Androgen Receptor Variants with Short Glutamine or Glycine Repeats May Identify Unique Subpopulations of Men with Prostate Cancer

Janette M. Hakimi, Mark P. Schoenberg, Rachel H. Rondinelli, Steven Piantadosi, and Evelyn R. Barrack

Department of Urology [J. M. H., M. P. S., R. H. R., E. R. B.] and The Oncology Center [S. P., E. R. B.], The Johns Hopkins University School of Medicine, Baltimore, Maryland 21287

ABSTRACT

The androgen receptor (AR) contains glutamine (CAG) and glycine (GGC) repeats that are each polymorphic in length. We screened clinically localized prostate cancers for somatic mutations in the length of the CAG and GGC repeats in the AR gene and characterized the length of these repeats in the germ-line AR gene. Somatic mutations were rare, and the range of germ-line repeat lengths in men with prostate cancer was within the range of normal in the general population. Most allele frequencies in Caucasian men with clinical prostate cancer were remarkably comparable to those in the general Caucasian population. However, a subpopulation of the men with clinical prostate cancer had a substantially higher frequency of AR alleles with 16 or 17 CAGs (6 of 59 men, 10%) than did the general population (6 of 370 alleles, 1.6%), and a different subpopulation of the men with prostate cancer had a higher frequency of AR alleles with 12 or 13 GGCs (7 of 54 men, 13%) than did the general population (1 of 110 alleles, 0.9%). Of the men with prostate cancer who had an AR gene with 16 or 17 CAGs, 83% had lymph node-positive disease, despite the lack of clinical evidence of metastatic spread. This suggests that a short AR CAG allele may be a risk factor for the development of clinically unsuspected lymph node-positive prostate cancer among men undergoing radical prostatectomy and raises the question of whether this short repeat length played an active role in the development of aggressive prostate cancer. The odds of having a germ-line AR gene with a short CAG repeat (≤17 CAGs) were substantially higher in Caucasian men with lymph node-positive prostate cancer than in Caucasian men with lymph node-negative disease or in the general Caucasian population. The odds of having a short germ-line AR CAG were the same for men with lymph node-negative prostate cancer as for the general Caucasian population. The odds of having a germ-line AR gene with a short glycine repeat (≤14 GGCs) were substantially higher in men with prostate cancer than in the general population, but the frequency of alleles with a short GGC repeat was the same in men with lymph node-positive versus lymph node-negative disease. This suggests that a short GGC repeat may be a risk factor for the development of clinical prostate cancer, a hypothesis that needs to be tested in cohort and case-control studies.

INTRODUCTION

Prostate cancer is the most commonly diagnosed cancer in men in the United States, accounting for 36% of new cancer cases in men, and it is the second most common cause of cancer deaths in this group (1). For 1995, it was estimated that there would be 244,000 new cases of prostate cancer and 40,000 deaths from prostate cancer (1). The disparity between the number of new cases and the number of deaths per year reflects the fact that a large percentage of men who have prostate cancer do not die from this disease (2), likely because there are different types of prostate cancer that differ in their natural history. The majority of men with prostate cancer have no known risk factors for disease development or to predict whether their tumor will remain indolent, progress slowly, or be aggressive, becoming high grade or metastatic even while still small (3–5). With the advent of screening, increased numbers of prostate cancer cases are being diagnosed (6, 7), thereby increasing the need for indicators of risk assessment.

The etiology of prostate cancer is unknown, but there is compelling evidence to support the hypothesis of a hormonal etiology involving androgen action (8–10). Androgen, acting via the AR, is required for differentiation and growth of the prostate in utero and at puberty (8). It is also presumed to play a role in prostate carcinogenesis, because men castrated before puberty do not develop prostate cancer, and because androgen is required for the experimental induction of prostate cancer in animals (8–10). We have been investigating the possibility that androgen might mediate abnormal prostate growth via a mutated AR gene (11–13). On the basis of available data, somatic AR gene mutations occur in only a subset of early-stage clinical prostate cancers (11–13). Therefore, we have concluded that the

1 Supported by National Cancer Institute Grants CA58236 and CA69645. A preliminary report of this work was presented at the 8th Annual Meeting of the Society for Basic Urologic Research, San Francisco, California, May 13–14, 1994 (56).

2 To whom requests for reprints should be addressed, at The Johns Hopkins Hospital-Medical School, 600 North Wolfe Street, Baltimore, MD 21287-2011. Phone: (410) 955-4491; Fax: (410) 955-0833; E-mail: ebarrack@ghate1.jhmi.edu.

3 The abbreviations used are: AR, androgen receptor; HD, Huntingtin; SBMA, spinal and bulbar muscular atrophy.
role of androgen in prostate carcinogenesis, progression to clinical disease, and growth of prostate tumor cells must be mediated predominantly by normal, wild-type AR (13).

The AR is a member of the steroid receptor superfamily of ligand-dependent nuclear transcription factors (Fig. 1). The AR protein consists of three functional domains: (a) the COOH-terminal hormone-binding domain confers ligand specificity; (b) the central DNA-binding domain binds to androgen-responsive target genes; and (c) the NH2-terminal domain affects transcriptional efficiency. Deletion of the NH2-terminal domain renders the AR transcriptionally inactive, despite its ability to bind androgen with high affinity (14, 15). The NH2-terminal domain interacts with the hormone-binding domain; this stabilizes the AR by slowing the rate of ligand dissociation and AR degradation (16, 17).

