

# Polymorphic CAG and GGN Repeat Lengths in the Androgen Receptor Gene and Prostate Cancer Risk: A Population-based Case-Control Study in China

Ann W. Hsing,<sup>1</sup> Yu-Tang Gao, Guan Wu, Xin Wang, Jie Deng, Yuh-Ling Chen, Isabell A. Sesterhenn, F. Kash Mostofi, Jacques Benichou, and Chawnshang Chang

Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20852-7234 [A. W. H.]; Shanghai Cancer Institute, Shanghai, China 200032 [Y. T. G., J. D.]; University of Rochester, Rochester, New York, 14642 [G. W., X. W., Y. L. C., C. C.]; Armed Forces Institute of Pathology, Washington, D.C. 20306 [I. A. S., F. K. M.]; and University of Rouen, 76031 Rouen, France [J. B.]

## ABSTRACT

The length of the polymorphic CAG trinucleotide repeat in the polyglutamine region of the androgen receptor (AR) gene is inversely correlated with the transactivation function of the AR. Because increased androgenic activity has been linked to prostate cancer and because an ethnic variation exists in the CAG repeat length, this polymorphism has been suggested to explain part of the substantial racial difference in prostate cancer risk. We conducted a population-based case-control study in China to investigate whether CAG and other polymorphisms of the AR gene are associated with clinically significant prostate cancer in this low-risk population. Genomic DNA from 190 prostate cancer patients and 304 healthy controls was used for direct sequencing to evaluate the relationship of CAG and GGN (polyglycine) repeat length in the AR gene. Relative to western men, our study subjects had a longer CAG repeat length, with a median of 23 and only 10% of the subjects having a CAG repeat length shorter than 20. Men with a CAG repeat length shorter than 23 (median length) had a 65% increased risk of prostate cancer (odds ratio, 1.65; 95% confidence interval, 1.14–2.39), compared with men with a CAG repeat length of 23 or longer. For the GGN tract (GGT<sub>3</sub>GGG<sub>1</sub>GGT<sub>2</sub>GGC<sub>n</sub>), based on the sequencing results from 481 samples, we are the first to show that although GGC regions in the polyglycine tract are highly variable, there are no mutations or polymorphisms in the GGT and GGG regions. More than 72% of the subjects had a GGN repeat length of 23, and those with a GGN repeat length shorter than 23 had a 12% increased risk of prostate cancer (95% confidence interval, 0.71–1.78), compared with those with  $\geq 23$  GGN repeats. Our study not only confirms that Chinese men do have a longer CAG repeat length than western men but also represents the first population-based study to show that even in a very low-risk population, a shorter CAG repeat length confers a higher risk of clinically significant prostate cancer. These results imply that CAG repeat length can potentially serve as a useful marker to identify a subset of individuals at higher risk of developing clinically significant prostate cancer. Larger studies are needed to evaluate the combined effect of CAG and GGN repeats. Because of the significance of AR in prostate cancer, investigation of factors that interact with the polyglutamine region of the AR gene to alter AR function and modulate prostate cancer risk is an important area for future research.

## INTRODUCTION

The incidence of clinical prostate cancer differs substantially between ethnic groups, with African Americans having a 10- to 40-fold higher incidence than Asians (1–3). Such disparity in incidence of clinical prostate cancer cannot be explained entirely by population differences in screening. An earlier study shows that after adjustment for screening, there is still a 3- to 4-fold difference in incidence rates between United States and Japanese men, whose rates are among the highest in Asians (4). Despite the dramatic racial variation in clinical

prostate cancer incidence, the prevalence of latent carcinoma appears to be similar across populations (5), suggesting that there exist differences in factors (either genetic or environmental) that promote the progression of microscopic tumors to clinically overt carcinoma.

The growth, differentiation, and proliferation of prostatic cells are regulated by androgens (6). The biological effects of androgens are mediated through binding to the intracellular AR,<sup>2</sup> which in turn regulates the transcription of target genes with the assistance of transcriptional coactivators (7). The AR protein, consisting of 918 amino acids and encoded singly by the AR gene located on the X chromosome (Xq11-12), has three major functional domains: a transactivating amino-terminal domain, a DNA binding domain, and a ligand (steroid) binding domain (8). The open reading frame of the AR gene is separated over eight exons and has a length of 2730 bp. The sequence encoding the large amino-terminal transactivating domain is found in the first exon, the DNA binding domain is encoded by exons 2 and 3, and the information for the ligand binding domain is distributed over exons 4–8 (8).

