DOI:10.1158/1055-9965.EPI-06-0868

Cancer Epidemiology, Biomarkers & Prevention

Trinucleotide Repeat Polymorphisms in the Androgen Receptor Gene and Risk of Ovarian Cancer

Joellen M. Schildkraut, Susan K. Murphy, Rachel T. Palmieri, et al.

Cancer Epidemiol Biomarkers Prev 2007;16:473-480. Published online March 19, 2007.

DOI:10.1158/1055-9965.EPI-06-0868

Trinucleotide Repeat Polymorphisms in the Androgen Receptor Gene and Risk of Ovarian Cancer

Joellen M. Schildkraut,[1] Susan K. Murphy,[2] Rachel T. Palmieri,[7] Edwin Iversen,[3] Patricia G. Moorman,[1] Zhiqing Huang,[2] Susan Halabi,[4] Brian Calingaert,[6] Alison Gusberg,[2] Jeffrey R. Marks,[5] and Andrew Berchuck[2]

[1]Department of Community and Family Medicine; [2]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology; [3]Institute of Statistics and Decision Sciences; Departments of [4]Biostatistics and Bioinformatics and [5]Surgery; and [6]Comprehensive Cancer Center, Duke University Medical Center, Durham, North Carolina; and [7]Department of Epidemiology, The University of North Carolina, Chapel Hill, North Carolina

Abstract

Introduction: **Androgens may play a role in the development of ovarian cancers. Two trinucleotide repeat polymorphisms have been described in exon 1 of the androgen receptor (AR) gene that may affect its function. Previous studies of ovarian cancer and AR repeat polymorphisms have been inconsistent.**
Methods: **We analyzed CAG and GGC repeat length polymorphisms in the AR gene using data from a population-based case-control study of ovarian cancer that included 594 cases and 681 controls. Repeat lengths were determined by fluorescent DNA fragment analysis using ABI GeneScan software. Change point models were used to determine appropriate repeat length cutoff points by race (African American versus Caucasian) for both the shorter and longer CAG and GGC repeats.**
Results: **No relationship was observed between CAG repeat length and ovarian cancer among Caucasians. Among African Americans, having a short repeat length on either allele was associated with a 2-fold increase in ovarian cancer risk (age-adjusted odds ratio, 2.2; 95% confidence interval, 1.1-4.1). Having short CAG repeat lengths for both alleles was associated with a 5-fold increased risk for developing ovarian cancer (age-adjusted odds ratio, 5.4; 95% confidence interval, 1.4-1.7). No relationship with the GGC repeat length polymorphisms was observed.**
Conclusion: **These results suggest that having a short CAG repeat length in AR increases ovarian cancer risk in African Americans. The failure to observe this relationship in Caucasians may be due to the rarity of such short CAG alleles in this population or could reflect racial differences in disease etiology.** (Cancer Epidemiol Biomarkers Prev 2007;16(3):473–80)

Introduction

It has been suggested that androgens may play a role in the development of ovarian cancer (1, 2). Androgen is produced by ovarian theca lutein cells and androgen receptors (AR) are found in the normal surface epithelium of the ovaries. Most ovarian cancers express AR and antiandrogens inhibit ovarian cancer growth (3-7). Epidemiologic studies also support a role of androgen in ovarian cancer in which increasing waist-to-hip ratio (8, 9) and polycystic ovarian syndrome (9, 10), which may be correlated with elevated androgen levels in women, have been associated with increased risk of ovarian cancer. In one study, higher levels of serum androstenedione were reported among women diagnosed with ovarian cancer compared with controls (11). Additionally, oral contraceptive use, which is inversely associated with ovarian cancer risk, suppresses testosterone production by 35% to 70% (1, 2).

Two highly polymorphic trinucleotide repeat polymorphisms in exon 1 of the AR gene have been studied in relation to cancer risk (12). The CAG trinucleotide repeat of *AR* encodes a polyglutamine tract (13), the length of which has been shown to be inversely associated with the ability of the AR-ligand complex to transactivate androgen-responsive genes. Molecular analyses have shown that the transactivational capacity of the AR decreases with increasing number of glutamines encoded by the CAG repeat tract (14). Indeed, shorter AR CAG repeat lengths are associated with a higher risk of prostate cancer (15). Racial differences in *AR* CAG repeat length have been noted, with African Americans having a lower mean CAG length as compared with Caucasians (16, 17). A second *AR* GGC trinucleotide repeat polymorphism codes for a polyglycine tract of variable length (18) but its functional significance has not been extensively examined. Data from one study suggest that whereas AR transactivation activity may not be affected by GGC repeats, translation of *AR* mRNA may be inversely related to GGC repeats with increased AR protein produced from alleles with shorter GGC repeats (19). This suggests that shorter GGC repeats may result in an increased capacity to respond to androgen exposure.

There are five published studies that have addressed the association between CAG repeat length and ovarian cancer. Two case-control studies reported an increased risk of ovarian cancer associated with increasing CAG repeat length among Caucasian women (20, 21). The data from both studies suggest that women who carry two alleles with ≥22 CAG repeats are more likely to develop ovarian cancer than those with two alleles with <22 repeats. However, other published studies have not found evidence to support the association between longer CAG repeat length and ovarian cancer (12, 22, 23), although there may have been little power to detect an association due to the small number of ovarian cancer cases in two of the studies (12, 23). Kadouri et al. (12) also examined the relationship between GGC repeat length and 29 ovarian cancer cases and did not find evidence to support a relationship.

In view of the conflicting data about the relationship between *AR* repeat polymorphisms and ovarian cancer risk,

Received 10/15/06; revised 12/13/06; accepted 1/10/07.

Grant support: NIH grant 2-R01-CA76016 and Department of Defense grant DAMD17-02-1-0666.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Joellen M. Schildkraut, Duke University Medical Center, Box 2949, Durham, NC 27710. Phone: 919-681-4761; Fax: 919-681-4766. E-mail: schil001@mc.duke.edu

doi:10.1158/1055-9965.EPI-06-0868

DOI:10.1158/1055-9965.EPI-06-0868

we examined this relationship in a large population-based, case-control study of ovarian cancer in North Carolina. In contrast to prior studies, change point statistical analysis was used to determine the appropriate threshold for dichotomizing repeat lengths. In addition, this is the first study to examine the relationship between AR repeat length polymorphisms and ovarian cancer risk in a relatively large group of African American women. This is of interest because African Americans have shorter CAG repeat lengths relative to Caucasians, which might mediate an increased respond to androgen exposure (16).