There is only one AR gene (located at Xq11-q12; Ref. 18), but there are multiple allelic variants of the AR gene in the general population; i.e., the AR is highly polymorphic (19–25). This polymorphism affects the protein coding sequence of the NH2-terminal domain of the AR, such that it contains a glutamine repeat and a glycine repeat that each vary in length among individuals in the general population (Fig. 1). The glutamine repeat, (Gln)n (Gln), starts at amino acid 58 and is encoded by a CAG repeat that varies in length (Ref. 20: 11–31 CAGs, n = 792 AR alleles; Ref. 22: 10–35 CAGs, n = 410; Ref. 23: 14–31 CAGs, n = 110; Ref. 24: 7–30 CAGs, n = 243; Ref. 25: 9–29 CAGs, n = 122). (Gly)n (Gly) is the polymorphic portion of the glycine repeat, and it is encoded by a GGC repeat that varies in length (Ref. 21: 10–18 GGCs, n = 73; Ref. 22: 4–24 GGCs, n = 410; and Ref. 25: 8–17 GGCs, n = 115).

Fig. 1 Schematic of the human AR protein. HBD, hormone-binding domain; DBD, DNA-binding domain. (Gln)n, is the polymorphic portion of the glutamine repeat, which is encoded by a CAG repeat that varies in length (Ref. 20: 11–31 CAGs, n = 792 AR alleles; Ref. 22: 10–35 CAGs, n = 410; Ref. 23: 14–31 CAGs, n = 110; Ref. 24: 7–30 CAGs, n = 243; Ref. 25: 9–29 CAGs, n = 122). (Gly)n (Gly) is the polymorphic portion of the glycine repeat, and it is encoded by a GGC repeat that varies in length (Ref. 21: 10–18 GGCs, n = 73; Ref. 22: 4–24 GGCs, n = 410; and Ref. 25: 8–17 GGCs, n = 115).

CAG repeat length modulates AR-mediated effects of androgen (30–32). CAG repeat lengths progressively greater than normal (>35 CAGs) progressively decrease the transcriptional activity of the AR (30, 31), and deletion of the glutamine repeat creates an AR with increased transcriptional activity (31). Longer CAG repeat lengths have also been reported to reduce AR mRNA and protein expression (32). The AR glutamine repeat is about twice the normal length (40–62 CAGs) in the germ-line DNA of men with X-linked SBMA (also known as Kennedy’s disease), an adult-onset degenerative disease that affects motor neurons in the spinal cord and brainstem (33). Androgen action is indeed blunted in men with this AR defect, because the ability of a synthetic androgen to lower plasma testosterone, luteinizing hormone, and follicle-stimulating hormone levels is significantly lower in SBMA patients than it is in controls (34), and men with SBMA exhibit signs of mild androgen insensitivity at adulthood, such as gynecomastia and infertility (35). The lower effectiveness of androgen in SBMA thus may be due to decreased AR activity (30) and/or decreased AR expression (32) associated with the longer-than-normal CAG (glutamine) repeat.

We speculate that even the normal range of CAG repeat lengths may affect AR function and androgen action. Because androgen plays an important role in development, growth, and behavior, normal AR polymorphism could help to account for variations in the degree of androgenization and manifestations of androgenic activity in the general population. Potential differences in the AR activity of normal AR alleles could even affect the role of androgen in prostate cancer. Glutamine repeats also affect the transcriptional activity of other transcription factors (35). To our knowledge, the role of glycine repeats has never been investigated.

The frequency of AR alleles with different CAG and GGC repeat lengths has been measured in the germ-line DNA of the general population, i.e., unrelated apparently healthy individuals not selected for any particular phenotype, except race (20, 21, 24, 25). In whites (Caucasians), the AR CAG allele frequency distribution has major peaks at 21 CAGs and at 24 CAGs (based on the combined data from Refs. 20, 24, and 25). In blacks (Africans and African Americans), the frequency distribution has major peaks at 18 CAGs and 22 CAGs (20, 24, 25). We calculated the cumulative allele frequencies in whites versus blacks (using raw data obtained from Refs. 20, 24, and 25) and found that the median CAG repeat length is 21 CAGs in whites versus 19 CAGs in blacks (P = 0.0001). In addition, based on data reported previously (21, 25), we determined that the median GGC repeat length also is significantly shorter in African Americans (14–15 GGCs) than it is in Caucasians (16 GGCs; P = 0.0001).