The first exon of the AR gene contains two polymorphic trinucleotide repeat segments that encode polyglutamine and polyglycine tracts localized in the NH<sub>2</sub>-terminal transactivation domain of the AR protein. The polyglutamine tract is encoded by a CAG trinucleotide repeat, and the polyglycine stretch is encoded by a GGN repeat. The number of CAG repeats ranges from about 8 to 35 repeats in normal individuals. Longer CAG repeat lengths appear to result in reduced AR transcriptional activity both *in vivo* and *in vitro* (9, 10). Otherwise healthy men whose AR has a CAG repeat length at the long end of the normal range (>28) have an increased incidence of impaired spermatogenesis and infertility (11), conditions that are extremely androgen-dependent (12). Expansion of the CAG repeat length to >40 repeats is related to a rare neuromuscular disorder, spinal and bulbar muscular atrophy (Kennedy syndrome), which is also associated with androgen insensitivity, decreased virilization, testicular atrophy, reduced sperm production, and infertility (13–15). Together, these clinical data suggest that a longer CAG repeat length decreases the functional competence of AR.

The length of the polyglycine (GGN) tract varies from about 10 to 30 repeats. The functional consequences of variation in the GGN tract are less clear. Deletion of the polyglycine tract reduces AR transcriptional activity by ~30% in transient transfection assays (16), although there is no significant correlation between polyglycine tract length and infertility (11).

Shorter AR polyglutamine tracts, and thus a more transcriptionally active AR, have been associated with increased prostate cancer risk (17–22), higher cancer grade at diagnosis (23), earlier age of cancer onset in white men (24, 25), and aggressive early-stage prostate cancer (defined as clinically unsuspected metastatic disease in men undergoing radical prostatectomy) (22). In addition, several epidemiological studies have shown that a shorter length of both CAG and GGN repeats confers a higher risk of prostate cancer (17, 20, 22).

<sup>2</sup> The abbreviations used are: AR, androgen receptor; CI, confidence interval; OR, odds ratio; BPH, benign prostatic hyperplasia.

Received 2/7/00; accepted 7/19/00.

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.

<sup>1</sup> To whom requests for reprints should be addressed, at Division of Cancer Epidemiology and Genetics, National Cancer Institute, EPS-MSC 7234, 6120 Executive Boulevard, Bethesda, MD 20852-7234. Fax: (301) 402-0916; E-mail: hsinga@exchange.nih.gov.

Previous studies have shown that the CAG repeat length is shortest in African Americans, intermediate in whites, and longest in Asians, which corresponds well with the high, intermediate, and low incidence of prostate cancer in these populations (26, 27). Because of the ethnic variation in CAG and GGN repeat lengths of the *AR* gene and the role of *AR* in androgenic activity, it has been suggested that these polymorphisms may help explain part of the large racial difference in prostate cancer risk. However, to date, data supporting the relationship between *AR* polymorphisms and prostate cancer came exclusively from white men, and presently there are no data on Asians or African Americans. To assess the importance of *AR* polymorphisms in prostate cancer, as part of our population-based case-control study conducted in China, we herein examine the polymorphic length of CAG and GGN repeats in relation to prostate cancer risk.

## MATERIALS AND METHODS

**Study Subjects.** Details of the study have been described previously.<sup>3</sup> Briefly, cases of primary prostate cancer (International Classification of Diseases 9, Code 185) newly diagnosed between 1993 and 1995 were identified through a rapid reporting system established between the Shanghai Cancer Institute and 28 collaborating hospitals in urban Shanghai. Cases were permanent residents in 10 urban districts of Shanghai (henceforth referred to as Shanghai) who had no history of any other cancer. Of the 268 eligible cases (representing 95% of the cases diagnosed in urban Shanghai during the study period), 243 (91%) were interviewed in person by trained interviewers. Four of the cases were later classified as having BPH and excluded from the study after a consensus review by both Chinese and United States pathologists.

Based on the records at the Shanghai Resident Registry, which contains personal identification cards for all adult residents over age 18 in urban Shanghai, healthy subjects who were free of cancer were selected randomly from permanent residents of Shanghai (6.5 million) and frequency-matched to the expected age distribution (5-year category) of prostate cancer cases. Of the 495 eligible controls without a history of cancer, 472 (95%) were interviewed.

Information on potential risk factors was elicited through an in-person interview by trained interviewers using a structured questionnaire. The interview included information on demographic characteristics; dietary history; smoking history; consumption of alcohol and other beverages; medical history; family history of cancer; physical activity; body size; and sexual behavior.

**Blood Collection and DNA Extraction.** Two hundred cases (84% of those interviewed) and 330 controls (70%) provided 20 ml of fasting blood for the study. The blood samples were processed within 3 h of collection at a central laboratory in Shanghai and stored at  $-70^{\circ}\text{C}$ . The frozen buffy coat samples (separated from 5 ml of blood) were later shipped to the United States on dry ice for DNA extraction at the American Type Culture Collection (Manassas, Virginia) with standard protocols. DNA purity, yield, and length were satisfactory, and there was no evidence of DNA degradation or RNA contamination. After DNA extraction, 191 cases and 304 controls had sufficient DNA for *AR* genotyping at the University of Rochester. DNA samples for cases and controls were grouped into pairs to minimize the effect of day-to-day laboratory variation. Laboratory personnel were blinded to the case-control status.

