How Strong Is the Association Between CAG and GGN Repeat Length Polymorphisms in the Androgen Receptor Gene and Prostate Cancer Risk?

Maurice P. Zeegers,1,2,3 Lambertus A.L.M. Kiemeney,4 Alan M. Nieder,5 and Harry Ostrer6

1Department of Public Health and Epidemiology, School of Medicine, University of Birmingham, Birmingham, United Kingdom; 2Department of General Practice, Comprehensive Cancer Institute Limburg, Catholic University of Leuven, Leuven, Belgium; 3Department of Epidemiology, Faculty of Health Sciences, Maastricht University, Maastricht, Netherlands; 4Department of Urology, School of Medicine, University of Miami, Miami, Florida and 6Human Genetics Program, School of Medicine, New York University, New York, New York

Abstract

Objective: Although narrative reviews have suggested an association between (CAG)n and (GGN)n polymorphisms in the AR gene and prostate cancer, it has never been quantified systematically. The purpose of this meta-analysis was to provide relative and absolute quantitative summary estimates with sufficient power.

Method: Publications were identified through database searches for epidemiologic studies published until February 2004. For each study, mean differences in repeat length between cases and controls were calculated as well as continuous odds ratios (OR) per one repeat length decrement and discrete ORs to compare prostate cancer risk in men with short CAG repeats (<21 repeats) versus long CAG repeats (>21 repeats) and short GGN repeats (<16 repeats) versus long GGN repeats (>16 repeats). The study-specific estimates were combined by random effects meta-regression analyses.

Results: Nineteen case-control studies were included in this review comprising a total of 4,274 cases and 5,275 controls. Prostate cancer cases had on average 0.26 fewer CAG repeats and 0.09 fewer GGN repeats than controls. The continuous ORs of prostate cancer per one decrement were 1.02 and 1.01 for CAG and GGN repeats, respectively. The summary discrete OR (95% confidence interval) were 1.19 (1.07-1.31) and 1.31 (1.06-1.61) for CAG and GGN repeat polymorphisms, respectively.

Conclusion: Although the presence of shorter repeats seemed to be modestly associated with prostate cancer risk, the absolute difference in number of repeats between cases and controls is <1 repeat. We question whether such a small difference is enough to yield measurable biological impact in prostate carcinogenesis. (Cancer Epidemiol Biomarkers Prev 2004;13(11):1765–71)

Introduction

Prostate cancer is the most commonly diagnosed non-cutaneous malignancy among men and the second leading cause of cancer-related deaths in men in Western countries (1). Until today, little is known about its etiology. Only age, family history, and ethnicity are considered to substantially influence prostate cancer risk.

Global ethnic variation contributes to the explanation of the substantial differences in prostate cancer incidence existing throughout the world, being highest in Western countries, especially among African American men, and lowest in developing countries. However, little is known about why certain populations are more susceptible for developing prostate cancer. Some have suggested that varying levels of androgens across ethnic groups may be responsible for these differences (2).

In our recent meta-analysis of 32 population-based studies, we have shown that family members of a prostate cancer patient experience a 2.46-fold increased risk of prostate cancer (3). The nature of this familial clustering is such that the risk is higher for family members with a brother with prostate cancer compared with those with an affected father.

Both observations of ethnic and familial clustering provided the rationale to study whether polymorphisms in the human androgen receptor (AR) gene located at the X chromosome (Xq11.2-q12) may explain this variation in susceptibility to prostate cancer. The AR gene comprises eight exons that encode for four functional domains including the amino-terminal transcription activation (transactivation) domain, the DNA binding domain, a hinge region, and the carboxyl-terminal ligand binding domain (4). The DNA binding domain and ligand binding domain are highly conserved through evolution, whereas the transactivation domain, which regulates expression of target genes, is highly polymorphic. Within this latter domain, there are three microsatellite trinucleotide repeats, two of which are polymorphic in length. Variability in the reported molecular size of the AR is due in part to the upstream

Received 4/20/04; revised 5/11/04; accepted 5/28/04.

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: Maurice P. Zeegers, Department of Public Health and Epidemiology, School of Medicine, Public Health Building, University of Birmingham, Birmingham B15 2TT, United Kingdom. Phone: 44-121-414-6721; Fax: 44-121-414-7878. E-mail: M.P.Zeegers@bham.ac.uk

Copyright © 2004 American Association for Cancer Research.
CAG and downstream GGN repeat polymorphisms, encoding polyglutamine and polyglycine regions, respectively (4-6).