Materials and Methods

Subjects. Study subjects were enrolled through the ongoing North Carolina Ovarian Cancer study, a population-based, case-control study of newly diagnosed epithelial ovarian cancer. Epithelial ovarian cancer cases were identified through the North Carolina Central Cancer Registry, a statewide population-based tumor registry, using rapid case ascertainment. Pathology reports for all ovarian cancer cases diagnosed in the study area were forwarded to the North Carolina Central Cancer Registry and then to the study office within 2 months of diagnosis. Eligibility criteria for ovarian cancer cases include diagnosis since January 1, 1999; age 20 to 74 years at diagnosis; no prior history of ovarian cancer; and residence in a 48-county area of North Carolina. For data included in the current analyses, the last diagnosis among Caucasians was November 2003. To maximize the number of African American subjects, the last date of diagnosis was extended to October 2005. All participants were English-speaking, mentally competent to complete an interview, and able to give informed consent. Physician permission was obtained before an eligible case was contacted. All cases underwent standardized pathologic and histologic review by the study pathologist to confirm diagnosis. Both invasive and borderline epithelial ovarian cancer cases were included. The response rate among eligible cases was 75%. Nonresponders were classified as patient refusal (7%), inability to locate the patient (9%), physician refusal (4%), death (4%), or debilitating illness (2%).

Population-based controls were identified from the same 48-county region as the cases and were frequency matched to the ovarian cancer cases on the basis of race (African American and Caucasian) and age (5-year age categories) using list-assisted random digit dialing. As required for the cases, controls had to be English-speaking, mentally competent to complete an interview, and able to give informed consent. Potential controls were screened for eligibility and were required to have at least one intact ovary and no prior diagnosis of ovarian cancer. Seventy-three percent of controls identified by random digit dialing, who passed the eligibility screening, agreed to be contacted and sent additional study information. Among those who sent additional study information, the response rate was 64%. Nonresponders were classified as refusal, 27%, and unable to contact, 9%. The study protocol was approved by the Duke University Medical Center Institutional Review Board and the Human Subjects committees at the North Carolina Central Cancer Registry and each of the hospitals where cases were identified.

Questionnaire Data. Trained nurse interviewers obtained written informed consent from study subjects at the time of the interview, which was usually conducted in the home of the study subject. A 90-min questionnaire was administered to obtain information on known and suspected ovarian cancer risk factors including family history of cancer in first-degree and second-degree relatives, menstrual characteristics, pregnancy and breast-feeding history, infertility, hormone use, and lifestyle characteristics such as smoking, alcohol consumption, physical activity, and occupational history. A life events calendar, which marked significant life events including marriage and education, was used to improve recall of reproductive and contraceptive history. Additionally, anthropometric descriptors (height, weight, waist and hip circumference) were measured and a blood sample (30 mL) was collected.

Laboratory Analyses

DNA Extraction. Germ line DNA was extracted from peripheral blood lymphocytes using PureGene DNA isolation reagents according to manufacturer's instructions (Gentra Systems, Minneapolis, MN).

AR Trinucleotide Repeat Length Analysis. Thirty nanograms of genomic DNA were used as template for PCR amplification of the region containing the CAG and GGC trinucleotide repeats in 25-μL reaction volumes. The CAG repeat was amplified using primers previously reported (12) with the exception that the forward primer was modified by the addition of a 5′ fluorescent label (6-carboxyfluorescein; 6-FAM). The GGC repeat was amplified using two rounds of PCR with primers as described (24). For the GGC repeat analysis, the forward primer used in the second round of PCR was labeled with 6-FAM.

PCR for the CAG repeat was done using Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA) with conditions as follows: 94°C for 3 min, then five rounds (four cycles each) of 94°C for 30 s, 64°C for 30 s for round 1, then decreased by 2°C each round down to 56°C, 72°C for 30 s, followed by 29 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 30 s, followed by a final 5-min extension at 72°C. For the GGC repeat, Pfu DNA polymerase (Stratagene, La Jolla, CA) was used with PCR conditions as follows: round 1, 98°C for 45 s, then 17 cycles of 98°C for 1 min and 70°C for 5 min, then a 10-min extension at 70°C. One microliter of the first-round PCR products was used as template for the second round of PCR under the same conditions except that 34 cycles of PCR were done.

The PCR products for both repeats were diluted 1:100 in nuclease-free water and these dilutions were run on an Applied Biosystems 3100 Automated Capillary Instrument followed by fragment analysis using GeneScan Analysis software (Applied Biosystems; Foster City, CA). To independently validate the fragment length call, a subset of samples were also analyzed by nucleotide sequencing after purification from high-resolution agarose gels of individual amplicons produced from each allele. Unlabeled forward primers were used for sequencing the amplicons using an ABI 3730 Prism capillary DNA sequencer for the CAG repeat ($n = 6$) or, for the GGC repeat ($n = 15$), the Thermosequenase Radiolabeled Terminator Cycle Sequencing Kit (U.S. Biochemical Corp., Cleveland, OH). The GGC sequencing reactions were resolved on denaturing 5% polyacrylamide sequencing gels followed by exposure at −80°C to Kodak BioMax MR radiographic film with an intensifying screen. The length of the trinucleotide repeats by sequencing was found to be longer relative to the repeat length determined by fragment length call of the GeneScan software in all cases. Because the sequencing results provide direct visualization of the number of repeats present, we systematically adjusted the GeneScan fragment lengths by the addition of 12.0 nucleotides (4 repeats) for the CAG repeat analysis and 8.4 nucleotides (2.8 repeats) for the GGC repeat analysis.

