (L) = ≥22 repeats.
[c] Short (S) = <17 repeats, long (L) = ≥17 repeats.

Table 3   Association of OC use with breast cancer risk, separately by CAG genotype

	All women				CAG S/S[a]							CAG any L[a]								
	Cases	Controls	OR[b]	95% CI	Cases		Controls		OR[b]	95% CI	P[c]	Cases		Controls		OR[b]	95% CI	P[c]	P[d]	
	%	%			N	%	N	%				N	%	N	%					
OC use																				
Never[e]	17.4	22.1	1.0	Reference	21	17.4	34	27.9	1.0	Reference		70	17.4	68	20.1	1.0	Reference			
Ever	82.6	77.9	1.3	0.9–1.8	100	82.6	88	72.1	1.9	1.0–3.6		333	82.6	271	79.9	1.1	0.8–1.6		0.16	
OC duration																				
Never[e]	17.4	22.1	1.0	Reference	21	17.4	34	27.9	1.0	Reference		70	17.4	68	20.1	1.0	Reference			
<5 yrs	43.1	40.7	1.3	0.9–1.8	53	43.8	46	37.7	1.8	0.9–3.6		173	42.9	141	41.6	1.1	0.7–1.7			
5 to <10 yrs	25.4	22.5	1.4	0.9–2.1	24	19.8	23	18.9	1.9	0.8–4.4		109	27.0	81	23.9	1.2	0.8–1.9			
≥10 yrs	14.1	14.7	1.2	0.7–1.8	23	19.0	19	15.6	2.1	0.9–5.0	0.66	51	12.7	49	14.5	0.9	0.5–1.6	0.55	0.43	
OC recency																				
Never used[e]	17.4	22.1	1.0	Reference	21	17.4	34	27.9	1.0	Reference		70	17.4	68	20.1	1.0	Reference			
≥10 yrs ago	53.2	48.3	1.3	0.9–1.8	60	49.6	55	45.1	1.6	0.8–3.3		219	54.3	167	49.3	1.2	0.8–1.8			
5 to <10 yrs	13.2	13.4	1.2	0.8–1.9	17	14.0	11	9.0	2.2	0.8–5.7		52	12.9	51	15.0	0.9	0.5–1.6			
<5 yrs/current	16.2	16.2	1.4	0.9–2.1	23	19.0	22	18.0	2.3	1.0–5.6	0.34	62	15.4	53	15.6	1.1	0.6–1.8	0.96	0.42	
Age at first OC use																				
Never used[e]	17.4	22.1	1.0	Reference	21	17.4	34	27.9	1.0	Reference		70	17.4	68	20.1	1.0	Reference			
≥22 yrs	16.0	13.2	1.4	0.9–2.2	18	14.9	16	13.1	1.5	0.6–3.7		66	16.4	45	13.3	1.3	0.8–2.2			
18–21 yrs	50.0	46.3	1.4	1.0–1.9	57	47.1	50	41.0	2.0	1.0–4.0		205	50.9	163	48.1	1.2	0.8–1.7			
<18 yrs	16.6	18.4	1.1	0.7–1.7	25	20.7	22	18.0	1.9	0.8–4.4	0.68	62	15.4	63	18.6	0.8	0.5–1.4	0.18	0.36	

[a] Short (S) = <22 repeats, long (L) = ≥22 repeats.
[b] Adjusted for age at reference and reference year.
[c] P for trend among "ever" users within CAG genotype.
[d] P for interaction of OC use and CAG genotype.
[e] "Never" includes women who used <6 months.

1.41; 95% CI, 0.95–2.09). A recent Canadian study of 255 cases, 272 general population controls, and 189 hospital controls observed an approximate 50% reduction in the risk of breast cancer for women with a smaller cumulative number of (CAG)$_n$ repeats (21). This reduction in risk was primarily confined to postmenopausal women; there was little evidence of any association in premenopausal women. A larger study of

British women (508 cases and 426 controls) found no significant or consistent association between (CAG)$_n$ repeat length and breast cancer risk (23). Neither of these last two studies is readily comparable with the data presented here, however, because these studies focused primarily on older women in contrast to the younger age of our study population. To date, the only study that has shown a significant increased risk of breast

Table 4 Association of OC use with breast cancer risk, separately by GGC genotype