Several studies have suggested that one or both of the trinucleotide repeat polymorphisms in AR may be associated with prostate cancer (7-41). However, to our knowledge, no meta-analysis of all previous studies on this relationship has been conducted thus far. Earlier narrative reviews (4, 42-47) have concluded that prostate cancer risk tend to be highest for men with shorter (CAG)n or (GGN)n alleles. Inconsistency between study results has been ascribed to small sample sizes (low power) or poor research strategies for those studies not observing an association. Although these reviews reported study-specific relative effects [i.e., relative risk or odds ratio (OR)], they did not attempt to quantify these in a summary measure. In addition, the reviews did not report on the absolute mean difference (MD) in number of CAG or GGN repeats between prostate cancer cases and controls.

The objective of this study is to increase power by meta-analyzing all epidemiologic studies up to February 2004 to provide both additive and relative quantitative summary estimates of the association between (CAG)n and (GGN)n repeat polymorphisms in the AR gene and prostate cancer risk and to evaluate changes in summary estimates by differences in study characteristics.

Materials and Methods

Search Strategy. Publications were independently identified by two authors (M.Z. and H.O.) through computerized Medline, Embase, Cancerlit, and Current Contents searches for retrospective or prospective (i.e., nested) studies published until February 2004 with no language restriction. The keywords used were a combination of prostate*, cancer, carcinoma, androgen receptor, polyglutamine, polyglycine, CAG, GGN, AR, and repeat. Furthermore, references cited in published original and review articles (4, 42-47) were examined until no further study was identified.

For inclusion in this analysis, the articles had to present data to estimate an OR or MD and the corresponding standard error (SE), which compares prostate cancer cases with unrelated controls on variation in CAG or GGN repeat length in the AR gene.

A distinction was made between an article and a study. In case the same study population was referred to in more than one article, these articles were considered as part of a single study. On the other hand, if an article presents data of different groups of research subjects, the results of these analyses were seen as separate studies.

Qualitative Data Extraction. Four reviewers (M.Z., L.K., A.N., and H.O.) extracted qualitative information from each article. They independently assessed the following qualitative items: design (retrospective case-control or nested case-control); population characteristics—ethnic label (Hispanic, Caucasian, Asian, or African) and geographic area (country of study); patient characteristics—number, age (years), and histologic grade prostate cancer (percentage advanced prostate cancer, T3-T4, M0; T0-T4, M1); and control characteristics—number, age (years), and disease status [healthy or benign prostatic hyperplasia (BPH)]. When continuous data were presented across different subgroups, we calculated a weighed average using the median of each subgroup weighted by the frequency of observations in these strata. For example, if a population’s age distribution is presented in strata of 10 younger (61-65 years) men and 15 older men (66-70 years), the average age of the population is estimated to be (10 × 63 + 15 × 68) / 25 = 66 years. When disagreement existed on an item, it was discussed until consensus was reached. These qualitative items are used to provide covariables for inclusion in metaregression models to explore reasons for observed heterogeneity in results between the studies.

Quantitative Data Extraction. Two reviewers (M.Z. and H.O.) independently extracted data, allowing us to calculate mean number of repeats in cases and controls, MDs in numbers of repeats between cases and controls, ORs of prostate cancer for short versus long repeats, and ORs of prostate cancer per decrement of one repeat.

When mean number of repeats and corresponding SD for cases and controls were not reported, these were calculated using published tables with frequency distributions of repeats for cases and controls. When these tables were not published, these were reconstructed by measuring published frequency graphs. When no information was available to calculate the SD, it was estimated as a quarter of the repeat range or calculated by using the P of the unpoold l test comparison of means between cases and controls: SEMD = \sqrt{\frac{\text{VAR}_{\text{cases}} + \text{VAR}_{\text{controls}}}{n_{\text{cases}} + n_{\text{controls}}}}, where \text{VAR}_{\text{cases}} is the squared SD for cases and \text{VAR}_{\text{controls}} is the squared SD for controls.

We extracted discrete ORs (ORd) to compare prostate cancer risk in men with short CAG repeats (≤21 repeats) versus long CAG repeats (>21 repeats) and short GGN repeats (≤16 repeats) versus long GGN repeats (>16 repeats). The cutoff values were chosen because these were the pooled median value of repeats in the control populations. If these could not be recapitulated, other nearby cutoff points were used. Preferably, ORd and corresponding SE of the ln(ORd) were extracted directly from the original reports. For studies that reported separate ORs for more than two strata of numbers of repeats, we combined the repeat-specific ORs by using the prevalence of the controls as weight to calculate ORd (48). If these were not available, two-way contingency tables were constructed for each study based on published or reconstructed exposure frequency distributions to calculate the ORd. The SE of latter ln(ORd) was calculated using the method of Woolf (49).