If ARs with shorter repeats have more activity, then the effect of androgen may be greater in individuals with such AR alleles. This could help to account for the apparently higher age-specific reference ranges of serum prostate-specific antigen levels in African-American men without prostate cancer than in Caucasian men without prostate cancer (36). If short repeats enhance AR activity, and AR activity is required for prostate carcinogenesis, then the higher frequency of short repeats in African Americans could also help to account for the higher incidence, higher mortality rate, and more aggressive nature of prostate cancer in African Americans than in Caucasians (9, 10, 37, 38). We previously reported finding a somatic mutation in the length of the AR CAG repeat in a Caucasian man with clinically localized stage B prostate cancer (12). The nontumor tissue had an AR gene with 24 CAGs, whereas his tumor contained cells with 24 CAGs (wild type) and cells with 18 CAGs (mutant), representing a somatic in-frame contraction of the AR CAG repeat (CAG24→CAG18) in the tumor (12). By coincidence,
this somatically mutated CAG₁₈ AR is the most common allele length in blacks (20, 24). Because this mutation was present in a large proportion of the tumor, we reasoned that it might have conferred a growth advantage on the cell in which it originally occurred (12), especially if, for example, ARs with a shorter repeat were a more effective mediator of androgen action.

In the present study, we screened prostate cancers for somatic mutations in the AR CAG and GGC repeats and characterized the lengths of these repeats in the germ-line AR gene. We studied only Caucasian men.

**MATERIALS AND METHODS**

Prostate Cancer Patients. Genomic DNA was obtained from Caucasian men undergoing radical retropubic prostatectomy at The Johns Hopkins Hospital for the treatment of clinically localized stage B (organ-confined) prostate cancer. Patients were from different parts of the country. The patients had no clinical evidence of metastatic disease, but at the time of staging lymphadenectomy, some were discovered to have prostate cancer in their pelvic lymph nodes. Tumor and normal tissues (peripheral lymphocytes, nonmalignant prostate, or seminal vesicles) obtained at the time of surgery were processed by cryostat sectioning and digested with proteinase K (39). DNA was extracted and purified by the phenol-chloroform method, and the concentration was determined. Both tumor DNA and nontumor DNA were analyzed. Thirty-four of these men had pathological stage D disease, i.e., they were lymph node-negative. Twenty-five patients had pathological stage D disease, i.e., they were lymph node-positive. Most tumors (n = 51) had a Gleason score of 7, 8, or 9; one patient with lymph node-positive disease had a Gleason score of 6, and seven patients with lymph node-negative disease had a Gleason score of 5 or 6.

AR CAG Repeat Length Analysis. The glutamine repeat in the AR protein, (Gln)₇₋₃₅(Gln), is encoded by the CAG repeat, which is adjacent to a CCG repeat, was achieved by an independent PCR reaction that contained fluorescent sequencing. In addition, the length of several samples was verified by direct DNA sequencing, one with 18 CAGs and the other with 24 CAGs. The average of six independent measurements of CAG repeat length based on PCR product size analysis was 18.2 ± 0.5 (SE) for the 18 CAG internal control and 24.0 ± 0.4 for the 24 CAG internal control. Variation results from the fact that DNA fragments of a given length may not migrate at the same rate in different regions of a given gel. Nevertheless, specimens with a repeat of 17 CAGs always ran faster than the 18 CAG standard. Therefore, we infer that our measurements of CAG repeat length are accurate.

AR GGC Repeat Length Analysis. The glycine repeat in the AR protein, (Gly)₉₋₁₄(Gly), is encoded by (GGT)₂₄₋₃₄(GGT), an invariant six-glycine (GGT/ GGG) repeat followed by a GGC repeat that is polymorphic in length (26, 28, 29). PCR primers JM113 and JM114 (22) were used to amplify the polymorphic GGC repeat. PCR reactions (25 μl) contained 100 ng of genomic DNA; 20 mm Tris-HCl (pH 8.2); 10 mm KCl; 6 mm (NH₄)₂SO₄; 1.5 mm MgCl₂; 0.1% (w/v) Triton X-100; 10% DMSO; 200 nm each primer (denatured at 95°C for 5 min before addition); 0.2 mm dATP, dTTP, and dCTP; 0.2 mm 7-deaza-dGTP/dGTP (3:1 ratio); 2.5 μCi of [α-³²P]dCTP; and 2.5 units of Pfu polymerase (Stratagene, La Jolla, CA). Amplification conditions were as follows: 94°C for 10 min; 96°C for 1 min; 40 cycles of 61°C for 1 min, 75°C for 2 min, and 96°C for 1 min; and a final cycle of 61°C for 1 min and 75°C for 5 min. Samples were denatured and analyzed on denaturing gels as described for the CAG repeat. To test the accuracy of our repeat length determinations, we always included in our analysis a PCR product with 18 GGCs, the length of which had been independently verified by direct DNA sequencing. In addition, the length of several samples was verified by an independent PCR reaction that contained fluorescent primers and was analyzed in the presence of fluorescent standards in each lane (The Johns Hopkins University Genetic Resources Core Facility).

**HD CAG Repeat Length Analysis.** The HD protein contains a polymorphic glutamine repeat that is encoded by (CAG)ₙ (n = 9–30; Ref. 40). Efficient amplification of the HD CAG repeat, which is adjacent to a CCG repeat, was achieved...
using primers that flank the repeat and PCR conditions that were the same as those used to amplify the AR GGC repeat.