Statistical Analysis. We did a two-stage analysis of the association between repeat length and ovarian cancer. In the first stage, we used Bayesian model selection and model averaging to determine the weight of evidence in the data for each possible cutoff point in repeat length and to estimate an average (over threshold values) measure of association (25). This approach allows us to determine if the association is significant marginal to the choice of threshold and obviates the

need for a multiple comparisons adjustment. In the second stage, we fit multivariate models of association fixing the repeat length threshold to its most probable value a posteriori. The purpose of these second stage analyses was to verify that the observed associations between repeat length and ovarian cancer were not confounded.

In stage 1, separate Bayesian change point models were fit to the self-reported African American and Caucasian short and long CAG repeat alleles, CAG_S and CAG_L, respectively, and short and long GGC repeat alleles, GGC_S and GGC_L, respectively. The designation of CAG_S and CAG_L, as well as of GGC_S and GGC-L, reflects the comparison of the repeat length of the two alleles within an individual. The change point model specifies that odds of disease is constant before and after a threshold value, but is different in the two regions. We used uniform prior probabilities over the possible discrete thresholds of the data and on whether or not there is a change point and used a $\beta(2, 2)$ prior over pre- and post-threshold probabilities of disease. The $\beta(2, 2)$ distribution has a mean of 0.5 and a SD of 0.22. This is equivalent to adding two cases and two controls in each of the pre- and post-threshold samples. Under this model, we calculated (*a*) the probability that there was a change point and, given that there was, (*b*) the probability that it occurred at each of the possible values. Calculation (*a*) was equivalent to a Bayesian hypothesis test of Ho: the case control fraction does not depend on a thresholded short CAG repeat length versus Ha: that it does. We report the posterior probability for association of disease to CAG_S length, the posterior probabilities of the thresholds given that the threshold model is the true model, and the odds ratio (OR) for disease given a CAG_S repeat smaller than each probable threshold. In addition, we calculate estimates of ORs that account for uncertainty in the threshold value. This was accomplished by summarizing the marginal (over threshold) posterior distribution on the OR for the association between repeat length and ovarian cancer. The resulting OR is a threshold independent measure of association. Parallel analyses were calculated to determine the relationship between the CAG_L, GGC_S, and GGC_L repeat alleles in the AR gene and ovarian cancer risk.

The stage 1 change point models were fit under the assumption that there was no potential confounding by other covariates. In our stage 2 analysis, we checked this assumption by fitting multivariable unconditional logistic regression models controlling for potential confounders to determine whether confounding bias would explain any observed association between CAG and GGC repeat length and epithelial ovarian cancer. We examined the CAG_S and CAG_L repeat polymorphism variables using the cutoff points identified through the change point analysis as having the highest probabilities. Additionally, because of evidence from prior reports, we present the association between the number of CAG repeats $\geq$22 and ovarian cancer risk among Caucasian study participants. To control for confounding, we simultaneously adjusted for variables with known associations with disease status. These variables included age at diagnosis/interview, tubal ligation (yes or no), months of oral contraceptive use, body mass index (BMI; kg/m^2) 1 year before diagnosis/interview, waist-to-hip ratio, family history of breast or ovarian cancer in first-degree relatives (yes or no), and total months pregnant. We report both age-adjusted ORs and 95% confidence intervals (95% CI) as well as ORs adjusted for additional potential confounders.

An interim analysis after the first 3 years of data collection revealed a statistically significant association with the CAG repeat polymorphism, but no evidence of an association with the GGC polymorphism. We therefore discontinued the analysis of the GGC repeats in this data set. Thus, our final sample size for the CAG polymorphism is approximately twice as large as that for GGC. For much of our sample, we had additional genotype data on 99 unrelated single-nucleotide polymorphisms unlinked to disease status and selected from 22 chromosomes. Among the self-reported African American subjects, we had this genotype data on 77 of 99 cases and 88 of 141 controls. Among the self-reported Caucasian subjects, we had this genotype data on 473 of 495 cases and all of the controls. To address the possibility of population-stratification, the genotype data from these 99 single-nucleotide polymorphisms along with the location of each single-nucleotide polymorphism within its chromosome were input into the program Structure (version 2.0) to estimate the degree of racial admixture for each individual. Structure is a program that implements a model-based clustering method for inferring population structure (26). It allows the user to select the number of parent populations represented in the sample. We set this variable to 2 to allow for African and European ancestral populations. Structure estimated the admixture fractions for each individual. These admixture fractions were also used as an alternate to self-reported race status performing race-specific change point analyses in stage 1. Change point analyses were done using R.[8] All other analyses were done using SAS 9.1 (SAS Institute, Inc., Cary, NC).

Results

The demographic features, epidemiologic risk factors, and pathologic characteristics of cases and controls are shown in Table 1, stratified by self-reported race. Tubal ligation was the only risk factor that had a statistically significant association in both races, in which an inverse relationship is observed. Waist-to-hip ratio was higher in cases than in controls both in Caucasians ($P < 0.001$) and among African Americans ($P = 0.059$). Fewer months of pregnancy and months of oral contraceptive use are observed among cases compared with controls for both races although these differences are statistically significant among Caucasians only. Tumor behavior was invasive in 76% and borderline in 24% for both African American and Caucasian cases. The distribution of histologic subtype was similar in both racial groups.

The CAG repeat length distributions in cases and controls by race are found in Table 2. No differences in the mean CAG repeat length in the AR gene were detected for either the short or long repeat alleles in Caucasian cases and controls. Among African Americans, the mean CAG_S and CAG_L repeat lengths were lower among cases compared with controls. The mean CAG_S length among cases was 16.8 (SD, 2.6) compared with 18.0 (SD, 2.7) among controls ($P = 0.001$), and the mean CAG_L length for cases was 20.7 (SD, 2.9) compared with 21.4 (SD, 2.5) for controls ($P = 0.044$). Both the mean CAG_S and CAG_L lengths were both significantly lower among African Americans compared with Caucasians ($P < 0.001$).

Parallel analyses to determine the association between GGC repeat and ovarian cancer risk were conducted in a subset of the population, which included 186 and 213 self-reported Caucasian ovarian cancer cases and controls, respectively, and 59 and 67 self-reported African American ovarian cancer cases and controls (see Table 2). There was no evidence of differences in the mean GGC_S or GGC_L allele length between cases and controls in either racial group.