Continuous ORs (ORc) per one CAG or GGN repeat increment (ORc) were preferably extracted directly from the published articles. ORs per one repeat increment were inverted. If ORc and corresponding SE were not
presented, they were derived directly from the estimated $m_{\text{controls}}$, $m_{\text{cases}}$, and $\text{VAR}_{\text{controls}}$ using $\text{OR}_r = \exp \left( \frac{m_{\text{cases}} - m_{\text{controls}}}{\text{VAR}_{\text{controls}}} \right)$ and $\text{SEln(OR)_r} = \frac{\text{SEln(OR)_cases}}{\text{VAR}_{\text{controls}}}$ according to the method of Chene and Thompson (50).

Statistical Analysis. To detect publication or related biases, we explored heterogeneity in funnel plots [i.e., plots of MD against their estimated precision (reciprocal of the variance)]. We examined funnel plot asymmetry visually and measured the degree of asymmetry by using the Egger et al. (51) unweighed regression asymmetry test.

Because of potential heterogeneity between studies, we estimated the summary MD (SMD), summary OR, and summary OR together with their corresponding 95% confidence intervals (CI) with random effects meta-analysis by using the Stata statistical software (52). The between-study variance was estimated by a noniterative procedure using a method of moments estimator. To explore reasons for heterogeneity in MD, we used three methods: (a) we calculated the $I^2$ statistics (i.e., the percentage of variability in point estimates that is due to heterogeneity; ref. 53) for each summary measure of CAG or GGN repeat polymorphisms and across specific strata based on different levels of the extracted study characteristics. Heterogeneity exists when $I^2$ within stratified analyses is higher than within stratified analyses; we did (b) metaregression analyses by including the study characteristics as covariates in the regression model (statistical significance indicates an effect modifier) and (c) stratification.

**Results**

**Study Characteristics.** The search strategy revealed 35 articles reporting epidemiologic studies on the association between AR receptor polymorphisms and prostate cancer (7-41). Thirteen articles were excluded because prostate cancer cases and controls were genealogically related ($n = 3$; refs. 28-30), the association between AR receptor polymorphisms and prostate cancer characteristics was investigated among cases only ($n = 8$; refs. 31-37, 54), or no sufficient data could be extracted ($n = 2$; refs. 38, 39). Three articles were combined in the analysis because the same study was published more often (9-11). Three articles were considered to represent six separate studies because different study populations were investigated (20, 24, 40). The remaining 23 articles described 19 retrospective case-control studies (7-9, 12, 14-18, 20-26, 40) and 5 prospective case-control studies, which were nested in a cohort (8, 13, 19, 27, 41), comprising a total of 4,274 cases and 5,275 controls (Table 1).

All studies were published in English. The majority (79%) of studies were done in Caucasian populations (8-15, 17-27, 40, 41), one study was done among Hispanic men (7), two studies were done among African men (24, 40), and two studies used an Asian population (16, 20). Most studies were done in Western countries, including America ($n = 12$; refs. 7-11, 13, 15, 18, 25-27, 40, 41), Europe ($n = 7$; refs. 12, 17, 19, 20, 22, 23), and Australia ($n = 1$; ref. 21). Two studies were done in Asia (16, 20) and two studies were done in Africa (24). The average mean (SD) age at diagnosis of the patients across