Statistical Analysis. Fisher's exact test (StatXact, Version 3 for Windows; Cytel Software, Cambridge, MA) was used to analyze 2×2 tables. Odds ratios were calculated to determine the relative likelihood of having certain AR alleles. Linkage equilibrium between CAG and GGC repeat lengths was tested using the following relationships: \( P(AB) = P(A) \times P(B) \); \( P(ab) = [1 - P(A)] \times [1 - P(B)] \); \( P(aB) = [1 - P(A)] \times [1 - P(B)] \); and \( P(ab) = [1 - P(A)] \times [1 - P(B)] \) in which \( P \) is allele frequency, \( A \) is a short CAG allele (≤17 CAGs), \( a \) is a long CAG allele (>17 CAGs), \( B \) is a short GGC allele (≤14 GGCs), and \( b \) is a long GGC allele (>14 GGCs); \( P(A), [1 - P(A)], P(B), \) and \( [1 - P(B)] \) are the frequencies of alleles with a short CAG, long CAG, short GGC, and long GGC, respectively; and \( P(AB), P(ab), P(aB), \) and \( P(ab) \) are the frequencies of alleles with short CAG/short GGC, short CAG/long GGC, long CAG/short GGC, and long CAG/long GGC, respectively.

General Caucasian Population. The allele frequencies of germ-line AR gene CAG repeat length in the general Caucasian population were obtained by combining data from three samplings of the general population of unrelated apparently healthy Caucasians from different geographic areas \( (n = 240, \) Ref. 20; \( n = 91, \) Ref. 24; and \( n = 39, \) Ref. 25). These three data sets were combined \( (n = 370) \) to generate a composite CAG allele frequency distribution. [A fourth data set was excluded because it represented a compilation of data from Caucasians and African Americans (22). Data on CAG repeat length in African Americans cannot be combined with data on repeat length in Caucasians because the frequency distributions are significantly different from each other (see “Introduction”).]

The allele frequencies of germ-line AR GGC repeat length in the general Caucasian population were also obtained from published studies. The frequency distributions in two samplings of the general population of unrelated apparently healthy Caucasians, one based on 73 individuals (Ref. 21) and another based on 37 individuals from a different geographic area (Ref. 25), were combined \( (n = 110) \) to generate a composite GGC allele frequency distribution.

It was not possible to determine repeat length frequencies in a control population of men without prostate cancer. Such a control population would be extraordinarily difficult to identify because prostate cancer is so common. Nine percent of men develop clinical evidence of prostate cancer within their lifetime, and 50% of men >70 years old have histological evidence of prostate cancer, the potential clinical aggressiveness of which is not known (6). Therefore, in a general population of apparently healthy men without clinical evidence of prostate cancer, a large percentage actually have prostate cancer.

RESULTS

Comparison of Repeat Length in Tumor and Nontumor DNA. In the present study, by comparing the mobility of PCR products from prostate cancer and nontumor DNA from the same patient in adjacent lanes, we found no somatic mutations in 59 patients analyzed for AR CAG (glutamine) repeat length mutations or in 54 patients analyzed for AR GGC (glycine) repeat length mutations. Therefore, in the present patient sam-

Fig. 2 Frequency of AR CAG allele lengths in germ-line DNA. A, frequency of alleles in 59 Caucasian patients with prostate cancer (Total PCa, ○), compared to the general Caucasian population (●, \( n = 370 \) AR alleles; data combined from Refs. 20, 24, and 25). B, frequency of alleles in men with clinically unsuspected lymph node-positive prostate cancer (LN+, ○, \( n = 25 \) men) compared to men with lymph node-negative prostate cancer (LN−, ○, \( n = 34 \) men). C, cumulative frequency plot of germ-line AR CAG repeat lengths in Caucasian men with lymph node-positive prostate cancer (LN+, ○, \( n = 25 \) men) compared to men with lymph node-negative prostate cancer (LN−, ○, \( n = 34 \) men). The cumulative frequency is the frequency of alleles with a CAG repeat length less than or equal to a given value.
frequencies as the general Caucasian population (Fig. 2C); this node-positive prostate cancer (n patients into two groups based on whether they had lymph node-positive disease. We divided our prostate cancer but at the time of surgery, some of these men are found to have prostate cancer (i.e. more aggressive disease than suspected. 

clinically localized prostate cancer (Total PCa), Caucasian men with repeat in the germ-line AR gene are shown for all Caucasian men with clinically localized prostate cancer (Total PCa), Caucasian men with lymph node-negative prostate cancer (LN− PCa), Caucasian men with lymph node-positive prostate cancer (LN+ PCa), and the general Caucasian population (Gen Pop) reported in Refs. 20, 24, and 25.

Table 1 Short AR CAG repeat as a potential risk factor in prostate cancer

<table>
<thead>
<tr>
<th></th>
<th>Short ≤17 CAG</th>
<th>Long &gt;17 CAG</th>
<th>Odds Ratioa (CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total PCa</td>
<td>6</td>
<td>53</td>
<td>3.7 (1.3−10.5)</td>
<td>0.02</td>
</tr>
<tr>
<td>Gen Pop</td>
<td>11</td>
<td>359</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN− PCa</td>
<td>1</td>
<td>33</td>
<td>1.0 (0.1−8.1)</td>
<td>1.0</td>
</tr>
<tr>
<td>Gen Pop</td>
<td>11</td>
<td>359</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN+ PCa</td>
<td>5</td>
<td>20</td>
<td>8.2 (2.6−26)</td>
<td>0.002</td>
</tr>
<tr>
<td>Gen Pop</td>
<td>11</td>
<td>359</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers of individuals with either a short or long CAG repeat in the germ-line AR gene are shown for all Caucasian men with clinically localized prostate cancer (Total PCa), Caucasian men with lymph node-negative prostate cancer (LN− PCa), Caucasian men with lymph node-positive prostate cancer (LN+ PCa), and the general Caucasian population (Gen Pop) reported in Refs. 20, 24, and 25.