	All women				GGC S/S[a]						GGC any L[a]						
	Cases	Controls	OR[b]	95% CI	Cases		Controls		OR[b]	95% CI	P[c]	Cases		Controls		OR[b]	95% CI	P[c]	P[d]
	%	%			N	%	N	%				N	%	N	%				
OC use																			
Never[e]	17.4	22.1	1.0	Reference	74	18.3	67	21.0	1.0	Reference		16	14.4	29	23.0	1.0	Reference		
Ever	82.6	77.9	1.3	0.9–1.8	331	81.7	252	79.0	1.1	0.8–1.7		95	85.6	97	77.0	1.7	0.9–3.5		0.33
OC duration																			
Never[e]	17.4	22.1	1.0	Reference	74	18.3	67	21.0	1.0	Reference		16	14.4	29	23.0	1.0	Reference		
<5 years	43.1	40.7	1.3	0.9–1.8	175	43.2	137	42.9	1.1	0.7–1.7		49	44.1	46	36.5	1.8	0.9–4.0		
5 to <10 years	25.4	22.5	1.4	0.9–2.1	98	24.2	75	23.5	1.2	0.7–1.8		31	27.9	25	19.8	2.2	1.0–5.1		
≥10 years	14.1	14.7	1.2	0.7–1.8	58	14.3	40	12.5	1.2	0.7–2.1	0.60	15	13.5	26	20.6	1.0	0.4–2.5	0.21	0.25
OC recency																			
Never used[e]	17.4	22.1	1.0	Reference	74	18.3	67	21.0	1.0	Reference		16	14.4	29	23.0	1.0	Reference		
≥10 yrs ago	53.2	48.3	1.3	0.9–1.8	211	52.1	158	49.5	1.1	0.7–1.7		63	56.8	61	48.4	1.9	0.9–4.0		
5 to <10 yrs	13.2	13.4	1.2	0.8–1.9	53	13.1	44	13.8	1.1	0.6–1.8		16	14.4	16	12.7	1.7	0.7–4.5		
<5 yrs/current	16.2	16.2	1.4	0.9–2.1	67	16.5	50	15.7	1.3	0.8–2.3	0.26	16	14.4	20	15.9	1.3	0.5–3.5	0.45	0.71
Age at first OC use																			
Never used[e]	17.4	22.1	1.0	Reference	74	18.3	67	21.0	1.0	Reference		16	14.4	29	23.0	1.0	Reference		
≥22 yrs	16.0	13.2	1.4	0.9–2.2	72	17.8	41	12.9	1.5	0.9–2.5		11	9.9	20	15.9	0.9	0.3–2.4		
18–21 yrs	50.0	46.3	1.4	1.0–1.9	192	47.4	155	48.6	1.1	0.7–1.7		64	57.7	49	38.9	2.4	1.1–5.2		
<18 yrs	16.6	18.4	1.1	0.7–1.7	67	16.5	56	17.6	1.0	0.6–1.7	0.30	20	18.0	28	22.2	1.2	0.5–2.9	0.86	0.05

[a] Short (S) = <17 repeats, long (L) = ≥17 repeats.
[b] Adjusted for age at reference and reference year.
[c] P for trend among "ever" users within GGC genotype.
[d] P for interaction of OC use and GGC genotype.
[e] "Never" includes women who used <6 months.

cancer in relation to longer *AR* (CAG)$_n$ repeat length is that of Rebbeck *et al.* (18), who reported an increased breast cancer risk in relation to the presence of (CAG)$_n$ alleles of repeat length ≥28 in a population of *BRCA1* mutation-positive women. Our study included very few known *BRCA1* or *BRCA2* carriers, and exclusion of those women did not affect overall results.

Neither of the two prior studies to date that have examined (GGC)$_n$ repeat length and breast cancer risk reported any association (19, 23). The results presented here, which indicate that among young women, the presence of an L (GGC)$_n$ allele (repeat length ≥17) is associated with a significantly reduced risk of breast cancer compared with the homozygous S genotype, are the first. Specifically, our results demonstrate a significant 30% reduction in breast cancer risk for heterozygotic carriers of a long (GGC)$_n$ allele and a 50% reduction in risk for those carrying two long (GGC)$_n$ alleles. These results, as with the (CAG)$_n$ repeat results, may not be directly comparable with previous studies because the latter have focused on women of all ages and/or specifically on *BRCA1/2* carriers (19, 23).

We did not observe any combined effects for (CAG)$_n$ and (GGC)$_n$ when they were considered jointly. However, we note that the sample size for such analyses was likely insufficient.