<table>
<thead>
<tr>
<th>Endnote ID</th>
<th>Design*</th>
<th>Population</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ethnicity</td>
<td>Age</td>
<td>Advanced</td>
</tr>
<tr>
<td>Balic et al. (7)</td>
<td>Retrospective</td>
<td>Hispanic</td>
<td>United States</td>
<td>64.0</td>
</tr>
<tr>
<td>Chen et al. (8)</td>
<td>Prospective</td>
<td>Caucasian</td>
<td>United States</td>
<td>61.2</td>
</tr>
<tr>
<td>Ingles et al. (9-11)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>United States</td>
<td>57.8</td>
</tr>
<tr>
<td>Geur et al. (12)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>Austria</td>
<td>65.9</td>
</tr>
<tr>
<td>Giovannucci et al. (13)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>United States</td>
<td>30.7</td>
</tr>
<tr>
<td>Correa-Cerro et al. (14)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>France/Germany</td>
<td>68.2</td>
</tr>
<tr>
<td>Chang et al. (15)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>United States</td>
<td>60.9</td>
</tr>
<tr>
<td>Hsing et al. (16)</td>
<td>Retrospective</td>
<td>Asian</td>
<td>China</td>
<td>72.2</td>
</tr>
<tr>
<td>Mononen et al. (17)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>Finland</td>
<td>68.1</td>
</tr>
<tr>
<td>Modugno et al. (18)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>United States</td>
<td>68.9</td>
</tr>
<tr>
<td>Latil et al. (19)</td>
<td>Prospective</td>
<td>Caucasian</td>
<td>Sweden</td>
<td>69.0</td>
</tr>
<tr>
<td>Ekman et al. (20)</td>
<td>Retrospective</td>
<td>Asian</td>
<td>Japan</td>
<td>71.0</td>
</tr>
<tr>
<td>Beilin et al. (21)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>Australia</td>
<td>67.0</td>
</tr>
<tr>
<td>Chatt et al. (22)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>Sweden</td>
<td>70.2</td>
</tr>
<tr>
<td>Edwards et al. (23)</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>United Kingdom</td>
<td>68.1</td>
</tr>
<tr>
<td>Panz et al. (24)</td>
<td>Retrospective</td>
<td>Retrospective</td>
<td>African</td>
<td>South Africa</td>
</tr>
<tr>
<td>Stanford et al. (25)</td>
<td>Retrospective</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>United States</td>
</tr>
<tr>
<td>Hakimi et al. (26)</td>
<td>Retrospective</td>
<td>Retrospective</td>
<td>United States</td>
<td>62.1</td>
</tr>
<tr>
<td>Platz et al. (27)</td>
<td>Prospective</td>
<td>Caucasian</td>
<td>United States</td>
<td>62.0</td>
</tr>
<tr>
<td>Dos Santos et al. (28)</td>
<td>Retrospective</td>
<td>Retrospective</td>
<td>African</td>
<td>Brazil</td>
</tr>
<tr>
<td>Visvanathan et al. (41)</td>
<td>Prospective</td>
<td>Retrospective</td>
<td>Caucasian</td>
<td>United States</td>
</tr>
</tbody>
</table>

*Retrospective or prospective (=nested) case-control study.

†Percentage advanced prostate cancer (T3-T4, M0; T3-T4, M1).
all studies was 66.2 (4.8) years ranging from 54.9 to 76.0 years. On average, 42.4% of the cases suffered from advanced prostate cancer. The average percentage of cases with a Gleason score ≥7 across all studies was 32.6% (data not shown). The average mean (SD) age of controls was 63.7 (6.7) years ranging from 54.0 to 73.6 years. Most studies selected healthy controls, although three studies selected controls with BPH (12, 20; Table 1).

Publication Bias. We could not identify heterogeneity in funnel plots neither visually (Fig. 1) nor in terms of statistical significance ($P = 0.39$ and 0.96 for CAG and GGN repeats, respectively).

CAG Repeats. Nineteen studies conveyed data on the mean number and SD of CAG repeats among cases and controls (7, 8, 12-17, 19-27, 40; Fig. 2). Statistical pooling showed that prostate cancer cases seemed to have on average 0.26 fewer CAG repeats than controls (SMD, $-0.26$; 95% CI, $-0.53$ to 0.00). The summary ORd (95% CI) comparing men with ≤21 CAG repeats to men with >21 CAG repeats was 1.19 (1.07-1.31) based on the results of 22 studies (7-26, 40, 41). The ORc (95% CI) of prostate cancer per one CAG decrement was 1.02 (0.99-1.06) based on the results from the same studies that were used to calculate the SMD (7, 8, 12-17, 19-27, 40; Fig. 2).