**Molecular Analysis and Assessment of the CAG and GGN Repeats.** As part of an ongoing molecular analysis of the *AR* gene, genomic DNA from the 495 subjects was used to determine the usual sense codon sequence and the exact number of CAG and GGN repeats in exon 1 of the *AR* gene through PCR amplification and end-labeled sequencing. For the CAG repeat analysis, we designed a set of oligonucleotide primers that flank the CAG repeat, 5'-GCTCTGG-GACGCAA-CCTCTCT-3' and 5'-GCAGCGACTACCGCATCATCA-3', for PCR amplification. We selected a pair of nested primers, 5'-CGGG-GTA-AGGGAAGTAGGTGGAAG-3' and 5'-CTCTACGATGGGCTTGGGGAG-AAC-3', for DNA sequencing. For GGN analysis, we used the oligonucleotide primers 5'-ACCCTCAGCCGCCGCTTCCTCATC-3' and 5'-CTGGGAT-AGGGCACTCTGCTCAAC-3' for both PCR amplification and sequencing.

The PCR products of the CAG and GGN repeats were amplified, using the Advantage 2 Polymerase System (Clontech) and the Advantage-GC Genomic Polymerase System (Clontech), respectively. Subsequently, these PCR products were purified, using the PCR Product Purification Kit (Qiagen), and sequenced directly, using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems). All reactions were optimized to reach consistent results, using genomic DNA samples extracted from cell lines. For the polyglutamine tract ((CAG)<sub>n</sub>CAA), the number of CAG triplets was counted to yield the length of CAG repeats. For the polyglycine tract (GGT<sub>3</sub>GGG<sub>1</sub>GGT<sub>2</sub>GGC<sub>n</sub>), the usual sense codon sequence of the GGN tract is: three GGT, one GGG, two GGT, followed by a variable number of GGC repeats. For example, a GGN repeat length of 23 in our study corresponded to a PCR fragment of 217 bp, encompassing 3 GGT, 1 GGG, 2 GGT, and 17 GGC triplets.

Because the PCR procedure is prone to contamination, a negative control (water blank) was always included in each batch of PCR reactions (usually 9–18 samples plus one negative control). The assay for one batch (9 samples) was repeated with new reagents because of an indication of minor contamination. Because exon 1 of the *AR* gene is GC-rich with CAG and GGN repeats, this region is difficult to amplify. Several samples had to be amplified and sequenced more than once. Overall, 5 (1%) of the 495 samples could not be typed for CAG repeats because of insufficient DNA or sequencing problems, whereas 14 (2.8%) could not be typed for GGN repeats for similar reasons. The percentages of samples that were unsuccessfully genotyped were similar in cases and controls.

Twenty-four split samples from the same individual were included as quality control samples to assess the reproducibility of genotyping. Of the 24 quality control samples, 23 and 20 were amplified and sequenced successfully for the CAG and GGN repeats, respectively. Of the 23 samples with CAG results, 21 (91%) had the same repeat length of 23, one had 24, and one had 22. Of the 20 samples with GGN results, 19 (95%) had the same repeat length of 23 and the other had a length of 24.

**Statistical Analysis.** The mean numbers of CAG and GGN repeats were compared in cases and controls using the *t* test. Unconditional logistic regression models were used to estimate ORs and their corresponding 95% CIs for prostate cancer in relation to CAG and GGN repeat lengths (28). Repeat lengths were examined first as continuous variables and later as categorical variables. The distributions of the number of CAG or GGN repeats among controls were used to derive the median or tertile cutoffs used to calculate ORs. In addition, the combined effects of CAG and GGN were evaluated based on the median lengths within the controls. The relationships between age, CAG and GGN repeat length, and other variables were assessed by Spearman correlation and ANOVA.

## RESULTS

Selected characteristics of cases and controls are shown in Table 1. Compared with controls, cases had higher caloric intake, higher levels of education, and a higher waist:hip ratio and were less likely to use cigarettes or alcohol. Age at diagnosis ranged from 50 to 94 (median 73) for cancer cases. Sixty-nine cases (36%) were diagnosed as having localized cancer, and most tumors (72%) were moderately or poorly differentiated.

Because the *AR* gene is located on the X chromosome, only one copy of the gene is present in men. For the polyglutamine tract (CAG)<sub>n</sub>CAA, there was no variation in the CAA sequence among the 490 samples analyzed. The number of CAG repeats ranged from 10 to 34. About 65% of the study subjects had a CAG repeat length that ranged from 21 to 24, but only 1% of the subjects had a CAG length longer than 30 repeats (Table 2). Although the median number of CAG repeats in controls was only slightly larger than that in cases (23.0 versus 22.0), there was a shift toward longer repeat length among controls (Fig. 1). For CAG repeat length shorter than 23, cases had higher percentages than controls in 6 of the 10 categories. However, for CAG repeat length longer than 22, controls had higher percentages than cases in 8 of the 12 categories. Age at diagnosis and stage of cancer were not related to CAG repeat length, with similar

<sup>3</sup> A. W. Hsing, J. Deng, T. Xie, I. A. Sesterhenn, F. K. Mostofi, and Y.-T. Gao. Body size and prostate cancer: a population-based case-control study in Shanghai, China, submitted for publication.

