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, 1 Yu-Tang Gao, Guan Wu, Xin Wang, Jie Deng, Yuh-Ling Chen, Isabell A. Sesterhenn, F. Kash Mostofi, Jacques Benichou, and Chaowansheng 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, GGG, GGT, GGC)n, based on the sequencing results from 481 samples, we are the first to show that although GGC regions in the polyglutamate 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 ≥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, 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 NH2-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 polyglutamine (GGN) tract varies from about 10 to 30 repeats. The functional consequences of variation in the GGN tract are less clear. Deletion of the polyglutamine tract reduces AR transcriptional activity by ~30% in transient transfection assays (16), although there is no significant correlation between polyglutamine 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).
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. 3 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 centralized laboratory in Shanghai and stored at −70°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 molecular analysis and assessment of the CAG and GGN repeats.

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 part of an ongoing molecular analysis of the AR gene is located on the X chromosome, only one copy of the gene is present in men. For the polyglutamine tract (CAGn 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 longer 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

Because the AR gene is located on the X chromosome, only one copy of the gene is present in men. For the polyglutamine tract (CAGn 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 longer 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
distribution and average number of CAG and GGN repeat lengths in various age categories and three clinical stages.

For the polyglycine tract (GGTn,GGn,GGTn,GGGn), there was no variation in the codon usage or the number of GGT and GGG 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 GGG repeats among study subjects ranged from 14 to 30 repeats) in the...
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 domains; 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 factors.

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

<table>
<thead>
<tr>
<th>No. of GGN repeats</th>
<th>Cases (n = 187)</th>
<th>Controls (n = 295)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>10.2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td>22</td>
<td>10</td>
<td>5.3</td>
</tr>
<tr>
<td>23</td>
<td>136</td>
<td>72.7</td>
</tr>
<tr>
<td>24</td>
<td>10</td>
<td>5.3</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>23</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>No. of CAG and GGN repeats</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>OR*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous (per decrement of one CAG repeat)</td>
<td>190</td>
<td>300</td>
<td>1.07</td>
<td>1.00–1.15</td>
</tr>
<tr>
<td>Median</td>
<td>74</td>
<td>154</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>&lt;23</td>
<td>116</td>
<td>146</td>
<td>1.65</td>
<td>1.14–2.39</td>
</tr>
<tr>
<td>Tertile</td>
<td>52</td>
<td>108</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>≥24</td>
<td>79</td>
<td>113</td>
<td>1.45</td>
<td>0.93–2.25</td>
</tr>
<tr>
<td>22–23</td>
<td>59</td>
<td>79</td>
<td>1.55</td>
<td>0.96–2.49</td>
</tr>
<tr>
<td>&lt;22</td>
<td>52</td>
<td>108</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

Linear trend P = 0.06

<table>
<thead>
<tr>
<th>No. of CAG repeats</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>OR*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous (per decrement of one GGN repeat)</td>
<td>187</td>
<td>294</td>
<td>1.07</td>
<td>0.96–1.20</td>
</tr>
<tr>
<td>Median</td>
<td>147</td>
<td>239</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>&lt;23</td>
<td>39</td>
<td>56</td>
<td>1.12</td>
<td>0.71–1.78</td>
</tr>
<tr>
<td>Combined number of CAG and GGN repeats</td>
<td>53</td>
<td>120</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>CAG ≥23, GGN ≥23</td>
<td>19</td>
<td>29</td>
<td>1.48</td>
<td>0.76–2.88</td>
</tr>
<tr>
<td>CAG &lt;23, GGN ≥23</td>
<td>94</td>
<td>115</td>
<td>1.85</td>
<td>1.21–2.82</td>
</tr>
<tr>
<td>CAG &lt;23, GGN &lt;23</td>
<td>20</td>
<td>26</td>
<td>1.75</td>
<td>0.90–3.41</td>
</tr>
</tbody>
</table>

*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)n repeats in the polyglycine tract (GTTGGGTGGTGGTGGCn) 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


Downloaded from cancerres.aacrjournals.org on August 15, 2017. © 2000 American Association for Cancer Research.


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


**Updated version**
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/60/18/5111

**Cited articles**
This article cites 39 articles, 10 of which you can access for free at:
http://cancerres.aacrjournals.org/content/60/18/5111.full#ref-list-1

**Citing articles**
This article has been cited by 35 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/60/18/5111.full#related-urls

**E-mail alerts**
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.