Interestingly, whereas no previous studies have observed an effect in terms of breast cancer, several studies do link (GGC)$_n$ repeat length with prostate cancer risk (7, 14–17). Only one study (31) failed to find an association between GGC repeat length and prostate cancer risk. These data suggest a nonspecific role for the (GGC)$_n$ repeat in risk of hormonal cancers overall.

Whereas a number of risk factors were found not to vary in effect by *AR* genotype, there was some suggestion of modification by genotype of the effect of OCs on breast cancer risk. Specifically, among women who were homozygous for the (CAG)$_n$ S allele and among women with any L (GGC)$_n$ allele, those who ever used OCs had a respective 90% and 70% excess

risk of breast cancer compared with those of the same genotype who never used OCs. No association with OC use was seen among women with any L (CAG)$_n$ allele and women homozygous for the S (GGC)$_n$ allele. Although there was some fluctuation in risk estimates for aspects of OC use such as duration and recency within the two genotypic groups that exhibited an increased risk in relation to ever use of OCs, there was no substantive evidence that the increased risk seen for ever use was further altered or explained by increased duration of use or time since last use. This persistent elevation in risk across increasing duration of use, for instance, could suggest that for women with certain genotypic profiles, any exposure to OCs, regardless of the duration of use, might be related to an increased risk of breast cancer.

Our results for OC use should be considered within the context of the multicenter study from which these women were drawn (25). The multicenter study evaluated the relationship between OCs and breast cancer risk in 1648 cases and 1505 controls under the age of 45 years (25). Main findings from that study included the fact that breast cancer risk was modestly increased in relation to ever use of OCs and that risk did not increase further in relation to duration of use. In both data from the Seattle study site used for analyses of AR and the larger multicenter study results reported by Brinton *et al.* (25), similar increases in risk were seen for ever users of OCs [OR = 1.3 (95% CI, 0.9–1.8) and OR = 1.3 (95% CI, 1.1–1.5), respectively; data not shown]. The results presented here suggest that a proportion of the overall increase in risk related to OC use may be attributable to *AR* genotype. They do not, however, exclude the possibility of a role for other genes and/or additional environmental exposures.

Our data are consistent with a role for *AR* repeats and interactions with hormones relevant to breast growth and development. However, no precise model can easily be drawn that explains our results in aggregate, due in part to the still incomplete understanding of the precise functional role that the *AR*

repeats play in hormone regulation (32). The role of androgens themselves in breast cancer is also not well understood, with higher levels of adrenal androgens noted in women who develop the disease in the postmenopausal years compared with those who develop it earlier in life (33). $(CAG)_n$ repeat expansion has consistently been associated with a reduction in the ability of the *AR* to activate transcription (9, 10, 34). However, analysis of $(CAG)_n$ repeat reduction and function of the neighboring $(GGN)_n$ repeat have only been partly investigated. Within the $(GGN)_n$ repeat, the non-normal $(GGC)_n$ repeat distribution found in our study, largely due to the high frequency of 16- and 17-repeat alleles (observed in 69.9% and 15.3% of the population, respectively), may provide some insight. Gao *et al.* (34) hypothesized that preference for a particular repeat size may be due to the physical and structural organization of the protein. Their study found a decrease in transcription activation levels in the absence of either the $(GGN)_n$ or $(CAG)_n$ repeat or in the presence of $(CAG)_n$ repeat lengths other than 20 (35). However, earlier studies contradict this result, showing increased transcription activation when the $(CAG)_n$ repeat was eliminated or reduced in size (9, 10).

A few studies have attempted to directly assess the possible association between *AR* $(CAG)_n$ repeat length and endogenous hormone levels. Ancillary to their main case-control study, Haiman *et al.* (22) examined the relationship of the $(CAG)_n$ repeat length with circulating (serum) steroid hormone levels in approximately 450 postmenopausal controls. Although no associations with repeat length were statistically significant, there were suggestive relationships with testosterone and dihydroepiandrosterone sulfate. A small study of 239 premenopausal women recently reported an association between short repeat alleles and higher testosterone levels in the follicular phase of the menstrual cycle (35). At least one large study in men has demonstrated a significant relationship between percentage change in free testosterone and increased number of $(CAG)_n$ repeats (36). Clearly, additional studies are needed before any conclusions can be drawn regarding the functional significance of normal length variation, particularly with regard to the $(GGC)_n$ repeat length within the $(GGN)_n$ repeat.