GGN Repeats. The calculation of the SMD for the number of GGN repeats was based on eight studies (8, 14-16, 23, 25-27; Fig. 2). On average, prostate cancer cases had 0.09 fewer GGN repeats than controls (SMD, $-0.09$; 95% CI, $-0.20$ to 0.03). Men with ≤16 GGN repeats seemed to have a 31% higher risk of prostate cancer than...
Table 2. $I^2$ for CAG repeats both unstratified and stratified for specific study characteristics, $P$ of interaction for individual study characteristics resulting from metaregression analysis, and MD and corresponding 95% CI per stratum

<table>
<thead>
<tr>
<th>No. studies</th>
<th>MD (95% CI)</th>
<th>$I^2$ (%)</th>
<th>$P_{interaction}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstratified</td>
<td>-0.26 (-0.53 to 0.00)</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Stratified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prospective</td>
<td>-0.08 (-0.53 to 0.37)</td>
<td>61</td>
<td>0.52</td>
</tr>
<tr>
<td>Retrospective</td>
<td>-0.31 (-0.63 to 0.01)</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Ethnicity of study population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>-0.10 (-0.26 to 0.07)</td>
<td>22</td>
<td>0.02</td>
</tr>
<tr>
<td>Other</td>
<td>-0.67 (-1.94 to 0.60)</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Country of study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>-0.46 (-1.17 to 0.24)</td>
<td>86</td>
<td>0.31</td>
</tr>
<tr>
<td>Other</td>
<td>-0.16 (-1.37 to 0.05)</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Mean age of cases (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65</td>
<td>-0.55 (-1.46 to 0.35)</td>
<td>90</td>
<td>0.37</td>
</tr>
<tr>
<td>$\geq$65</td>
<td>-0.17 (-0.35 to 0.02)</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Advanced prostate cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50%</td>
<td>-0.18 (-0.50 to 0.13)</td>
<td>48</td>
<td>0.90</td>
</tr>
<tr>
<td>$\geq$50%</td>
<td>-0.29 (-0.69 to 0.12)</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Mean age of controls (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65</td>
<td>-0.60 (-1.71 to 0.51)</td>
<td>93</td>
<td>0.41</td>
</tr>
<tr>
<td>$\geq$65</td>
<td>-0.18 (-0.37 to 0.02)</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Disease status of controls*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>-0.35 (-0.65 to -0.04)</td>
<td>74</td>
<td>0.41</td>
</tr>
<tr>
<td>BPH</td>
<td>-0.01 (-0.63 to 0.64)</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

*The disease status of the controls in one study (8) could not be recapitulated.

men with $>16$ GGN repeats (ORd, 1.31; 95% CI, 1.06-1.61). This estimate is based on the results of eight studies (8-11, 15, 16, 23, 25-27). The ORd (95% CI) of prostate cancer per one GGN decrement was 1.01 (0.98-1.04) based on the results from the same studies that were used to calculate the SMD (8, 14-16, 23, 25-27; Fig. 2).

**Heterogeneity.** We further examined heterogeneity in study results by categories of study characteristics including study design, ethnicity of study population, country of study, study age of cases and controls, severity of disease, and type of disease status of controls. The definitions of the specific categories can be found in Table 2. The analyses were limited to studies reporting the association between number of CAG repeats and prostate cancer risk only, because the data on GGN repeats were too sparse. Except for the ethnicity of the study population ($I^2 = 0.02$), none of the study characteristics seemed to be a statistically significant interaction term in metaregression analyses. However, quantitative differences in stratified MDs in CAG repeats were found. The MD in number of CAG repeats between cases and controls seemed to be larger in retrospective (-0.31) versus prospective (-0.08) case-control studies, non-Caucasian (-0.67) versus Caucasian (-0.10) study populations, the United States (-0.046) versus other countries (-0.16), younger (-0.55) versus older (-0.17) cases, more (-0.29) versus less (-0.18) cases with advanced prostate cancer, younger (-0.48) versus older (-0.18) controls, and healthy (-0.35) versus BPH (-0.01) controls. However, within all subgroups, still a considerable amount of the variation (17-93%) could be explained by random heterogeneity. Furthermore, none of the stratified results showed a meaningful difference in CAG repeat between cases and controls. All differences were <1 repeat (Table 2).

**Discussion**

Several biological observations have led to the expectation that $AR$ repeat polymorphisms may be associated with susceptibility to developing prostate cancer. First, it is known that androgens play a role in human prostate cancer growth, because dogs and males castrated before puberty do not get prostate cancer and because androgen-ablative therapy initially inhibits prostate tumor growth (46, 55). Second, it has been observed that the length of the CAG repeat region is inversely related to the transactivation activity on the receptor and the binding affinity for androgens (56-58). Irvine et al. (59) reported that increased CAG length inhibited both basal and coactivator-mediated $AR$ transactivation activity in cultured prostate epithelial cells that were cotransfected with $AR$ constructs of up to 42 repeats and p160 coactivator expression vectors. Therefore, decreased repeat length may make the prostate more vulnerable to chronic androgen stimulation leading to increased proliferative activity, which in turn may increase the rate of somatic mutations (46). Third, $AR$ repeat polymorphisms have been shown to be associated with the incidence of other androgen-related clinical conditions; specifically, a high number of repeats seem to adversely influence fertility and spermatogenesis (60) and bone density (61), and fewer repeats are associated with increased risk of baldness (62) and benign prostatic hyperplasia (63-66), although this has not always been confirmed (67).