Table 1 Selected characteristics of prostate cancer cases and population controls, China

Characteristics	Cases	Controls
	(n = 191) Mean (SD)	(n = 304) Mean (SD)
Age (yr)	72.2 (7.7)	71.9 (7.3)
Total calories (kcal/day)	2457.0 (647)	2342.0 (731)
Height (cm)	167.9 (6.0)	167.6 (5.8)
Weight (kg)	61.3 (8.4)	61.5 (10.1)
Body mass index (kg/m <sup>2</sup> )	21.8 (2.9)	21.9 (3.3)
Waist circumference (cm)	82.6 (10.4)	82.5 (10.7)
Hip circumference (cm)	90.7 (8.9)	92.6 (8.5)
Waist:hip ratio	0.91 (0.05)	0.89 (0.05)
% married	89.5	92.1
% with education greater than high school	34.6	25.3
% smokers	56.5	65.1
% alcohol users	31.4	42.1
Clinical stage (%)		
Localized	36.3	
Regional	30.5	
Remote	32.1	
Histologic grade (%)		
Well-differentiated	7.9	
Moderately differentiated	31.0	
Poorly differentiated	41.0	
Cannot be assessed	20.1	

distribution and average number of CAG and GGN repeat lengths in various age categories and three clinical stages.

For the polyglycine tract (GGT<sub>3</sub>GGG<sub>1</sub>GGT<sub>2</sub>GGC<sub>n</sub>), there was no variation in the codon usage or the number of GGT and GGC trinucleotides in all of the 481 samples analyzed, although the number of GGC repeats was highly variable. The pattern was always three GGT, one GGG, two GGT, followed by a variable number of GGC. The number of GGN repeats among study subjects ranged from 14 to 27 (the number of GGC repeats thus ranged from 9 to 21; Table 3). About 72% of the study subjects had a GGN repeat length of 23.

Risks of prostate cancer associated with CAG and GGN repeat lengths are shown in Table 4. When the number of CAG repeats was included in the model as a continuous variable, there was a 7% increase in the risk of prostate cancer for each decrement in length of one CAG repeat (OR, 1.07; 95% CI, 1.00–1.15). The risks associated with decrements of three and six repeats were 1.21 (95% CI, 1.14–1.32) and 1.42 (95% CI, 1.22–1.61), respectively. When the median repeat length was used to dichotomize study subjects, men with a CAG repeat length shorter than 23 had a 65% increased risk (OR, 1.65; 95% CI, 1.14–2.39), compared with men with a CAG repeat length of 23 or longer. Relative to the highest tertile of CAG repeat length ( $\geq 24$ ), men in the second and first tertiles (22–23 and  $< 22$ , respectively) had ORs of 1.45 and 1.55, respectively ( $P_{trend} = 0.06$ ).

Similarly, men with a shorter GGN repeat length had a higher risk of prostate cancer, although the excess risk was moderate and did not reach statistical significance. Each decrement of one GGN repeat length was associated with a 7% increase in risk (OR, 1.07; 95% CI, 0.96–1.20). Men with a GGN repeat length shorter than the median length of 23 had a 12% increase in prostate cancer risk, compared with those with  $\geq 23$  repeats. Because  $> 72\%$  of the subjects had 23 GGN repeats, we could not estimate ORs by tertiles for GGN repeats.

Also shown in Table 4 are the ORs associated with combined categories of CAG and GGN repeat lengths. Men with both CAG and GGN repeat lengths shorter than 23 had a 75% elevated risk of prostate cancer. There was little correlation between the number of CAG and GGN repeats ( $r = -0.03$ ;  $P > 0.05$ ).

The number of CAG or GGN repeats did not correlate with age, education, body mass index, waist:hip ratio, total calories, smoking, or drinking. These variables therefore were not included in the model for adjustment. The ORs were materially unchanged after further adjustment for BPH, although the cases had a higher prevalence of BPH

(57% versus 23%) and there was a nonsignificant moderate association between CAG or GGN repeat lengths and BPH (data are reported separately). Associations of CAG or GGN repeat length were similar across all stages of disease at diagnosis (data not shown).

## DISCUSSION

Results from our study, the first population-based study conducted in a low-risk population, confirm the hypothesis that a shorter CAG repeat length is associated with an increased risk of clinically significant prostate cancer. A shorter length of GGN repeat also appears to increase the risk of prostate cancer, but the magnitude of excess risk was smaller and did not reach statistical significance.