The results of this study should be assessed within the context of its limits. We note specifically that there are differences between tested and untested women. Tested cases were more likely to be alive, older, and to have a less advanced stage of cancer than untested cases. Thus, the generalizability of our results, although from a population-based study, must be viewed with caution because the underrepresentation in our study population of cases with advanced disease could result in potential bias if there are associations between genotype and prognosis. In addition, tested women, particularly controls, were more likely to use OCs than untested women ($P = 0.01$ for controls and $P = 0.026$ for cases). The fact that tested controls were more likely to be OC users could suggest that observed risk estimates may be underestimates of the true magnitude of risk. In addition, the total number of cases in many cells was limited. Repetition of these analyses with larger data sets allowing for larger numbers of women in various strata is warranted to best understand the relationship between *AR* polymorphisms, breast cancer, and OC use.

A further limitation of the study is that the $(GGC)_n$ repeat genotype categories were not in compliance with Hardy-Weinberg equilibrium. Rather, we observed an excess of both LL and SS homozygotes compared with SL heterozygotes. Several factors may account for this. It is formally possible that the observed polymorphisms in the *AR* gene may influence mating,

fertility, or fecundity and/or mortality, any of which could in turn influence changes in the gene frequencies in a population. It is more likely that the explanation is methodologic. We speculate that the presence of a single nucleotide polymorphism in one of the $(GGC)_n$ repeat primer sites could lead to an overestimation of homozygotes at the expense of heterozygotes by preferential amplification of the allele without the change. DNA sequencing of several putative homozygote samples did not reveal any such single nucleotide polymorphism (data not shown). We also note that there was a small set of 27 samples for which the presence of a second allele could neither be confirmed nor denied due to the existence of a broad "stutter band" on the gel associated with a different molecular weight allele. The presence of such stutter bands is commonly observed when amplifying segments of DNA that contain repeats. It is possible that inclusion of these samples could have altered our results slightly. However, reanalysis of the data after exclusion of individuals for whom only one genotype was present did not alter our results (data not shown). Finally, we note that the fact that X-inactivation occurs randomly early in development means that there will be variation in *AR* polymorphism expression between women with the same genotype in any organ of interest. This added degree of variation might account, at least in part, for the variability in results between studies.

This study does have a number of important strengths. Because our assessment of *AR* repeat effects on breast cancer risk includes a large and unselected proportion of all women with breast cancer before age 45 years during a defined time period and in a defined geographic area, it produces results that are applicable to the general population of women under age 45 years. In addition, the wealth of data collected for these women allowed for the consideration of important potential confounding and effect-modifying factors.

The strength of the association we observe between the *AR* $(GGC)_n$ repeat and the risk of breast cancer and the fact that this is the first study to examine this in relation to younger onset breast cancer suggest that additional studies of this association are needed. This is the first study, to our knowledge, that has shown an association between OC use and breast cancer risk among women with certain *AR* genotypes. These findings also warrant replication, although their interpretation is limited somewhat by our sample size. Resolution may ultimately require the pooling of data sets for questions such as this. These results lend further support to the need for more detailed functional studies aimed at understanding the role of *AR* repeats in the etiology of breast cancer in young women.

References

1. Lumbroso, R., Beitel, L. K., Vasiliou, D. M., Trifiro, M. A., and Pinsky, L. Codon-usage variants in the polymorphic $(GGN)_n$ trinucleotide repeat of the human androgen receptor gene. Hum. Genet., *101:* 43–46, 1997.

2. Faber, P. W., Kuiper, G. G., van Rooij, H. C., van der Korput, J. A., Brinkmann, A. O., and Trapman, J. The N-terminal domain of the human androgen receptor is encoded by one, large exon. Mol. Cell. Endocrinol., *61:* 257–262, 1989.

3. Jenster, G., van der Korput, H. A., Trapman, J., and Brinkmann, A. O. Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. J. Biol. Chem., *270:* 7341–7346, 1995.

4. Irvine, R. A., Yu, M. C., Ross, R. K., and Coetzee, G. A. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res., *55:* 1937–1940, 1995.

5. Ingles, S. A., Ross, R. K., Yu, M. C., Irvine, R. A., La Pera, G., Haile, R. W., and Coetzee, G. A. Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. J. Natl. Cancer Inst. (Bethesda), *89:* 166–170, 1997.