The odds ratio is equal to the ratio of incidence of short alleles to incidence of long alleles in one group of individuals divided by the same ratio of alleles in a second group of individuals. The 95% confidence interval (CI) for the odds ratio is given in parentheses.

P values were determined using Fisher’s exact test.

Men undergo radical prostatectomy for clinically localized prostate cancer (i.e., no clinical evidence of metastatic spread), but at the time of surgery, some of these men are found to have lymph node-negative disease. We divided our prostate cancer patients into two groups based on whether they had lymph node-negative prostate cancer (n = 25) or lymph node-negative disease (n = 34; Fig. 2, B and C). Caucasian men with lymph node-negative disease had virtually the same AR CAG allele frequencies as the general Caucasian population (Fig. 2C); this argues against a selection bias in our patient sample. The prevalence of germ-line AR alleles with ≤17 CAGs was 20% (5 of 25) in men with lymph node-positive prostate cancer, compared to 2.9% (1 of 34) in men with lymph node-negative disease and 3% (11 of 370) in the general Caucasian population.

Fig. 3 Frequency of AR GGC repeat lengths in germ-line DNA. A, frequency of alleles in Caucasian patients with prostate cancer (Total PCa, n = 54 men) compared to the general Caucasian population (○, combined data from Refs. 21 (n = 73) and 25 (n = 37)). B, cumulative frequency plot of AR GGC repeat lengths in Caucasian men with prostate cancer (Total PCa, ○, n = 54), Caucasian men with lymph node-negative disease (●, n = 30), Caucasian men with lymph node-positive disease (□, n = 24), and the general Caucasian population (●, n = 110).

The odds of having a germ-line AR gene with a short CAG repeat (≤17 CAGs) were eight times higher in Caucasian men with lymph node-positive prostate cancer than they were in Caucasian men with lymph node-negative disease (see Table 1). The odds of having a short germ-line AR CAG repeat were four times higher in men with prostate cancer (regardless of lymph node status) than they were in the general population (P = 0.02). These increased odds were virtually completely accounted for by men who had developed clinically unsuspected lymph node-positive disease. The odds of having a short germ-line AR CAG were significantly higher in men with lymph node-positive prostate cancer than they were in the general
population (odds ratio = 8; P = 0.002) but were not different for men with lymph node-negative prostate cancer compared to the general Caucasian population (odds ratio = 1; P = 1.0).

Thus, a short AR CAG allele may be a risk factor for the development of clinically unsuspected lymph node-positive prostate cancer among men undergoing radical prostatectomy. A prospective study of men undergoing radical prostatectomy for apparently localized disease will need to be done to test this hypothesis. In addition, cohort studies will need to be done to determine the relative risk of developing lymph node-positive disease among men in the general population with a short AR CAG allele in their germ-line DNA.

The hypothesis has been put forth that prostate cancer that is diagnosed in younger men may be more aggressive than disease diagnosed in older men (41). Therefore, we tested the possibility that the higher frequency of short AR CAG alleles in men who developed lymph node-positive disease reflected a higher frequency of short alleles in men who developed prostate cancer at a younger age. We divided our prostate cancer patients into two groups on the basis of age at the time of diagnosis, ≤60 years of age versus >60 years. The mean ± SE age of the young group (≤60 years) was 55.8 ± 1.2 years (n = 21 men; range, 38–60 years; median, 57–58 years). The mean age of the old group was 66.1 ± 0.6 years (n = 33; range, 61–71 years; median, 66–67 years). Men diagnosed at ≤60 years of age did not have a higher frequency of short alleles but rather had the same frequency of AR CAG alleles as men diagnosed at >60 years of age (data not shown). The odds of having a short germ-line AR CAG (≤17 CAGs) were the same for the two age groups (odds ratio = 1; P = 1.0). Therefore, short AR CAG alleles seem not to affect the odds of being diagnosed at a younger age, when younger age is defined as ≤60 years.

**Germ-Line AR GGC Repeat Length Analysis.** AR GGC repeat lengths in the germ-line DNA of 54 Caucasian patients with prostate cancer ranged from 10 to 22; hence, they were within the range of normal in the general population (4–24 GGCs; see Fig. 1). The frequency distribution of GGC alleles is shown in Fig. 3A; there was a major peak at 16–17 GGCs and a minor peak at 12–13 GGCs. The frequency of AR alleles with different GGC repeat lengths in the general Caucasian population is also shown. The major peak at 16–17 GGCs in men with prostate cancer completely overlaps the major peak at 16–17 GGCs in the general population (Fig. 3A). However, men with prostate cancer had a 4-fold higher frequency of short GGC alleles (12–13 GGCs; Fig. 3A); the prevalence of alleles with ≥14 GGCs was 14.8% (8 of 54) in men with prostate cancer, compared to 3.6% (4 of 110) in the general population. The odds of having a GGC repeat length ≤14 were significantly higher in men with prostate cancer than in the general population (odds ratio = 4.6; P = 0.02; Table 2).

