Androgen Receptor CAG Repeats and Prostate Cancer

Kevin A. Nelson and John S. Witte

Prostate cancer is the most common nonskin malignancy and the second leading cause of cancer deaths among men in the United States. Prostate cancer (Mendelian Inheritance in Man 176807) has a complex etiology; presently, age, ethnicity, and family history are the most consistently reported risk factors associated with disease. Other potential risk and protective factors have also been suggested. Androgen, acting through the androgen receptor (AR) is helpful in preserving the normal function and structure of the prostate. The AR (Mendelian Inheritance in Man 313700) is a structurally conserved member of the nuclear receptor superfamily of ligand-activated transcription factors. Androgens, such as testosterone, are strong tumor promoters, and work with the AR to augment the effect of any carcinogens present and stimulate cell division. The CAG repeats encode long glutamine homopolymeric amino acid chains in the amino-terminal domain of the AR gene. The authors focus on CAG repeat length because recent research suggests that men with shorter AR CAG lengths (e.g., ≤22 repeats) are at a greater risk of developing prostate cancer than are those with longer variants. Among populations studied to date, African Americans appear to have the highest frequency of short CAG repeats. Several potential interactions have also been explored, including molecular interactions, androgen deprivation therapy, and prostate-specific antigen expression. CAG repeat length can be determined with high sensitivity and specificity. Presently, there is no recommended population screening for AR CAG repeat length.
the transport of the androgen hormone (13).

It has been observed that shorter AR CAG repeats impose a higher transactivation activity on the receptor and have an increased binding affinity for androgens (6, 14). This may make the prostate more vulnerable to chronic androgen over-stimulation and increased proliferative activity, which, in turn, could increase the rate of somatic mutations among tumor suppressor genes (e.g., the AR). In contrast, the expansion of the CAG repeat (>40 repeats) leads to a below-normal AR level when measured by transient transfections assays with hormone-responsive-computerized axial tomography constructs (15–18) and by Scatchard analysis of AR in skin fibroblasts (19). It has also been found that there is an altered coactivator interaction with the AR (3). Thus, the CAG microsatellite encodes a polyglutamine tract, which has a length that is inversely and linearly related to AR activity (20, 21).

**GENE VARIANTS**

Table 1 presents the frequency by ethnicity of AR CAG repeat lengths among relevant studies detected in a search of Medline between 1990 and September 2001. We linked the keywords “androgen receptor” and “trinucleotide repeats” for one search and the keywords “androgen receptor” and “prostatic neoplasms” for another. (This search strategy was also applied to the remaining sections of this paper.) Note that estimates of population-specific allele frequencies might be affected by misclassification of ethnicity. For example, Edwards et al. (22) give estimates of CAG repeat frequency among European Americans, African Americans, Asian Americans, and Latinos using DNA extracted from a convenience sample of blood donated to Houston, Texas, blood banks. Blood-bank personnel noted European-American and African-American ethnicity based on visual appearances. Asian-American and Latino ethnicities were approximated based on surnames. Using visual or name identification methods could explain some of the ethnic variation (31). Some have also suggested that varying levels of testosterone and its intraprostatic metabolite dihydrotestosterone across ethnic groups may be responsible for these differences (32).

**DISEASES**

**Incidence and mortality**

Prostate cancer is the most common solid tumor and the second leading cause of cancer deaths among men in the United States: In 2001, approximately 198,100 men were diagnosed with prostate cancer, and about 31,500 will die of the disease. On the basis of data gathered during 1994–1996, a male has a one-in-six chance of developing prostate cancer at some point in his lifetime (26).

**Ethnicity**

Ethnicity is also one of the most widely accepted risk factors for prostate cancer. This helps to explain, in part, the wide geographic variation in the incidence of this disease. For instance, the United States has an incidence of prostate cancer that is eight times higher than that of Japan (27), and African Americans have the highest incidence and mortality in the world. For example, the Surveillance, Epidemiology, and End Results program data from the years 1988–1992 estimate age-adjusted incidence rates of 181/100,000 for African-American men versus only 135/100,000 among Caucasians; similarly, the age-adjusted prostate cancer mortality rate for African Americans is 54/100,000 versus only 24/100,000 in Caucasians (28). In all likelihood, racial and geographic differences are multifactorial in nature, with environmental, genetic, and possibly social factors (29, 30). Differences in diet and access to improved detection methods could explain some of the ethnic variation (31). Some have also suggested that varying levels of testosterone and its intraprostatic metabolite dihydrotestosterone across ethnic groups may be responsible for these differences (32).