Change point analysis was used to determine appropriate cutoff points for CAG repeat length for both the short and long CAG repeat alleles, stratified by self-reported race, either African American or Caucasian. Cutoff points were detected for both CAG_S and CAG_L repeat alleles among African Americans only. Tables 3 and 4 present model probabilities associated with the relationship between ovarian cancer and

[8] http://www.r-project.org

DOI:10.1158/1055-9965.EPI-06-0868

Table 1. Demographics and pathologic characteristics of ovarian cancer cases and controls from the North Carolina Ovarian Cancer Study by self-reported race

	Caucasians			African Americans		
	Cases (N = 495), n (%)	Controls (N = 540), n (%)	P	Cases (N = 99), n (%)	Controls (N = 141), n (%)	P
Age (y)						
20-49	160 (32)	191 (35)	0.664	39 (39)	50 (35)	0.720
50-64	222 (45)	196 (36)		44 (44)	65 (46)	
65-75	113 (23)	153 (28)		16 (16)	26 (18)	
Menopause status						
Pre/peri	180 (36)	225 (42)	0.081	42 (42)	62 (44)	0.775
Post	315 (64)	315 (58)		57 (58)	78 (56)	
Months pregnant						
0	83 (17)	59 (11)	0.004	8 (8)	7 (5)	0.425
1-8	26 (5)	19 (4)		4 (4)	10 (7)	
9-18	182 (37)	206 (38)		36 (36)	48 (34)	
19-36	177 (36)	215 (40)		37 (37)	51 (36)	
>36	26 (5)	40 (7)		14 (14)	25 (18)	
Oral contraceptive use (mo)						
None	165 (33)	161 (30)	0.050	40 (40)	60 (43)	0.134
<12	42 (8)	44 (8)		14 (14)	5 (4)	
12-36	107 (22)	120 (22)		23 (23)	35 (25)	
37-60	45 (9)	46 (9)		3 (3)	6 (4)	
>60	122 (25)	164 (30)		15 (15)	31 (22)	
User of unknown duration	14 (3)	5 (1)		4 (4)	4 (3)	
History of breast/ovarian cancer in first-degree relative						
Yes	89 (18)	87 (16)	0.440	28 (28)	24 (17)	0.037
No	406 (82)	451 (84)		71 (72)	117 (83)	
Tubal ligation						
Yes	122 (25)	164 (30)	0.040	32 (32)	75 (53)	0.001
No	373 (75)	376 (70)		67 (68)	66 (47)	
Polycystic ovarian syndrome						
Yes	2 (0)	4 (1)	0.688	0 (0)	2 (1)	0.513
No	493 (100)	536 (99)		99 (100)	138 (99)	
BMI 1 y before diagnosis/interview						
Quartile 1: <22.42	103 (21)	131 (25)	0.167	NA	NA	
Quartile 2: 22.42-25.739	133 (28)	133 (25)		NA	NA	
Quartile 3: 25.74-29.759	104 (22)	132 (25)		NA	NA	
Quartile 4: >29.76	143 (30)	133 (25)		NA	NA	
BMI 1 y before diagnosis/interview						
Quartile 1: <27.341	NA	NA		23 (24)	34 (25)	0.368
Quartile 2: 27.341-30.33	NA	NA		17 (18)	33 (24)	
Quartile 3: 30.34-36.4	NA	NA		26 (27)	35 (26)	
Quartile 4: >36.4	NA	NA		31 (32)	34 (25)	
Waist-to-hip ratio at interview						
Quartile 1: <0.739	82 (17)	133 (25)	<0.001	NA	NA	
Quartile 2: 0.739-<0.7871	103 (21)	134 (25)		NA	NA	
Quartile 3: 0.7871-<0.8351	152 (31)	134 (25)		NA	NA	
Quartile 4: >0.8351	150 (31)	134 (25)		NA	NA	
Waist-to-hip ratio at interview						
Quartile 1: <0.772	NA	NA		16 (16)	35 (25)	0.059
Quartile 2: 0.772-<0.828	NA	NA		25 (26)	35 (25)	
Quartile 3: 0.829-<0.876	NA	NA		20 (21)	35 (25)	
Quartile 4: >0.876	NA	NA		36 (37)	34 (24)	
Infertility, doctor diagnosed in female						
Yes	62 (13)	53 (10)	0.166	8 (8)	10 (7)	0.775
No	433 (87)	487 (90)		91 (92)	131 (93)	
Tumor behavior						
Borderline	117 (24)			24 (24)		
Invasive	378	(76)		75 (76)		
Tumor histology						
Serous	300 (61)				61 (62)	
Endometrioid	63 (13)				12 (12)	
Mucinous	48 (10)				11 (11)	
Clear cell	37 (7)				2 (2)	
Other	46 (9)			13 (13)		

CAG_S repeat length allele and CAG_L repeat allele, respectively, conditional on thresholded CAG length among African Americans and Caucasians. The tables present estimates of posterior model probabilities for each cutoff point. OR estimates and 95% CIs for the association between CAG repeat length allele and ovarian cancer for each cutoff point are also shown. Among African Americans, the posterior probability of a change point association in the *AR* CAG_S repeat allele is ~72% and the most likely threshold is between 15 and 16 with a posterior probability of 31% given that the change point class of models is correct. The OR for the association between CAG repeat length at the threshold between 15 and 16 is 2.77 (95% CI, 1.31-5.26). The posterior probability of a change point association in the *AR* CAG_L repeat allele is ~73% and the most likely threshold is between 18 and 19 with a posterior probability of 21%. In contrast, the data for Caucasian ovarian

Table 2. Mean and median CAG and GGC trinucleotide repeat length polymorphism lengths in ovarian cancer cases and controls enrolled in the North Carolina Ovarian Cancer study by self-reported race