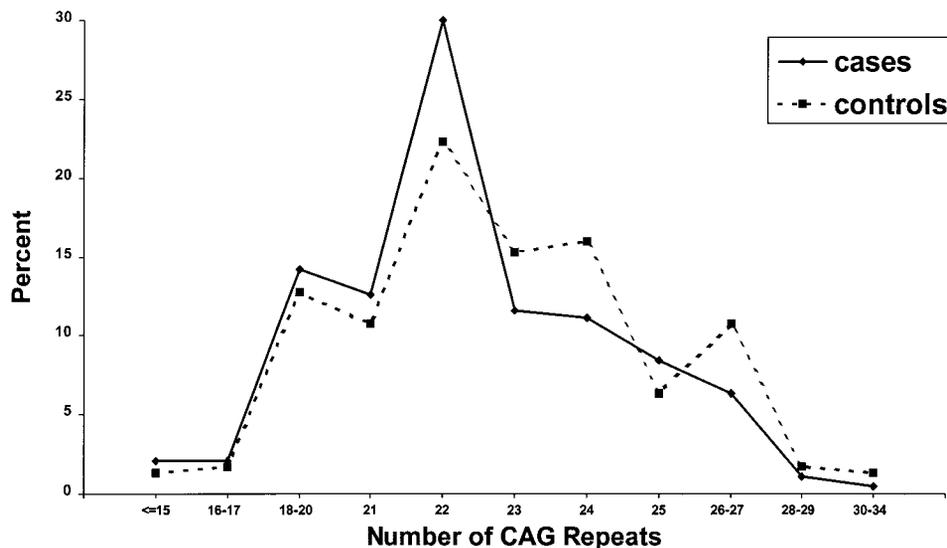
Our finding that a shorter CAG repeat length is associated with a higher risk of prostate cancer is consistent with four previous studies (17–22) but differs from one study in a French-German population that reported no such association (29). Despite the cross-sectional nature of our study, the observed inverse association may reflect some degree of genetic predisposition to clinically significant cancer (and thus to progression of prostate tumors) because most of the cases in our study had clinically advanced cancer and few, if any, were identified through screening (average serum levels of prostate-specific antigen among cases was 73). In addition, despite the large racial difference in clinical prostate cancer risk, the prevalence of latent prostate tumors has been reported to be similar across populations (5), suggesting that racial variation in the polymorphisms of AR may be related to differences in genetic susceptibility to progression rather than to initiation of prostate cancer.

The observed inverse association with AR polymorphisms is biologically plausible because laboratory studies have shown that a long polyglutamine chain ( $> 30$  repeats) in the AR gene is associated with androgen insensitivity and reduced AR transactivation activity (13, 14). *In vitro* transfection studies also have demonstrated that elimination of the polyglutamine tracts results in elevated transcriptional activities (9, 11, 16). Clinical studies have suggested that alteration in the AR function, through either polymorphisms of CAG repeat length or somatic mutations, may be associated with tumor progression. For example, the progression from latent to clinically invasive prostate cancer is initially androgen-dependent, although some tumors later

Table 2 Distribution of number of CAG repeats in the AR gene in prostate cancer cases and controls, China

No. of CAG repeats	Cases (n = 190)		Controls (n = 300)	
	n	%	n	%
10	0	0.0	1	0.3
14	0	0.0	1	0.3
15	4	2.1	2	0.7
16	3	1.6	2	0.7
17	1	0.5	3	1.0
18	7	3.7	9	3.0
19	7	3.7	12	4.0
20	13	6.8	17	5.7
21	24	12.6	32	10.7
22	57	30.0	67	22.3
23	22	11.6	46	15.3
24	21	11.1	48	16.0
25	16	8.4	19	6.3
26	9	4.7	17	5.7
27	3	1.6	15	5.0
28	2	1.1	2	0.7
29	0	0.0	3	1.0
30	0	0.0	1	0.3
31	0	0.0	2	0.7
32	0	0.0	0	0.0
33	0	0.0	1	0.3
34	1	0.5	0	0.0
Median	22		23	

Fig. 1. Percent distribution of number of CAG repeats.



become androgen-independent (thus becoming nonresponsive to hormonal treatment). Several non-germ-line-related changes of the *AR* gene, including amplification of the *AR* gene (usually a key step in the transition from a hormone-sensitive to a hormone-refractory state in prostate tumors; Refs. 30 and 31), *AR* somatic mutations (identified throughout transactivation, DNA binding, and ligand binding do-

main; Refs. 32 and 33), and contraction of CAG repeat length in cancer cells (31), have been shown to be associated with tumor aggressiveness, cancer progression, and failure of hormonal therapy. *AR* expression studies in the majority of prostate tumors, including those that have become refractory to hormonal therapy, also suggest that *AR* plays a key role in androgen-independent tumors (34, 35).