**Family history**

In addition, family history is a generally accepted risk factor for prostate cancer. For younger men, familial factors appear to be particularly influential, and the attributable risk of strong familial factors may be as high as 43 percent for men who are less than age 55 years (33). Nevertheless, family history of prostate cancer accounts for only about 9 percent of all cases (33). First-degree relatives of men diagnosed with prostate cancer have a two- to sixfold excess risk of the disease (34–38). Spitz et al. (37) and Steinberg et al. (39) suggest that the more closely a man is related to an affected relative and the greater the number of affected relatives in his family, the higher is his risk of acquiring prostate cancer. Narod et al. (40) and Monroe et al. (41) found that brothers of cases faced significantly higher risks of prostate cancer than did fathers.

**Other risk factors**

Age is an important risk factor for prostate cancer (30): Men are rarely diagnosed before age 40 years, and the rate of increase with aging—at approximately the ninth to tenth power of age—is greater than that for any other cancer (13). A large number of other putative risk factors for prostate cancer have been identified, although whether these are causal remains controversial. Among these are obesity (42), alcohol (43), sexual behavior (35), cigarette smoking (44), use of pesticides (45), and nuclear power (46). Other factors (for example, selenium (47); vitamins A, D, and E (48–51);
and phytoestrogens (52, 53)) have been inversely associated with prostate cancer.

**AR and prostate cancer**

Prostate cancer is one of the primary diseases that have been associated with somatic and germline polymorphisms in the AR gene. In the “Associations” section below, we discuss in detail the relation between CAG repeats in the AR gene and prostate cancer. Other AR polymorphisms potentially associated with prostate cancer include GGC/N repeats, Stu I, and numerous base substitutions (6, 11, 13, 24, 50, 54–56).

**CAG polymorphisms in the AR**

Shorter CAG repeats have been associated with benign prostatic hyperplasia (BPH) (57, 58). Short CAG repeat lengths have also been associated with androgenetic alopecia (59, 60), ankylosing spondylitis (61), mental retardation (62), younger age of male rheumatoid arthritis (63), and hepatitis B-related hepatic carcinoma (64). In contrast, extremely long AR CAG length (i.e., 38–52 repeats) has been associated with Kennedy’s disease, spinal and bulbar muscular atrophy (SBMA) (65, 66). SBMA is an adult-onset, motor neuron disease that is associated with reduced fertility, low virilization, testicular atrophy, and reduced sperm production (13). In addition, it is of interest to note that men with SBMA have also been diagnosed with prostate cancer (67). Expansions of the CAG repeat may also be involved in the carcinogenesis of uterine endometrial cells (68). Finally, Rebbeck et al. (69) found that BRCA1 mutation carriers who carry long CAG repeats have an earlier age of breast cancer diagnosis. However, Spurdle et al. (70) found no association between CAG repeat length and breast cancer risk in women under age 40 years.

**ASSOCIATIONS**

**CAG repeat length and occurrence of prostate cancer**

Table 1 also gives odds ratios for the association between CAG repeats in the AR gene and prostate cancer. These results suggest that the length of CAG repeats might be inversely associated with the risk of prostate cancer. Coetzee and Ross (14) initially suggested that variations in CAG repeat length are associated with prostate cancer. One possible explanation for this is that short alleles may impose higher transactivation on the receptor due to the inverse relation between the number of glutamine residues in the polyglutamine tract and transcriptional activity (6, 14, 18). This potential association may be modified by stage/grade at diagnosis (71), and men diagnosed at an older age appear to have longer CAG repeats (9), suggesting that AR CAG repeat length may also be associated with prostate cancer aggressiveness. The use of a lower cutpoint shifts men from the short to the long category of trinucleotide repeat length. Thus, the proportion of short alleles is reduced, while that of longer alleles is increased. Other things being equal, this would bias the association between shorter alleles and the occurrence of prostate cancer. One would also expect that the use of a lower cutpoint would accentuate short-long allele differences between racial groups because a higher proportion of African Americans have a low number of CAG repeats (9–17) compared with Caucasians or Asians. Reducing the short-long cutpoint should reclassify a greater proportion of Caucasians from the short to the long group compared with African Americans. The ratio of proportions for the shorter alleles for African Americans compared with Caucasians is higher in the study by Sartor et al. (72) than it is for the other studies that would be expected since a lower cutpoint is used.