GGC repeat length distributions were not significantly different between men with lymph node-positive and men with lymph node-negative prostate cancer (Fig. 3B; Table 2). The frequency of a short GGC allele (≤14 GGCs) was 16.7% (5 of 30) in men with lymph node-negative disease and 12.5% (3 of 24) in men with lymph node-positive disease (Fig. 3B). In addition, GGC allele frequencies were not different between men with prostate cancer that was diagnosed at ≤60 and men with prostate cancer that was diagnosed at >60 years; cumulative frequency plots overlapped completely (data not shown).

**HD CAG Repeat Length Analysis.** Because men with lymph node-positive prostate cancer seemed to have a higher frequency of short AR CAG repeats but the same frequency of short GGC repeats compared to men with lymph node-negative disease, we sought to determine whether there was also a difference between these subgroups in terms of the frequency of short CAG repeat lengths in other polymorphic genes. We chose to study the HD gene because the length of its CAG repeat (9–30 CAGs; Ref. 40), which codes for a polymorphic glutamine repeat in the HD protein, is in the same range as that of the AR gene (7–35 CAGs). However, the HD protein plays no known or suspected role in prostate cancer; therefore, we considered the HD gene as a negative control. We found that the frequency and cumulative frequency of HD CAG alleles were indistinguishable between men with lymph node-positive and men with lymph node-negative prostate cancer (data not shown). The odds ratio for a short HD CAG repeat (≤17 CAGs) was not significantly different for these two groups of men (odds ratio = 1.14; P = 0.84). Thus, there was no significant association between a short HD CAG repeat (≤17 CAGs) and the development of prostate cancer. The data are consistent with the lack of a role of the HD gene in prostate carcinogenesis.

**Analysis of Each AR at Two Loci: Linkage Equilibrium of AR CAG and GGC Repeats.** Up to now, we have been treating CAG and GGC repeat lengths as individual loci. Because both loci are located in the same exon of the AR gene, we characterized the genotype of each of the 53 men with prostate cancer for whom we had measured the lengths of both the CAG repeat and the GGC repeat. These men had 32 different AR alleles. The most common AR allele was CAG21/GGC16; 6 of 53 men had this allele.

Analysis of the AR genotype in each individual at both repeats allowed us to determine whether the men with a short CAG repeat were the same men who had a short GGC repeat. These data are summarized in Table 3. Of six men with a short (≤17) CAG repeat, only one had a short (≤14) GGC repeat. Of eight men with a short GGC repeat, only one had a short CAG.

### Table 2 Short AR GGC repeat as a potential risk factor in prostate cancer

<table>
<thead>
<tr>
<th></th>
<th>Short ≤14 GGC</th>
<th>Long &gt;14 GGC</th>
<th>Odds Ratiob (CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total PCa</td>
<td>8</td>
<td>46</td>
<td>4.6 (1.3–16.1)</td>
<td>0.02</td>
</tr>
<tr>
<td>Gen Pop</td>
<td>4</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN+ PCa</td>
<td>3</td>
<td>21</td>
<td>0.7 (0.2–3.3)</td>
<td>0.72</td>
</tr>
<tr>
<td>LN− PCa</td>
<td>5</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The numbers of individuals with either a short or long GGC repeat in the germ-line AR gene are shown for all Caucasian men with clinically localized prostate cancer (Total PCa). Caucasian men with lymph node-negative prostate cancer (LN− PCa), Caucasian men with lymph node-positive prostate cancer (LN+ PCa), and the general Caucasian population (Gen Pop) reported in Refs. 20, 24, and 25.

b The odds ratio is equal to the ratio of incidence of short alleles to incidence of long alleles in one group of individuals divided by the same ratio of alleles in a second group of individuals. The 95% confidence interval (CI) for the odds ratio is given in parentheses.

P values were determined using Fisher’s exact test.
Of 13 men with either a short CAG or a short GGC, only 1 had repeats that were both short (17 CAGs/12 GGCs). Therefore, men with a short CAG repeat are predominantly different individuals from those who have a short GGC repeat. This suggests that either a short CAG repeat or a short GGC repeat may increase the odds of developing clinical prostate cancer (Tables 1 and 2). The proportion of men in our sample who had either a short CAG or a short GGC was 13 of 53 (25%; see Table 3).

χ² analysis indicated that GGC repeat length was not significantly associated with CAG repeat length (P = 1.0). The odds of having a short GGC repeat (≤14) were the same in men who had a short CAG (≤17) as in men who had a long CAG (>17) repeat (odds ratio = 1.14; 95% confidence interval, 0.12–1.25). A test for linkage equilibrium indicated that our observed frequencies of individuals with short CAG/short GGC, short CAG/long GGC, long CAG/short GGC, and long CAG/long GGC were all close to the expected values (observed: expected ratio: 1:0.9, 5:5.09, 7:7.1, and 40:39.9, respectively). This analysis indicates that the CAG and GGC microsatellites of the AR gene seem to be in linkage equilibrium in men with prostate cancer, i.e., the frequency of one microsatellite is independent of the other.