The inverse relationship between CAG repeat length and *AR* transcriptional activity (thus androgen sensitivity) is the presently recognized underlying molecular mechanism by which *AR* polymorphisms modulate prostate cancer risk. Because transcriptional activation of the *AR* gene is influenced by not only polymorphisms in the *AR* gene but also a number of other factors, including tissue levels of dihydrotestosterone, estradiol, insulin-like growth factors, and *AR* coactivators (36–42), it is likely that these factors may also affect prostate cancer risk by mediating transcriptional activities. Several *AR* coactivators, including *AR*-associated proteins (*ARA70* and *ARA55*), amplified in breast cancer-1 (*AIB1*), cyclic AMP-responsive element-binding protein (*CBP*), *Rb*, and *BRCA1*, have been shown to enhance *AR*-mediated transcriptional activity from 2- to 10-fold, suggesting that *in vivo* coactivators are essential in attaining optimal *AR* transactivation in response to androgens (39–42). Future studies are needed to evaluate the effects of *AR* in conjunction with these

Table 3 Distribution of number of GGN repeats in the *AR* gene in prostate cancer cases and controls, China

No. of GGN repeats	Cases (n = 187)		Controls (n = 295)	
	n	%	n	%
14	0	0.0	1	0.3
15	1	0.5	1	0.3
16	2	1.1	1	0.3
17	2	1.1	0	0.0
18	0	0.0	1	0.3
19	19	10.2	20	6.8
20	2	1.1	2	0.7
21	3	1.6	0	0.0
22	10	5.3	24	8.2
23	136	72.7	212	72.1
24	10	5.3	24	8.2
25	2	1.1	1	0.3
27	0	0.0	1	0.3
Median	23		23	

Table 4 ORs and 95% CIs for prostate cancer in relation to the number of CAG and GGN repeats in the *AR* gene, China

No. of CAG and GGN repeats	No. of cases	No. of controls	OR <sup>a</sup>	95% CI
<b>No. of CAG repeats</b>				
Continuous (per decrement of one CAG repeat)	190	300	1.07	1.00–1.15
Median				
≥23	74	154	1.00	
<23	116	146	1.65	1.14–2.39
<b>Tertile</b>				
≥24	52	108	1.00	
22–23	79	113	1.45	0.93–2.25
<22	59	79	1.55	0.96–2.49
<i>Linear trend P</i> = 0.06				
<b>No. of GGN repeats</b>				
Continuous (per decrement of one GGN repeat)	187	294	1.07	0.96–1.20
Median				
≥23	147	239	1.00	1.00
<23	39	56	1.12	0.71–1.78
<b>Combined number of CAG and GGN repeats</b>				
CAG ≥23, GGN ≥23	53	120	1.00	
CAG ≥23, GGN <23	19	29	1.48	0.76–2.88
CAG <23, GGN ≥23	94	115	1.85	1.21–2.82
CAG <23, GGN <23	20	26	1.75	0.90–3.41

<sup>a</sup>Adjusted for age (continuous).

coactivators to clarify further the underlying mechanism of androgenic pathways in prostate carcinogenesis.

It has been suggested that variations in CAG repeat length in the *AR* gene between populations may explain part of the large racial difference in prostate cancer risk and that a shorter CAG repeat length reported for African Americans may contribute to some of their higher risk of prostate cancer, although presently no data are available from this population. Our results confirm that, relative to western men, Chinese men do indeed have a longer CAG repeat length. For example, 22% of the 1722 white men in two United States studies (17, 18) had a CAG repeat length shorter than 20 *versus* only 10% in our study and 55% reported for African Americans in a cross-sectional survey (26, 27). Our results, based on the population with the lowest reported incidence of prostate cancer in the world, cannot be generalized directly to African Americans. However, inverse associations have also been reported for Caucasians, suggesting that the underlying biological mechanism in various racial groups may be similar and that the polymorphisms of *AR* may be related, in part, to racial difference in prostate cancer risk.

The polymorphic CAG repeat length in the *AR* gene represents the first of a new class of common polymorphisms as genetic risk factors for prostate cancer. Rare genetic factors with high penetrance, such as *HPC1* on chromosome 1, conferring a much higher relative risk to a few individuals who carry them (for example, *HPC1* may explain about 10% of the prostate cancer cases in the United States), are unlikely to explain the large racial difference in prostate cancer risk. In contrast, the common CAG polymorphism of the *AR* gene confers variable risk upon all individuals, which in turn may result in a much larger proportion of prostate cancer cases attributable to having fewer CAG repeats. Assuming that the CAG polymorphism association is causal, we estimated that 25% (95% CI, 9–41%) of the cases in Shanghai can be attributed to a CAG repeat length shorter than 23. In an effort to provide insights into the reasons for the substantial racial difference in prostate cancer risk, using the CAG repeat length distribution in the two United States studies among white men (17, 18), we further estimated that 3–7% of cases among United States white men can be attributed to the CAG polymorphism (repeat length <23) and that this polymorphism alone potentially accounts for at least 5% of the difference in incidence between Chinese and United States men.

Similar to two previous studies (17, 18), we found that the number of GGN repeats clusters around 23 [in the study of Stanford *et al.* (17), only the number of GGC repeats was counted and 15 was the peak of the repeat, which corresponds to 21 GGN repeats], and that a shorter GGN repeat length appears to be associated with a moderate increase in prostate cancer risk. Twenty-three GGN repeats may represent the coding sequence for optimal *AR* protein conformation and activity, since >70% of the study subjects in our study as well as in studies of western men had a GGN repeat length of 23.