Two studies listed here (73, 74) have not detected a statistically significant association between shorter CAG length and the occurrence of prostate cancer. Brett et al. (75) concluded that CAG repeats are associated with young age at diagnosis, but not with higher risk of the disease. This study did not give the distribution of CAG repeat lengths among the controls in tabular form. As a result, the odds ratio could not be computed, and this study was not included in table 1. Eeles et al. (76) did not find an association between CAG repeat length and prostate cancer development or aggressiveness among 178 Caucasian cases and matched controls. (Note that this has been published only in abstract form and does not give sufficient details for inclusion in table 1).

Most studies reported to date suggest a positive association between shorter CAG repeat lengths and prostate cancer. However, the magnitude of any potential association appears to be relatively limited, and small sample sizes have restricted the precision of published results. Nevertheless, the large population frequency of short CAG repeat lengths suggests that, if causal, variants in CAG lengths could have substantial public health implications. More specifically, this potentially low-penetrance, high-frequency mutation could theoretically account for many more prostate cancer cases than a high-penetrance, low-frequency mutation.

**CAG repeat length and aggressiveness of prostate cancer**

With regard to prostate cancer aggressiveness, Giovannucci et al. (71) found that comparing CAG repeat lengths among subjects with a high grade/stage (n = 269) gave an odds ratio of 1.64 (95 percent confidence interval [CI]: 1.22, 2.22), whereas those with a low grade/stage (n = 309) had an odds ratio of 1.02 (95 percent CI: 0.77, 1.37). Stanford et al. (11) also looked at the impact of CAG repeat length on disease aggressiveness (i.e., stage C or D or Gleason score 8–10). Comparison of short versus long CAG repeats (cutpoint = 22) gave an odds ratio of 1.20 (95 percent CI: 0.79, 1.84) for more aggressive prostate cancer. These results suggest that shorter CAG repeat length may be involved not only with the development of prostate cancer but also the potential aggressiveness of the disease.

**CAG repeat length and BPH**

A recent study by Giovannucci et al. (57) found that shorter CAG repeats are associated with BPH. The data for this came from the Health Professionals Follow-up Study,
TABLE 1. Androgen receptor CAG repeat length and association with prostate cancer