In this meta-analysis, we conclude that the relative epidemiologic associations derived from previous literature confirm these biological expectations. We have shown that men with shorter CAG and GGN repeats experience a 1.19- and 1.31-fold increased risk of prostate...
cancer, respectively, although these increased risks should be interpreted as being very small. It should be noted that it is possible that CAG and GGC repeats are in linkage disequilibrium and it should therefore be considered that effects seen at one microsatellite might reflect effects at the other.

The results from the heterogeneity analyses were to some extent consistent with previous literature. For example, the association between CAG repeat polymorphism and prostate cancer seemed to be somewhat larger when controls were compared with cases with more advanced prostate cancer or with cancer that was diagnosed at an earlier age. According to Giovannucci (44), this could suggest that a heightened state of androgenicity directly influences a pool of relatively aggressive, early-onset androgen-driven prostate cancers. Possibly, prostate cancers that occur at older ages may be less driven by androgenicity and more related to pathologic processes such as oxidative insults (44). In addition, the MD in number of CAG repeats between cases and controls seemed to be larger for non-Caucasians and study populations from developing countries. Black males have a population average of 18 repeats, whereas Caucasian men have on average 20 repeats (47). Combining this with the fact that there is an inverse relationship between CAG length and AR transactivation function, such that the shorter the CAG, the better the AR is at activating gene transcription (68), Coetzee and Ross (57) hypothesized that the racial differences seen in prostate cancer susceptibility may be due in part to variability in the AR CAG repeat length (4). However, because the association between AR repeat length polymorphisms and prostate cancer is small, other factors must be involved as well.

The incidence of prostate cancer has increased in several countries because the availability of prostate-specific antigen testing in the late 1980s in the United States and early 1990s in other countries. As pointed out by Giovannucci (44), pre–prostate-specific antigen era patients might differ significantly from post–prostate-specific antigen era patients. However, in this meta-analysis, all studies have been conducted after the introduction of prostate-specific antigen screening. Because prostate-specific antigen screening of the general population is of more common use in the United States and less often recommended and carried out by other countries, the larger SMD in CAG repeat polymorphism between prostate cancer cases in the United States compared with other countries may be attributed to differences in screening policy.

Although previous epidemiologic studies have consistently reported that the magnitude of relative association between AR repeat polymorphisms and prostate cancer is relatively limited, most narrative reviews to date have suggested a positive association between shorter CAG or GGN repeat lengths and prostate cancer. Some have argued that even a modest increased risk could still be of importance from a public health perspective because of the large population frequency of short repeat lengths (46). Based on this reasoning, Nelson and Witte (46) suggested that this potentially low-penetrance, high-frequency polymorphism could theoretically account for many more prostate cancer cases than a high-penetrance, low-frequency mutation for which the etiologic fraction is estimated to be 9%. This may be true on a population scale for the GGN polymorphism but not for the CAG polymorphism. We have estimated that the etiologic fraction is 13.4% and 8.7% for GGN and CAG repeat polymorphisms, respectively, assuming a short repeat frequency of 50% (because median OR values were used) and the reported summary ORs.

However, our analyses also show that although cases seemed to have shorter repeats and shorter repeats seemed to be modestly associated with prostate cancer risk on a relative scale, the absolute difference in number of repeats between cases and controls is <1 repeat. The SMDs in number of CAG and GGN repeats were −0.27 and −0.09, respectively. Although the direction of results in the heterogeneity analysis seemed to be consistent with previous literature, none of the stratified differences between cases and controls exceeded one repeat. Based on these results, we question whether an average MD of <1 repeat is enough to yield measurable biological impact in prostate carcinogenesis.

References


47. Sta statistical software: release 6.0. College Station (TX): Stata Corporation.


How Strong Is the Association Between CAG and GGN Repeat Length Polymorphisms in the Androgen Receptor Gene and Prostate Cancer Risk?


Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/13/11/1765

Cited articles
This article cites 67 articles, 15 of which you can access for free at:
http://cebp.aacrjournals.org/content/13/11/1765.full#ref-list-1

Citing articles
This article has been cited by 7 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/13/11/1765.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.