Although it is well established that (GGC)<sub>n</sub> repeats in the polyglycine tract (GGT<sub>3</sub>GGG<sub>1</sub>GGT<sub>2</sub>GGC<sub>n</sub>) of the *AR* gene are polymorphic, to date there has been little information on variations in the GGG and GGT regions of the polyglycine tract because these regions are GC-rich and technically it has been difficult to amplify these regions. Our study represents the first successful effort to sequence the exact codon usage and number of the GGN trinucleotide repeats in a large number of population-based samples. We showed that GGT and GGG regions were quite stable and that there were no variations in these two regions in all of the 481 DNA samples analyzed.

Survival and selection biases in our study should be minimal because well over 90% of the eligible cases participated in the study and most cases were interviewed within 30 days after diagnosis. Seventy to 80% of the study subjects gave blood for the study, so it

is unlikely that response status among cases and controls was related to the number of CAG or GGN repeats.

In summary, results from our population-based multidisciplinary case-control study in China confirm the hypothesis that a shorter CAG repeat length is associated with clinically significant prostate cancer and that relative to western men, Chinese men do have longer CAG and GGN repeat lengths. Larger studies are needed to evaluate the combined effect of CAG and GGN repeats, especially among African Americans. Because of the importance of *AR* in prostate cancer etiology, investigation of factors that might interact with the polyglutamine region of the *AR* gene to alter *AR* function and modulate prostate cancer risk is an important area for future research.

## ACKNOWLEDGMENTS

We thank Jiaorong Cheng of the Shanghai Cancer Institute for specimen collection and processing; collaborating hospitals and urologists for data collection; pathologists for pathology review; Linda Lannom, John Heinrich, Nancy Odaka, Kimberly Viskul, Mille Bendel, and Harvey Co Chien of Westat for data preparation and management; and Mary McAdams, Jean Cyr, and Leslie Carroll of Information Management Systems, Inc. for data analysis.

## REFERENCES

- Parkin, D. M., Whelan, S. L., Ferlay, J., Raymond, L., and Young, L. (eds). Cancer Incidence in Five Continents. II. IARC Scientific Publ. No. 143. Lyon, France: IARC, 1997.
- Hsing, A. W., Tsao, L., and Devesa, S. S. International trends and patterns of prostate cancer incidence and mortality. *Int. J. Cancer*, 85: 60–67, 2000.
- Hsing, A. W., Devesa, S. S., Jin, F., and Gao, Y.-T. Rising incidence of prostate cancer in Shanghai, China. *Cancer Epidemiol. Biomark. Prev.*, 7: 83–84, 1998.
- Shimizu, H., Ross, R. K., and Bernstein, L. Possible underestimation of the incidence rate of prostate cancer in Japan. *Jpn. J. Cancer Res.*, 82: 483–485, 1991.
- Breslow, N., Chan, C. W., Dhom, G., Drury, R. A., Franks, L. M., Gellei, B., Lee, Y. S., Lundberg, S., Sparke, B., Sternby, N. H., and Tulinius, H. Latent carcinoma of the prostate of autopsy in seven areas. *Int. J. Cancer*, 20: 680–688, 1977.
- Roy, A. K., Lavrovsky, Y., Song, C. S., Chen, S., Jung, M. H., Velu, N. K., Bi, B. Y., and Chatterjee, B. Regulation of androgen action. *Vitam. Horm.*, 55: 309–352, 1999.
- Trapman, J., and Brinkmann, A. O. The androgen receptor in prostate cancer. *Pathol. Res. Pract.*, 192: 752–760, 1996.
- McPhaul, M. J. Molecular defects of the androgen receptor. *J. Steroid Biochem. Mol. Biol.*, 69: 315–322, 1999.
- Chamberlain, N. L., Driver, E. D., and Miesfeld, R. L. T. 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.
- Choong, C. S., Kempainen, J. A., Zhou, Z. X., and Wilson, E. M. Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Mol. Endocrinol.*, 10: 1527–1535, 1996.
- 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 *trans*-activation, impaired sperm production, and male infertility. *J. Clin. Endocrinol. Metab.*, 82: 3777–3782, 1997.
- Zirkin, B. R. Spermatogenesis: its regulation by testosterone and FSH. *Semin. Cell Dev. Biol.*, 9: 417–421, 1998.
- Fischbeck, K. H., Lieberman, A., Bailey, C. K., Abel, A., and Merry D. E. Androgen receptor mutation in Kennedy's disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 354: 1075–1078, 1999.
- Brooks, B. P., and Fischbeck, K. H. Spinal and bulbar muscular atrophy: a trinucleotide-repeat expansion neurodegenerative disease. *Trends Neurosci.*, 18: 459–461, 1995.
- 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)<sub>n</sub>-expanded neuropathies. *Hum. Mol. Genet.*, 4: 523–527, 1995.
- Gao, T., Marcellini, M., and McPhaul, M. J. Transcriptional activation and transient expression of the human androgen receptor. *J. Steroid Biochem. Mol. Biol.*, 59: 9–20, 1996.
- 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.
- Gioannucci, 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. *Proc. Natl. Acad. Sci. USA*, 94: 3320–3323, 1997.
- 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.*, 89: 166–170, 1997.