<table>
<thead>
<tr>
<th>Study (reference)</th>
<th>Place of study</th>
<th>Description of controls</th>
<th>Description of cases</th>
<th>Ethnic group</th>
<th>No. of subjects</th>
<th>CAG repeat frequency</th>
<th>Cases</th>
<th>Controls</th>
<th>OR*</th>
<th>95% CI*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Short</td>
<td>Long</td>
<td>CAG repeat cutpoint</td>
<td>No.</td>
<td>Frequency</td>
<td>No.</td>
</tr>
<tr>
<td>Edwards et al. (22)</td>
<td>Houston, TX</td>
<td></td>
<td>Caucasian</td>
<td>240</td>
<td>0.62</td>
<td>0.38</td>
<td>&gt;22</td>
<td>19</td>
<td>0.33</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>African American</td>
<td>250</td>
<td>0.76</td>
<td>0.24</td>
<td>≤22</td>
<td>38</td>
<td>0.67</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Asian</td>
<td>97</td>
<td>0.55</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hispanic</td>
<td>205</td>
<td>0.48</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irving et al.† (24)</td>
<td>Los Angeles, CA</td>
<td>Cross-sectional survey</td>
<td>Prostate cancer</td>
<td>39</td>
<td>0.63</td>
<td>0.38</td>
<td>&gt;22</td>
<td>19</td>
<td>0.33</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>patients</td>
<td>44</td>
<td>0.75</td>
<td>0.25</td>
<td>≤22</td>
<td>38</td>
<td>0.67</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>identified by</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingles et al.† (50)</td>
<td>Los Angeles, CA</td>
<td>Population-based</td>
<td>Same as Irving et</td>
<td>39</td>
<td>0.49</td>
<td>0.51</td>
<td>&gt;22</td>
<td>19</td>
<td>0.33</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>case-control study of</td>
<td>al. (24)</td>
<td>169</td>
<td>0.60</td>
<td>0.40</td>
<td>≤22</td>
<td>38</td>
<td>0.67</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bladder cancer,</td>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>matched to cases on</td>
<td>Non-Hispanic White</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>age, sex, and race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giovannucci et al. (71)</td>
<td>US</td>
<td>From the Physician’s</td>
<td>Caucasian (over</td>
<td>588</td>
<td>0.48</td>
<td>0.52</td>
<td>&gt;22</td>
<td>274</td>
<td>0.47</td>
<td>306</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Health Study</td>
<td>95%</td>
<td></td>
<td></td>
<td></td>
<td>≤22</td>
<td>313</td>
<td>0.53</td>
<td>282</td>
</tr>
<tr>
<td>Stanford et al. (11)</td>
<td>King County, WA</td>
<td>Population-based;</td>
<td>Caucasian</td>
<td>266</td>
<td>0.48</td>
<td>0.52</td>
<td>&gt;22</td>
<td>136</td>
<td>0.48</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ascertained by age-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤22</td>
<td>145</td>
<td>0.52</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>stratified random-digit telephone dialing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/aje/article-abstract/155/10/883/144142 by guest on 10 March 2019
Hakimi et al. (23) US From three “samplings” in different geographical areas Men undergoing radical retropubic prostatectomy Caucasian 370 0.03 0.97 >17 ≤17 53 0.90 359 0.03 1.00 Referent 3.7 1.3, 10.4
Sartor et al. (72) US Group of unaffected men Caucasian 130 0.28 0.72 >20 ≤20 65 0.57 0.43 1.00 Referent 0.96, 1.03
Edwards et al. London, England Women from a previous study of breast cancer Referred cases systematically ascertained Caucasian 390 0.54 0.46 ≥22 <22 74 0.46 178 0.46 1.00 Referent 0.54 1.00
Correa-Cerro et al. (73) Paris and Nancy, France Controls free of prostate cancer over 7 years Consecutive series French-German 105 0.52 0.48 ≥22 <22 69 0.48 50 0.52 1.00 Referent 0.83 0.73, 0.95
Ekman et al. (30) Stockholm, Sweden; Yamaguchi, Japan Men with BPH* From families with clustering of PCa Swedish Japanese 38 0.24 0.76 ≥15 <15 44 0.75 29 0.76 1.00 Referent 1.1 0.43, 2.84
Lange et al. MI (cases) Los Angeles, CA and Stanford et King County, al. (11) WA (controls) From families with two or more cases of prostate cancer Caucasian 133 >21 ≤21 1.00 Referent 0.99 0.60, 1.36

* OR, odds ratio; CI, confidence interval; SEER, Surveillance, Epidemiology, and End Results; BPH, benign prostate hypertrophy.
† These two studies used the same cases, with the paper by Irving et al. (24) simply adding more controls. In the paper by Edwards et al. (22), DNA was taken from blood banks from unrelated volunteers in a multilocus genotype survey. In the paper by Sartor et al. (72), subjects who had signs and symptoms of prostate cancer were excluded from the study. In the paper by Hakimi et al. (23), there was no listing of the number of observations per measured repeat length, which is necessary for recategorizing the data at a higher cutpoint (e.g., 22 CAG repeats), such as that used in other studies (e.g., the paper by Ingles et al. 50). In the paper by Lange et al. (84), the individual allele data were not provided, so we could not complete all columns.
an ongoing nationwide prospective cohort study of men aged 40–75 years. The subjects were first enrolled in 1986 and were employed in a variety of health professions. This study used 349 cases and 449 controls, and an odds ratio of 1.92 (95 percent CI: 1.22, 3.03) for BPH was observed for subjects with short alleles (<19 repeats) compared with those with long alleles (≥25 repeats) (57).

The potential public health relevance and the biologic rationale supporting the relation between AR CAG and prostate cancer has helped generate a broad interest in CAG repeats. A recent paper in Epidemiologic Reviews provides additional information about CAG repeats and prostate cancer (77). When results from studies that are investigating the potential associations between trinucleotide repeats and prostate cancer are available, we will incorporate them into the electronic version of this Human Genome Epidemiology review. (Results can be sent directly to Dr. John Witte at witte@darwin.cwru.edu.) Since numerous different cutoffs have been used to distinguish between short and long AR CAG and GGN repeats, we suggest that future reports give crude data that allow for using a range of repeat cutpoints.