	Caucasians			African Americans		
	Cases	Controls	*P*	Cases	Controls	*P*
CAG repeats	(*N* = 484)	(*N* = 522)		(*N* = 99)	(*N* = 140)	
CAG_S						
Mean (SD)	19.4 (2.3)	19.3 (2.2)	0.685	16.8 (2.6)	18.0 (2.7)	0.001
Median (range)	19.0 (6-25)	19.0 (5-25)		17.0 (8-23)	17.0 (10-26)	
CAG_L						
Mean (SD)	22.6 (2.6)	22.4 (2.5)	0.146	20.7 (2.9)	21.4 (2.5)	0.044
Median (range)	23.0 (15-34)	22.0 (15-32)		21.0 (15-29)	22.0 (14-27)	
GGC repeats	(*N* = 186)	(*N* = 213)		(*N* = 59)	(*N* = 67)	
GGC_S						
Mean (SD)	16.6 (1.53)	16.4 (1.86)	0.377	15.4 (1.74)	15.2 (1.77)	0.593
Median (range)	17.0 (6-18)	17.0 (6-18)		16.0 (9-18)	16.0 (9-18)	
GGC_L						
Mean (SD)	17.2 (0.89)	17.2 (1.04)	0.800	16.6 (0.83)	16.6 (0.97)	0.925
Median (range)	17.0 (14-19)	17.0 (12-20)		17.0 (14-18)	17.0 (14-18)	

NOTE: *P* values are from Student's *t* test.

cancer cases and controls do not support a change point model; for the *A*R CAG_S repeat allele, the posterior probability of this class of change point models is 31% and no threshold has a posterior probability exceeding 14%. For the CAG_L allele, it is 36% and no threshold has a posterior probability exceeding 13%.

We repeated the change point analysis in a subset of cases and controls defined as African Americans based on having admixture fraction <10% as determined from the Structure analysis. A total of 13 cases and 21 controls of the self-reported African Americans who had an admixture fraction ≤90% were omitted. An additional 22 cases and 52 controls of the self-reported African Americans were omitted from this analysis due to missing admixture information. In this analysis, the relationship between CAG_S length and ovarian cancer became even stronger, suggesting that admixture does not explain the observed association with ovarian cancer (data not shown). In this analysis, the posterior of a change point association was 0.871 and, given the association, the probability of a change point between 15 and 16 was 0.675. For the CAG_L, there was no evidence of an admixture effect. Similarly, we repeated the change point analysis among Caucasians, omitting those with an admixture of <10%, and found that the posterior probability of a threshold-based effect of the CAG_S allele on risk was 0.337, which was not different from the probability of 0.314 that estimated when using self-reported race to identify Caucasians.

Change point analysis did not detect an association between ovarian cancer and GGC repeat length for either the GGC_S or GGC_L repeats among those whose self-reported race was African American, those whose self-reported race was Caucasian, or those limited to having an African American admixture fraction of >90%. In fact, the probability for a change point did not exceed 41% for either the GGC_S or GGC_L allele in any of these groups (data not shown). Given a prior probability of 50% in favor of such an association, this is evidence against association.

Additional multivariable analyses to determine whether confounding could explain the association between CAG repeat length and ovarian cancer among African Americans are presented in Table 5. The age-adjusted OR for the association between the CAG_S repeat length allele <16 and ovarian cancer was 2.8 (95% CI, 1.4-5.9) in African Americans. A similar relationship was found between the CAG_L repeat length allele <19 and ovarian cancer (age-adjusted OR, 2.5; 95% CI, 1.3-4.8). Having both a CAG_S repeat <16 and a CAG_L repeat <19 was associated with a 5-fold increased risk of ovarian cancer (age-adjusted OR, 5.4; 95% CI, 1.6-17.9). Also

Table 3. Model probabilities conditional on thresholded CAG_S length and ORs for CAG_S repeat length less than versus greater than a threshold *t* by self-reported race

CAG_S length	Caucasians (484 cases and 522 controls)		African Americans (99 cases and 140 controls)	
	Posterior Pr(*T* = *t*/change)	OR (95% CI)	Posterior Pr(*T* = *t*/change)	OR(95% CI)
7.5	0.110	1.62 (0.19-6. 30)	NA	NA
9.5	NA	NA	0.029	4.26 (0.34-19.54)
10.5	0.098	1.80 (0.35-5.80)	NA	NA
11.5	0.099	1.91 (0.48-5.44)	0.029	2.91 (0.60-9.26)
12.5	0.106	1.96 (0.59-5.11)	0.067	3.98 (0.91-12.45)
13.5	0.134	2.03 (0.74-4.64)	0.072	3.05 (1.01-7.52)
14.5	0.053	1.26 (0.55-2.48)	0.029	2.12 (0.83-4.54)
15.5	0.048	0.88 (0.44-1.57)	0.306	2.77 (1.31-5.26)
16.5	0.039	0.91 (0.51-1.48)	0.053	1.84 (1.04-3.00)
17.5	0.030	0.89 (0.63-1.21)	0.045	1.79 (1.02-2.92)
18.5	0.017	1.02 (0.77-1.31)	0.156	2.15 (1.18-3.65)
19.5	0.019	0.94 (0.73-1.20)	0.049	2.02 (1.02-3.70)
20.5	0.025	0.90 (0.67-1.19)	0.030	2.09 (0.88-4.35)
21.5	0.025	0.93 (0.66-1.27)	0.025	2.28 (0.75-5.80)
22.5	0.047	0.82 (0.52-1.23)	0.041	3.74 (0.77-13.13)
23.5	0.070	1.50 (0.63-3.09)	0.024	3.68 (0.39-16.43)
24.5	0.081	1.40 (0.33-4.10)	0.024	3.66 (0.40-16.35)
25.5	NA	NA	0.020	2.15 (0.17-10.03)
Overall	0.314		0.724	

DOI:10.1158/1055-9965.EPI-06-0868

Table 4. Model probabilities conditional on thresholded CAG_L length and ORs for CAG_L repeat length less than versus greater than a threshold *t* by self-reported race