20. 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.
21. Kantoff, P., Giovannucci, E., and Brown, M. The androgen receptor CAG repeat polymorphism and its relationship to prostate cancer. *Biochim. Biophys. Acta*, *1378*: C1–C5, 1998.
22. 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.
23. Sartor, O., Zheng, Q., and Eastham, J. A. Androgen receptor gene CAG repeat length varies in a race-specific fashion in men without prostate cancer. *Urology*, *53*: 378–380, 1999.
24. Hardy, D. O., Scher, H. I., Bogenreider, T., Sabbatini, P., Zhang, Z. F., Nanus, D. M., and Catterall, J. F. Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. *Clin. Endocrinol. Metab.*, *81*: 4400–4405, 1996.
25. Bratt, O., Borg, A., Kristoffersson, U., Lundgren, R., Zhang, Q. X., Olsson, H. CAG repeat length in the androgen receptor gene is related to age at prostate cancer diagnosis and response to endocrine therapy, but not to risk of prostate cancer. *Br. J. Cancer*, *81*: 672–676, 1999.
26. 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.
27. Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., and Chakraborty, R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, *12*: 241–253, 1992.
28. Breslow, N. E., and Day, N. E. *Statistical Methods in Cancer Research. I. The Analysis of Case-Control Studies.* IARC Scientific Publ. No. 32. Lyon, France: IARC, 1980.
29. Correa-Cerro, L., Wöhr, 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)<sub>n</sub>CAA<sub>n</sub> 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.
30. Wallen, M. J., Linja, M., Kaartinen, K., Schleutker, J., and Visakorpi, T. Androgen receptor gene mutations in hormone-refractory prostate cancer. *J. Pathol.*, *189*: 559–563, 1999.
31. Koivisto, P. A., and Rantala, I. Amplification of the androgen receptor gene is associated with P53 mutation in hormone-refractory recurrent prostate cancer. *J. Pathol.*, *187*: 237–241, 1999.
32. Culig, Z., Hobisch, A., Cronauer, M. V., Cato, A. C., Hittmair, A., Radmayr, C., Eberle, J., Bartsch, G., and Klocker, H. Mutant androgen receptor detected in an advanced-stage prostatic carcinoma is activated by adrenal androgens and progesterone. *Mol. Endocrinol.*, *7*: 1541–1550, 1997.
33. Schoenberg, M. P., Hakimi, J. M., Wang, S., Bova, G. S., Epstein, J. I., Fischbeck, K. H., Isaacs, W. B., Walsh, P. C., Barrack, E. R. Microsatellite mutation (CAG24→18) in the androgen receptor gene in human prostate cancer. *Biochem. Biophys. Res. Commun.*, *198*: 74–80, 1994.
34. Jenster, G. The role of the androgen receptor in the development and progression of prostate cancer. *Semin. Oncol.*, *26*: 407–421, 1999.
35. Koivisto, P., Kononen, J., Palmberg, C., Tammela, T., Hyytinen, E., Isola, J., Trapman, J., Cleutjens, K., Noordzij, A., Visakorpi, T., and Kallioniemi, O. P. Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. *Cancer Res.*, *57*: 314–319, 1997.
36. Culig, Z., Hobisch, A., Cronauer, M. V., Hittmair, A., Radmayr, C., Bartsch, G., and Klocker, H. Activation of the androgen receptor by polypeptide growth factors and cellular regulators. *World J. Urol.*, *13*: 285–289, 1995.
37. McAbee, M. D., and DonCarlos, L. L. Estrogen, but not androgens, regulates androgen receptor messenger ribonucleic acid expression in the developing male rat forebrain. *Endocrinology*, *140*: 3674–3681, 1999.
38. Gupta, C. Modulation of androgen receptor (AR)-mediated transcriptional activity by EGF in the developing mouse reproductive tract primary cells. *Mol. Cell. Endocrinol.*, *152*: 169–178, 1999.
39. Cude, K. J., Dixon, S. C., Guo, Y., Lisella, J., and Figg, W. D. The androgen receptor: genetic considerations in the development and treatment of prostate cancer. *J. Mol. Med.*, *77*: 419–426, 1999.
40. Yeh, S., and Chang, C. Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. *Proc. Natl. Acad. Sci. USA*, *93*: 5517–5521, 1997.
41. Yeh, S., Miyamoto, H., Shima, H., and Chang, C. From estrogen to androgen receptor: a new pathway for sex hormones in prostate. *Proc. Natl. Acad. Sci. USA*, *95*: 5527–5532, 1998.
42. Anzick, S. L., Kononen, J., Walker, R. L., Azorsa, D. O., Tanner, M. M., Guan, X. Y., Sauter, G., Kallioniemi, O. P., Trent, J. M., and Meltzer, P. S. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science*, *277*: 965–968, 1997.