INTERACTIONS

Epidemiologic interactions

Some speculate that there could be an interaction between genetic factors and the ratio of estrogen to testosterone. This ratio increases with age and could affect the up-regulation of AR and, ultimately, the slope of age-incidence curves (13). Moreover, androgen deprivation therapy, a common therapy used in the later stages of prostate cancer, has been associated with the AR. In particular, amplification of the AR gene may come about during therapy and encourage tumor cell growth in a low-androgen-concentration environment (78).

Molecular and genetic interactions

Inappropriate RNA-binding protein interaction with specific messenger RNA may disturb cellular dynamics or change the regulation, transport, and expression of CAG containing RNA. This, in turn, may account for the observation of an inverse correlation between CAG repeat length and AR messenger RNA expression in studies involving lower pri- mates. Furthermore, the CAG polyglutamine region may interact with some cellular proteins, and differences in the polyglutamine length could affect the tenacity of such interac- tions (79). There may also be interaction between CAG and GGN repeats. For example, when the subgroup in which both (CAG)n and (GGN)n alleles were short (CAG, <22; GGN, ≤16) was compared with those in which both alleles were long (CAG, ≥22; GGN, >16), an odds ratio of 2.05 (95 percent CI:1.09, 3.84) was observed (11). Finally, there is evi- dence that allelic differences in AR-driven, prostate-specific antigen expression may affect prostate cancer risk (25).

The study of interactions between CAG repeats and environmental and/or genetic mechanisms in prostate cancer is hampered by lack of power. These limitations have been highlighted by Smith and Day (80) and observed in another the Human Genome Epidemiology review by Cotton et al. (81).

Laboratory tests

To determine AR CAG repeat length, extracted undi- gnosed DNA is amplified by polymerase chain reaction (PCR) in a series of two rounds using nested primers sur- rounding the repeat in exon 1 (24). The final products can then be analyzed by electrophoresis on a 5 percent denatur- ing polyacrylamide gel and subjected to autoradiography. One can then obtain the number of CAG repeats from the size of the predominant PCR product relative to a series of previously determined CAG size standards. Subsequently, all samples are ranked by size and reanalyzed by elec- trophoresis and autoradiography, with each allele of an equivalent size next to each other to validate the original size estimates (24). Bharaj et al. (82) improved on this tech- nique somewhat by utilizing a fully automated system for the electrophoretic separation of the PCR product, utilizing fluorescent (nonisotopic) detection of standard-sized mark- ers and using fragment analysis to establish the length of the PCR product. Estimates of sensitivity and specificity for this laboratory test have not been reported. Nevertheless, since the primers for this test are designed specifically for the CAG repeat, the chance of making an error in CAG number counts is very low, and the sensitiv- ity and specificity are extremely high; the main source of error is likely to come from human origins, for example, incorrectly mixing or labeling samples (S. Ingles, personal communication, 1999). Wada et al. (83) describe a promis- ing, but not widely used, alternative to the above method. This involves amplification via PCR, using a simple purifi- cation procedure and counting CAG repeats via matrix- assisted laser desorption/ionization time-of-flight mass spectrometry.

Population testing

On the basis of the evidence summarized here, testing for CAG trinucleotide repeat length in the general popula- tion as part of a population screening program is not presently warranted. This underscores the need for doing additional research on this topic. The association between AR CAG repeat length and prostate cancer is intriguing but is still somewhat limited. To understand prostate carcino- genesis better, we should work to fully explicate any rela- tion among genes such as CAG repeats and environmental factors that may interact to increase prostate cancer risk. The resulting information may ultimately improve our ability to prevent the development of prostate cancer and provide us with better prognostic information following diagnosis.

ACKNOWLEDGMENTS

Supported by grants from the National Institutes of Health (CA88164), DOD Prostate Cancer Research Program (DAMD17-98-1-8589), and the Urologic Research Foundation.
REFERENCES


APPENDIX. INTERNET SITES

Prostate cancer
American Cancer Society: http://www.cancer.org
CaP Cure: http://www.capcure.org
National Institute of Health: http://www.nigms.nih.gov/nci‌
National Cancer Institute: http://www.cancer.gov
National Prostate Cancer Coalition: http://www.4nppc.org
University of Pennsylvania Cancer Center OncoLink: http://oncolink.upenn.edu/disease/prostate

Genetic databases
The Androgen Receptor Gene Mutations Database: http://www.mcgill.ca/androgendb
The Genome Database (GDB): http://gdbwww.gdb.org