CAG length	Caucasians (484 cases and 522 controls)		African Americans (99 cases and 140 controls)	
	Posterior Pr($T = t$/change)	OR (95% CI)	Posterior Pr($T = t$/change)	OR (95% CI)
14.5	NA	NA	0.021	1.41 (0.10-5.93)
15.5	0.078	1.62 (0.18-6.28)	0.020	2.14 (0.24-8.35)
16.5	0.067	1.44 (0.24-4.80)	0.015	1.73 (0.43-4.79)
17.5	0.082	0.67 (0.17-1.71)	0.022	1.85 (0.81-3.69)
18.5	0.067	0.69 (0.33-1.26)	0.205	2.44 (1.24-4.41)
19.5	0.025	0.87 (0.57-1.29)	0.204	2.24 (1.22-3.80)
20.5	0.017	0.91 (0.68-1.20)	0.150	2.05 (1.18-3.34)
21.5	0.032	0.84 (0.65-1.08)	0.180	2.08 (1.20-3.39)
22.5	0.036	0.83 (0.65-1.06)	0.008	1.32 (0.72-2.23)
23.5	0.048	0.81 (0.62-1.04)	0.007	1.18 (0.61-2.12)
24.5	0.015	1.06 (0.77-1.43)	0.010	1.39 (0.56-2.99)
25.5	0.021	0.92 (0.61-1.34)	0.011	1.02 (0.34-2.45)
26.5	0.024	0.99 (0.58-1.58)	0.034	0.53 (0.08-1.74)
27.5	0.060	0.72 (0.36-1.27)	0.070	0.35 (0.03-1.27)
28.5	0.065	0.71 (0.27-1.49)	NA	NA
29.5	0.053	1.30 (0.35-3.48)	NA	NA
30.5	0.082	0.93 (0.13-3.24)	NA	NA
32.5	0.127	0.62 (0.05-2.35)	NA	NA
33.5	0.102	0.93 (0.07-3.84)	NA	NA
Overall	0.362		0.734	

shown in Table 5, simultaneously controlling for age, months pregnant, months of oral contraceptive use, BMI, family history of ovarian or breast cancers in a first-degree relative, and tubal ligation did not substantially change the relationship between CAG repeat length and ovarian cancer detected in the age-adjusted analyses. Additional analyses limited to invasive ovarian cancers as well as histologic subtype (serous, endometrial, and clear cell only) did not reveal any substantial differences in the relationship with CAG repeat length and ovarian cancers (data not shown). Although we did not detect evidence for a threshold in CAG repeat length in Caucasians, we calculated the age-adjusted OR for a CAG_S repeat length <16 of 0.8 (95% CI, 0.4-1.5) and for the CAG_L repeat length <19 of 0.6 (95% CI, 0.3-1.2).

We also conducted unconditional logistic regression analyses in Caucasian subjects using a cutoff point of ≥22 CAG repeats to compare our data to those of previously published reports (20, 21, 27). The age-adjusted ORs for the association between those who carry either one or two alleles with ≥22 CAG repeats versus those with two alleles with <22 repeats were 1.2 (95% CI, 0.9-1.6) and 1.2 (95% CI 0.8-1.7), respectively (see Table 6). Restricting this analysis to Caucasian women of <10% admixture, an age-adjusted OR of 1.3 (95% CI, 0.9-2.0) was calculated for the association with two alleles with ≥22 repeats.

Table 5. Relationship between AR CAG repeat polymorphisms and ovarian cancer among African American women enrolled in the North Carolina Ovarian Cancer Study

	Cases, *n* (%)	Controls, *n* (%)	OR* (95% CI)	OR† (95% CI)
CAG_S repeat <16				
No	76 (77)	126 (89)	1.0 (reference)	1.0 (reference)
Yes	23 (23)	15 (11)	2.8 (1.4-5.9)	2.5 (1.1-5.5)
CAG_L repeat <19				
No	72 (73)	121 (86)	1.0 (reference)	1.0 (reference)
Yes	27 (27)	20 (14)	2.5 (1.3-4.8)	2.7 (1.3-5.8)
No. with CAG_S <16 or CAG_L <19				
None	60 (61)	111 (79)	1.0 (reference)	1.0 (reference)
1 allele	28 (28)	25 (18)	2.2 (1.1-4.1)	2.1 (1.1-4.3)
2 alleles	11 (11)	5 (4)	5.4 (1.6-17.9)	4.8 (1.4-17.0)

*Age adjusted.
†Adjusted for age, months pregnant, months of oral contraceptive use, BMI, tubal ligation, family history of breast or ovarian cancer in a first-degree relative, waist-to-hip ratio.

Discussion

The mean and median AR CAG lengths for both the CAG_S and CAG_L alleles in Caucasian subjects in the North Carolina Ovarian Cancer study population are similar to lengths reported in previous studies (20, 21, 27). No relationship was found between CAG repeat length and ovarian cancer among the Caucasians in this study. To our knowledge, this is the first study to evaluate the association between CAG repeat length in *AR* and ovarian cancer risk in African American women. We found an increase in ovarian cancer risk associated with both CAG_S and CAG_L repeat length alleles in African Americans. These differences were evidenced by both the shorter mean repeat length of the CAG_S and CAG_L alleles as well as the higher prevalence of the CAG_S repeat length <16 and the CAG_L repeat length <19 among African American cases compared with controls.

The association between *AR* CAG repeat length and ovarian cancer risk in African Americans is further supported by an analysis of these data that omitted self-reported African American subjects with evidence of significant admixture of

Table 6. AR CAG repeat polymorphisms ≥22 versus ≤22 repeats among Caucasian cases and controls in the North Carolina Ovarian Cancer study

	Cases, *n* (%)	Controls, *n* (%)	OR* (95% CI)	OR† (95% CI)
CAG repeat ≥22				
0 allele	163 (34)	198 (38)	1.0 (reference)	1.0 (reference)
1 allele	237 (49)	240 (46)	1.2 (0.9-1.6)	1.2 (0.9-1.6)
2 alleles	84 (17)	84 (16)	1.2 (0.8-1.7)	1.2 (0.8-1.8)
Either 1 or 2 alleles	321 (66)	324 (62)	1.2 (0.9-1.5)	1.2 (0.9-1.5)

*Age adjusted.
†Adjusted for age, months pregnant, months of oral contraceptive use duration, BMI, tubal ligation, family history of breast or ovarian cancer in a first-degree relative, waist-to-hip ratio.