DISCUSSION

Microsatellite DNA sequences are common sites of genetic instability, especially in certain types of cancers with defective mismatch repair enzymes (42). By comparison, genome-wide microsatellite instability is uncommon in sporadic prostate cancer (43, 44), and the AR CAG and GGC microsatellite repeats are remarkably stable in sporadic cases of prostate cancer (Ref. 12; this study). There is, however, a baseline rate of AR CAG repeat length instability, which has been demonstrated by genotyping sperm (23). Zhang et al. (23) measured the AR CAG repeat length in each of 200–400 sperm from each of 7 healthy men and inferred a mutation rate of 0.9–3.9%. The mutation rate was higher in men with an AR CAG repeat length at the high end of the normal range (28–31 CAGs; average mutation rate, 3.2%) than it was in men with the most common AR CAG repeat lengths (20–22 CAGs; average mutation rate, 1.3%; Ref. 23). It would not be possible to detect these mutation rates in pooled sperm because of limitations of detection sensitivity; hence, it would not be possible to detect these mutation rates in tissue either. The baseline mutation rate of the AR GGC repeat has not been studied.

Men with clinical stage B prostate cancers are a heterogeneous group at the time of diagnosis; tumors differ in degree of aggressiveness, tumor volume, growth rate, time interval from tumor initiation to tumor detection, and time interval from tumor initiation to metastasis. For example, men with no clinical evidence of metastatic disease are candidates to undergo potentially curative radical prostatectomy, but at the time of surgery, some of these tumors are discovered to have already metastasized (lymph node-positive disease). And a proportion of men who seem to have lymph node-negative disease at the time of surgery actually have undetected micrometastatic disease that subsequently becomes manifest, despite radical prostatectomy. It is likely that even lymph node-positive tumors are heterogeneous and differ in aggressiveness, with some having become metastatic at a time when the tumor was still small, and others not having become metastatic until the tumor was large (4, 5). If AR activity was modulated by glutamine and glycine repeat lengths in the AR, we might expect different AR alleles to influence the role of androgen in the natural history of prostate cancer and perhaps to help account for the variety of tumor phenotypes among men with clinically localized prostate cancer.

The overall frequency of lymph node-positive disease in patients undergoing radical prostatectomy for clinically localized prostate cancer, with no clinical evidence of metastatic spread, is about 10% at our institution (45, 46). A criterion for inclusion of specimens in our tumor bank, which was the source of specimens for the present study, was that they have palpable tumors; this facilitates the subsequent enrichment of tumor DNA in a specimen and facilitates the detection of mutations or loss of heterozygosity. Coincidentally, all men included in the present study had tumors of moderate to high grade. These features probably account for the relatively high proportion (25 of 59, 42%) of aggressive early-stage tumors (i.e., clinically unsuspected metastatic disease) in our sample of men undergoing radical prostatectomy.

On the basis of our study of AR alleles in men with prostate cancer, we suggest that AR alleles with a short CAG or GGC repeat deserve special attention as potential prostate cancer susceptibility alleles. Finding a somatic mutation that shortened the CAG repeat length from 24 to 18 in a specimen of prostate cancer was our first clue that CAG repeat length might actually affect AR activity (12), contrary to the common presumption that polymorphism existed because it had no biological consequence. Subsequent studies confirm that CAG repeat length can indeed modulate AR-mediated effects of androgen (30–35) such that the shorter the repeat, the more active the AR. Our present study of germ-line AR alleles in men with prostate cancer again draws attention to AR alleles with a short repeat. The most striking feature of our sample of men with clinically unsuspected metastatic prostate cancer was that they had a much higher frequency of short CAG AR alleles than did men with lymph node-negative, organ-confined disease (Fig. 2B). Our observations have led us to hypothesize that a short CAG or GGC repeat may increase the risk of developing prostate cancer, and this hypothesis now needs to be tested in case-control and cohort studies. Our original study design had focused on prostate cancer patients, so an age-matched control sample of men without prostate cancer had not been identified at the start of the study. However, the remarkable reproducibility of CAG repeat length measurements and the striking concordance of allele frequency distributions among geographically different samples of the general Caucasian population (20, 24, 25) provided a rationale to compare allele frequencies in our patients with those in historical controls. The fact that most men with prostate cancer (Fig. 2A) and all men with lymph node-negative disease

### Table 3 Short CAG alleles and short GGC alleles identify different men

<table>
<thead>
<tr>
<th>Short CAG (≤17)</th>
<th>Long CAG (&gt;17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short GGC (≤14)</td>
<td>1</td>
</tr>
<tr>
<td>Long GGC (&gt;14)</td>
<td>5</td>
</tr>
</tbody>
</table>


(Fig. 2C) had an allele distribution like that of previously reported general populations supports the use of this general population as a control comparison population, pending confirmation from future case-control studies. Thus, our study may be considered hypothesis-generating rather than conclusive.