6. Giovannucci, E., Stampfer, M. J., Krithivas, K., Brown, M., Dahl, D., Brufsky, A., Talcott, J., Hennekens, C. H., and Kantoff, P. W. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer [published erratum appears in Proc. Natl. Acad. Sci. USA, *94:* 8272, 1997]. Proc. Natl. Acad. Sci. USA, *94:* 3320–3323, 1997.

7. Stanford, J. L., Just, J. J., Gibbs, M., Wicklund, K. G., Neal, C. L., Blumenstein, B. A., and Ostrander, E. A. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res., *57:* 1194–1198, 1997.

8. Modugno, F., Weissfeld, J. L., Trump, D. L., Zmuda, J. M., Shea, P., Cauley, J. A., and Ferrell, R. E. 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.

9. Chamberlain, N. L., Driver, E. D., and Miesfeld, R. L. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res., *22:* 3181–3186, 1994.

10. Kazemi-Esfarjani, P., Trifiro, M. A., and Pinsky, L. 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, 1995.

11. Tut, T. G., Ghadessy, F. J., Trifiro, M. A., Pinsky, L., and Yong, E. L. Long polyglutamine tracts in the androgen receptor are associated with reduced transactivation, impaired sperm production, and male infertility. J. Clin. Endocrinol. Metab., *82:* 3777–3782, 1997.

12. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature (Lond.), *352:* 77–79, 1991.

13. Mhatre, A. N., Trifiro, M. A., Kaufman, M., Kazemi-Esfarjani, P., Figlewicz, D., Rouleau, G., and Pinsky, L. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. Nat. Genet., *5:* 184–188, 1993.

14. Hsing, A. W., Gao, Y. T., Wu, G., Wang, X., Deng, J., Chen, Y. L., Sesterhenn, I. A., Mostofi, F. K., Benichou, J., and Chang, C. Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. Cancer Res., *60:* 5111–5116, 2000.

15. Hakimi, J. M., Schoenberg, M. P., Rondinelli, R. H., Piantadosi, S., and Barrack, E. R. Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. Clin. Cancer Res., *3:* 1599–1608, 1997.

16. Platz, E. A., Giovannucci, E., Dahl, D. M., Krithivas, K., Hennekens, C. H., Brown, M., Stampfer, M. J., and Kantoff, P. W. The androgen receptor gene GGN microsatellite and prostate cancer risk. Cancer Epidemiol. Biomark. Prev., *7:* 379–384, 1998.

17. Chang, B-L., Zheng, S. L., Hawkins, G. A., Isaacs, S. D., Wiley, K. E., Turner, A., Carpten, J. D., Bleecker, E. R., Walsh, P. C., Trent, J. M., Meyers, D. A., Isaacs, W. B., and Xu, J. Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk. Hum. Genet., *110:* 122–129, 2002.

18. Rebbeck, T. R., Kantoff, P. W., Krithivas, K., Neuhausen, S., Blackwood, M. A., Godwin, A. K., Daly, M. B., Narod, S. A., Garber, J. E., Lynch, H. T., Weber, B. L., and Brown, M. Modification of BRCA1-associated breast cancer risk by the polymorphic androgen-receptor CAG repeat. Am. J. Hum. Genet., *64:* 1371–1377, 1999.

19. Kadouri, L., Easton, D. F., Edwards, S., Hubert, A., Kote-Jarai, Z., Glaser, B., Durocher, F., Abeliovich, D., Peretz, T., and Eeles, R. A. CAG and GGC repeat polymorphisms in the androgen receptor gene and breast cancer susceptibility in BRCA1/2 carriers and non-carriers. Br. J. Cancer, *85:* 36–40, 2001.

20. Menin, C., Banna, G. L., De Salvo, G., Lazzarotto, V., De Nicolo, A., Agata, S., Montagna, M., Sordi, G., Nicoletto, O., Chieco-Bianchi, L., and D'Andrea, E. Lack of association between androgen receptor CAG polymorphism and familial breast/ovarian cancer. Cancer Lett., *168:* 31–36, 2001.

21. Giguère, Y., Dewailly, E., Brisson, J., Ayotte, P., Laflamme, N., Demers, A., Forest, V. I., Dodin, S., Robert, J., and Rousseau, F. Short polyglutamine tracts in the androgen receptor are protective against breast cancer in the general population. Cancer Res., *61:* 5869–5874, 2001.