>10% and found a similar association. Therefore, this result is not likely to be explained by confounding due to population stratification. Additionally, our results remained significant when simultaneously controlling for other potential confounders. Our results suggest that having one short CAG repeat length in African Americans, which is associated with higher levels of androgenic activity, more than doubles the risk of ovarian cancer. Additionally, those with both short CAG_S and short CAG_L repeats may have a 5-fold increased risk of ovarian cancer. This is also the largest study to evaluate the relationship between the *AR* GGC repeat length polymorphism and ovarian cancer. Similar to the finding of a small case-control study by Kadouri et al. (12), we did not detect a relationship in either racial group.

The observed association between CAG repeat length and ovarian cancer is biologically plausible in view of the inverse relationship between CAG length and transactivation activity on the receipt and binding affinity of androgens (14). Short CAG repeat alleles may facilitate greater chronic androgen stimulation leading to increased proliferative activity. Shorter CAG repeat alleles have also been associated with other hyperandrogenic clinical conditions including risk of baldness and having prostatic hyperplasia in men and hirsutism (28), annovulation (29), and acne in women (30). Additionally, data from a nested case-control study by Helzlsouer et al. (11) found that increased serum androgen levels were associated with an increased risk of ovarian cancer.

Change point analysis of our data does not support a choice of a threshold. However, two recent studies by Terry et al. (20) and Santarosa et al. (21) support an association between having two alleles with ≥22 CAG repeats and ovarian cancer risk in Caucasian subjects. The relationship in the study by Santarosa et al. was stronger, with an OR of 3.45 (95% CI, 1.42-8.34) compared with 1.31 (95% CI, 1.01-1.59) in the study by Terry et al. (20). For comparative purposes, we did analyses among Caucasians using the cutoff point reported in these studies and found an OR of 1.2 (95% CI, 0.9-1.6) for self-reported Caucasian and 1.3 (95% CI, 0.9-2.0) among those with <10% admixture. These ORs, although weak, are similar in magnitude and precision as that reported by Terry et al. (20). As suggested by Terry et al., a possible explanation for the discrepancies between previously published reports that may also explain the findings in the current study includes differences in the prevalence in the carriage of subjects having two AR CAG repeats ≥22. It is known that allele frequency varies according to ethnicity (16, 31). In our study, we found that the prevalence of two CAG repeats ≥22 differed markedly between Caucasian and African American controls, 16% and 9%, respectively. The prevalence of two AR CAG repeats >22 in our Caucasian subjects also differs from the prevalences among Caucasians in the studies by Terry (prevalence, 24%) and Spurdle et al. (prevalence, 26%), a positive and a negative study, respectively, but is more similar to that of Santarosa et al. (prevalence, 18%), which was a positive study. Due to the known ethnic variation in CAG repeat length, it is possible that the results of the association with *AR* CAG length could be due to chance. However, our analyses of admixture among the Caucasian subjects did not support that population stratification was a major concern; only 8% of cases and 11% of controls had evidence of significant (>10%) admixture, and when we reanalyzed our data, omitting subjects with >10% admixture, our findings did not change.

Strengths of this study include the fact that this is a large population-based study of both Caucasian and African American women. Our approach using the change point analysis provided a more objective and thorough evaluation of a cutoff point in the association between *AR* CAG repeat length and ovarian cancer risk, avoiding multiple comparisons at different thresholds. The nature of the change point analyses sidesteps the issue of multiple comparisons. In this approach, we simultaneously calculate (*a*) the posterior probability of a change point association versus the alternative of no association and (*b*) the posterior probability of each of a discrete set of thresholds being the location of the change point given that there is one. This decouples the question of whether or not there is a change point association from the question of the appropriate threshold. In addition, we estimated the association between *AR* CAG repeat length while simultaneously controlling for other potential confounders, thus providing evidence that confounding bias is unlikely to account for the association. We were able to determine that it was unlikely that population stratification among African Americans biased our results. It is also unlikely that selection bias related to genotype would have occurred and influenced our results. Limitations of our study include a somewhat small sample of African American subjects. We attempted to find an independent data set that could be used for a validation of the association between *AR* CAG repeat length and ovarian cancer in African Americans, but we were unsuccessful. Finally, we were not able to conclusively determine why the findings among African American women and Caucasian women differed. The failure to observe the relationship in Caucasians may be due to the rarity of the short CAG alleles in this population or could reflect racial differences in disease etiology. For example, compared with Caucasian women, differences in the prevalence of other genetic variants and other characteristics, such as BMI and waist-to-hip ratio, among African American women may play a role. Further exploration of these factors may help increase the understanding of ovarian cancer etiology.

Similar to studies in ovarian cancer, analyses of the relationship between the short *AR* CAG repeat length polymorphism and prostate cancer risk also have yielded conflicting results. Likewise, differences in the association with prostate have been noted between racial groups (17, 32-35). Pettaway (32) has suggested that racial differences in genetic variation in several genes in the androgen/AR pathway may be related to clinically observed differences in the biology of prostate cancer among racial groups (32). For example, in addition to CAG repeat length in *AR*, genetic variants in the 5α-reductase type 2 also differ between African Americans and Caucasians. However, it has also been suggested that racial differences and inconsistent findings in studies of prostate cancer may be due to linkage disequilibrium between *AR* CAG repeat length polymorphisms and another susceptibility locus on the X chromosome (36). These possible explanations are also relevant to studies of ovarian cancer.

In summary, our finding of an association between short *AR* CAG repeat lengths and ovarian cancer among African Americans warrants replication in a larger data set, and further study is needed to more fully understand the complexities of this relationship. We did not detect a relationship between CAG repeat length and ovarian cancer among Caucasian women and were not able to confirm previous reports for such an association. Additionally, we were unable to detect a relationship between the GGC repeat polymorphism and ovarian cancer in either African American or Caucasian women. However, we believe that further study of the positive finding in African American women may provide insight into the etiology of ovarian cancer.

References

1. Risch HA. Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. J Natl Cancer Inst 1998; 90:1774–86.
2. Modugno F. Ovarian cancer and polymorphisms in the androgen and progesterone receptor genes:a HuGE review. Am J Epidemiol 2004;159: 319–35.
3. Lau KM, Mok SC, Ho SM. Expression of human estrogen receptor-α and -β, progesterone receptor, and androgen receptor mRNA in normal and malignant ovarian epithelial cells. Proc Natl Acad Sci U S A 1999;96: 5722–7.

DOI:10.1158/1055-9965.EPI-06-0868

4. Kuhnel R, de Graaff J, Rao BR, Stolk JG. Androgen receptor predominance in human ovarian carcinoma. J Steroid Biochem 1987;26:393–7.
5. Chadha S, Rao BR, Slotman BJ, van Vroonhoven CC, van der Kwast TH. An immunohistochemical evaluation of androgen and progesterone receptors in ovarian tumors. Hum Pathol 1993;24:90–5.
6. Slotman BJ, Baak JP, Rao BR. Correlation between nuclear DNA content and steroid receptor status in ovarian cancer. Eur J Obstet Gynecol Reprod Biol 1991;38:221–7.
7. Edmondson RJ, Monaghan JM. The epidemiology of ovarian cancer. Int J Gynecol Cancer 2001;11:423–9.
8. Hoyo C, Berchuck A, Halabi S, et al. Anthropometric measurements and epithelial ovarian cancer risk in African American and White women. Cancer Causes Control 2005;16:955–63.
9. Mink PJ, Folsom AR, Sellers TA, Kushi LH. Physical activity, waist-to-hip ratio, and other risk factors for ovarian cancer:a follow-up study of older women. Epidemiology 1996;7:38–45.
10. Schildkraut JM, Schwingl P, Bastos E, Evanoff A, Hughes C. Epithelial ovarian cancer risk among women with polycystic ovarian disease. Obstet Gynecol 1996;88:554–9.
11. Helzlsouer KJ, Alberg AJ, Gordon GB, et al. Serum gonadotropins and steroid hormones and the development of ovarian cancer. JAMA 1995;274:1926–30.
12. Kadouri L, Easton DF, Edwards S, et al. CAG and GGC repeat polymorphisms in the androgen receptor gene and breast cancer susceptibility in BRCA1/2 carriers and non-carriers. Br J Cancer 2001;85:36–40.
13. Sleddens HF, Oostra BA, Brinkmann AO, Trapman J. Trinucleotide repeat polymorphism in the androgen receptor gene (AR). Nucleic Acids Res 1992;20:1427.
14. Chamberlain NL, Driver ED, Miesfeld RL. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res 1994;22:3181–6.
15. Nelson KA, Witte JS. Androgen receptor CAG repeats and prostate cancer. Am J Epidemiol 2002;155:883–90.
16. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 1992;12:241–53.
17. Platz EA, Rimm EB, Willett WC, Kantoff PW, Giovannucci E. Racial variation in prostate cancer incidence and in hormonal system markers among male health professionals. J Natl Cancer Inst 2000;92:2009–17.
18. Sleddens HF, Oostra BA, Brinkmann AO, Trapman J. Trinucleotide (GGN) repeat polymorphism in the human androgen receptor (AR) gene. Hum Mol Genet 1993;2:493.
19. Ding Y, Tong M, Liu S, Moscow JA, Tai HH. NAD+-linked 15-hydroxyprostaglandin dehydrogenase (15-PGDH) behaves as a tumor suppressor in lung cancer. Carcinogenesis 2005;26:65–72.
20. Terry KL, De Vivo I, Titus-Ernstoff L, Sluss PM, Cramer DW. Genetic variation in the progesterone receptor gene and ovarian cancer risk. Am J Epidemiol 2005;161:442–51.
21. Santarosa M, Bidoli E, Gallo A, Steffan A, Boiocchi M, Viel A. Polymorphic CAG repeat length within the androgen receptor gene:identification of a subgroup of patients with increased risk of ovarian cancer. Oncol Rep 2002;9:639–44.
22. Spurdle AB, Chen X, Abbazadegan M, et al. CYP17 promotor polymorphism and ovarian cancer risk. Int J Cancer 2000;86:436–9.
23. Menin C, Banna GL, De Salvo G, et al. Lack of association between androgen receptor CAG polymorphism and familial breast/ovarian cancer. Cancer Lett 2001;168:31–6.
24. Irvine RA, Yu MC, Ross RK, Coetzee GA. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res 1995;55:1937–40.
25. Barry D, Hartigan JA. Bayesian Analysis for Change Point Problems. J Am Stat Assoc 1993;88:309–19.
26. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–59.
27. Spurdle AB, Webb PM, Chen X, et al. Androgen receptor exon 1 CAG repeat length and risk of ovarian cancer. Int J Cancer 2000;87:637–43.
28. Franks S. Polycystic ovary syndrome. N Engl J Med 1995;333:853–61.
29. Mifsud A, Ramirez S, Yong EL. Androgen receptor gene CAG trinucleotide repeats in anovulatory infertility and polycystic ovaries. J Clin Endocrinol Metab 2000;85:3484–8.
30. Sawaya ME, Shalita AR. Androgen receptor polymorphisms (CAG repeat lengths) in androgenetic alopecia, hirsutism, and acne. J Cutan Med Surg 1998;3:9–15.
31. Buchanan G, Yang M, Cheong A, et al. Structural and functional consequences of glutamine tract variation in the androgen receptor. Hum Mol Genet 2004;13:1677–92.
32. Pettaway CA. Racial differences in the androgen/androgen receptor pathway in prostate cancer. J Natl Med Assoc 1999;91:653–60.
33. Hsing AW, Gao YT, Wu G, et al. 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 2000;60:5111–6.
34. Gsur A, Preyer M, Haidinger G, et al. Polymorphic CAG repeats in the androgen receptor gene, prostate-specific antigen polymorphism and prostate cancer risk. Carcinogenesis 2002;23:1647–51.
35. Balic I, Graham ST, Troyer DA, et al. Androgen receptor length polymorphism associated with prostate cancer risk in Hispanic men. J Urol 2002;168:2245–8.
36. Li C, Gronberg H, Matsuyama H, et al. Difference between Swedish and Japanese men in the association between AR CAG repeats and prostate cancer suggesting a susceptibility-modifying locus overlapping the androgen receptor gene. Int J Mol Med 2003;11:529–33.