22. Haiman, C. A., Brown, M., Hankinson, S. E., Spiegelman, D., Colditz, G. A., Willett, W. C., Kantoff, P. W., and Hunter, D. J. The androgen receptor CAG repeat polymorphism and risk of breast cancer in the Nurses' Health Study. Cancer Res., *62:* 1045–1049, 2002.

23. Dunning, A. M., McBride, S., Gregory, J., Durocher, F., Foster, N. A., Healey, C. S., Smith, N., Pharoah, P. D., Luben, R. N., Easton, D. F., and Ponder, B. A. No association between androgen or vitamin D receptor gene polymorphisms and risk of breast cancer. Carcinogenesis (Lond.), *20:* 2131–2135, 1999.

24. Spurdle, A. B., Dite, G. S., Chen, X., Mayne, C. J., Southey, M. C., Batten, L. E., Chy, H., Trute, L., McCredie, M. R., Giles, G. G., Armes, J., Venter, D. J., Hopper, J. L., and Chenevix-Trench, G. Androgen receptor exon 1 CAG repeat length and breast cancer in women before age forty years. J. Natl. Cancer Inst. (Bethesda), *91:* 961–966, 1999.

25. Brinton, L. A., Daling, J. R., Liff, J. M., Schoenberg, J. B., Malone, K. E., Stanford, J. L., Coates, R. J., Gammon, M. D., Hanson, L., and Hoover, R. N. Oral contraceptives and breast cancer risk among younger women. J. Natl. Cancer Inst. (Bethesda), *87:* 827–835, 1995.

26. Waksberg, J. Sample methods for random digit dialing. J. Am. Stat. Soc., *73:* 40–46, 1978.

27. Malone, K. E., Daling, J. R., Thompson, J. D., O'Brien, C. A., Francisco, L. V., and Ostrander, E. A. BRCA1 mutations and breast cancer in the general population: analyses in women before age 35 years and in women before age 45 years with first-degree family history. JAMA, *279:* 922–929, 1998.

28. Malone, K. E., Daling, J. R., Neal, C., Suter, N. M., O'Brien, C., Cushing-Haugen, K., Jonasdottir, T. J., Thompson, J. D., and Ostrander, E. A. Frequency of BRCA1/BRCA2 mutations in a population-based sample of young breast carcinoma cases. Cancer (Phila.), *88:* 1393–1402, 2000.

29. Breslow, N. E., and Day, N. E. Statistical Methods in Cancer Research. Volume I: The Analysis of Case-Control Studies. IARC Sci. Publ. No. 32, pp. 5–338. Lyon, France: IARC, 1980.

30. Guo, S. W., and Thompson, E. A. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics, *48:* 361–372, 1992.

31. Correa-Cerro, L., Wohr, G., Haussler, J., Berthon, P., Drelon, E., Mangin, P., Fournier, G., Cussenot, O., Kraus, P., Just, W., Paiss, T., Cantu, J. M., and Vogel, W. $(CAG)_n$CAA and GGN repeats in the human androgen receptor gene are not associated with prostate cancer in a French-German population. Eur. J. Hum. Genet., *7:* 357–362, 1999.

32. Brys, M. Androgens and androgen receptor: do they play a role in breast cancer? Med. Sci. Monit., *6:* 433–438, 2000.

33. Adams, J. B. Adrenal androgens and human breast cancer: a new appraisal. Breast Cancer Res. Treat., *51:* 183–188, 1998.

34. Gao, T., Marcelli, M., and McPhaul, M. J. Transcriptional activation and transient expression of the human androgen receptor. J. Steroid Biochem Mol. Biol., *59:* 9–20, 1996.

35. Westberg, L., Baghaei, F., Rosmond, R., Hellstrand, M., Landen, M., Jannson, M., Holm, G., Bjorntorp, P., and Eriksson, E. Polymorphisms of the androgen receptor β gene are associated with androgen levels in women. J. Clin. Endocrinol. Metab., *86:* 2562–2568, 2001.

36. Krithivas, K., Yurgalevitch, S. M., Mohr, B., Wilcox, C. J., Batter, S. J., Brown, M., Longcope, C., McKinlay, J. B., and Kantoff, P. W. Evidence that the CAG repeat in the androgen receptor gene is associated with the age-related decline in serum androgen levels in men. J. Endocrinol., *162:* 137–142, 1999.