CAG allele frequencies in men with clinically localized prostate cancer were mostly concordant with those in the general population, with the notable exception that the frequency of AR alleles with a short (≤17) CAG repeat was substantially higher in men with prostate cancer than it was in the general population, and these short alleles predominantly identified a subpopulation of men who were thought to have organ-confined disease but who actually had clinically unsuspected metastatic (lymph node-positive) prostate cancer. This suggests to us that there may be something unique about the tumors of men who have a germ-line AR gene with a short CAG repeat. If AR proteins with shorter glutamine repeats are more active at transducing androgenic signals, then men with short AR CAG repeats in their germ-line DNA might be more susceptible to develop a form of prostate cancer that is intrinsically more aggressive from an earlier time in tumor development and that becomes metastatic when tumor volume is still low. Our data suggest that a short AR CAG allele may be a risk factor for the development of clinically unsuspected lymph node-positive prostate cancer among men undergoing radical prostatectomy. We have not attempted to address the broader question of whether a short CAG repeat (or a short GGC repeat) increases prostate cancer risk among men in the general population.

On the basis of our analysis of GGC repeat length, a short GGC (glycine) repeat length in the germ-line AR gene (≤14 GGCs) identified a subpopulation of men with clinically localized prostate cancer. We suggest that there may be something unique about the tumors of men with a short AR GGC repeat, especially if the glycine repeat length in the AR affects the androgenicity of endogenous androgen levels. The odds of having a short GGC repeat were not higher in men with clinically unsuspected lymph node-positive prostate cancer than those in men with lymph node-negative disease. In addition, the odds of having a short GGC repeat were not higher in men diagnosed at a younger age (≤60 years) than they were in men diagnosed at an older age. Therefore, a short GGC repeat may not increase the risk of developing a more aggressive disease, when aggressiveness is defined as the presence of tumor cells in pelvic lymph nodes in the absence of clinical evidence of metastasis or diagnosis at a young age. A short GGC repeat might nevertheless be associated with increased aggressiveness, if aggressiveness were measured at a different time in the natural history of the tumor, or if there were other ways to assess aggressiveness. On the other hand, we cannot rule out the alternative possibility that a short GGC repeat may identify a subpopulation of men who develop prostate cancer that has decreased aggressiveness. A proportion of patients with pathologically lymph node-negative disease at the time of radical prostatectomy actually have metastatic disease that may become manifest after prostatectomy. Thus, lymph node-negative status at the time of surgery may underestimate the true metastatic status. Conversely, lymph node-positive status at the time of surgery may represent a heterogeneous population of tumors with differing aggressiveness. Of the tumors we have studied, the follow-up interval after prostatectomy is too short at this time to be able to say whether the patients with a short AR GGC repeat represent a group with more aggressive disease or less aggressive disease. Additional studies are needed to identify the unique features of prostate cancers that develop in Caucasian men with a short GGC repeat in their germ-line AR gene.

Men with a short CAG repeat were different individuals from the men who had a short GGC repeat. Thirteen of 53 (25%) men with clinically localized prostate cancer had an AR allele in their germ-line DNA with ≤17 CAGs or ≤14 GGCs. By comparison, only 6.6% of the general population have either ≤17 CAGs or ≤14 GGCs; this percentage was derived by adding the frequency of ≤17 CAG (3%) and the frequency of ≤14 GGC (3.6%) in the general population and therefore is probably an underestimate, because there are no data on CAG and GGC repeat lengths in the same individuals in the general population. Thus, short CAG and short GGC repeat lengths may be independent risk factors and may both be informative markers for prostate cancer risk assessment.

While our study was in progress, Irvine et al. (25) reported AR CAG and GGC repeat lengths in the germ-line DNA of 39 randomly selected control Caucasian men (>35 years old; mean and median age not stated) and 57 Caucasian men with prostate cancer who had been recruited to a study to identify families with and without a family history of prostate cancer. They concluded that the odds of having an AR allele with <22 CAGs, a GGC repeat length other than 16, or both <22 CAGs and a GGC repeat length other than 16 were not significantly higher in men with prostate cancer than in controls (25). Their definitions of a high-risk allele, which differed from ours, were chosen arbitrarily, and their group of men with prostate cancer may not be comparable to ours. Some of their patients had advanced disease, defined as having tumor invading and extending beyond the prostate capsule, and others had localized disease; however, it was not stated whether these descriptions were based on clinical or pathological criteria (25). Therefore, the conclusions of our study and the study of Irvine et al. (25) cannot be directly compared.

It is of interest to consider how differences in glutamine and/or glycine repeat length in the AR might affect androgen action. The endogenous AR in genital skin fibroblasts from patients with SBMA, which has a glutamine repeat that is twice the normal length, has a lower-than-normal affinity for the androgen R1881 (methyltrienolone; Ref. 47). The transcriptional activity of an AR protein with a mutation in the hormone-binding domain is markedly affected by the length of the glutamine repeat (12 versus 20 Gln) in the NH2-terminal domain (48). These effects of glutamine repeat length may be a consequence of the ability of the NH2-terminal domain of the AR to interact with the COOH-terminal hormone-binding domain (17) and of the effect of differences in the glutamine repeat length (within the normal range) on the phosphorylation state of the AR (49). Hence, differences in glutamine and/or glycine repeat length could affect structural changes in the NH2-terminal domain and thereby affect interactions with the hormone-binding domain of the AR or with other proteins. Transcription requires the assembly of specific multiprotein complexes (50–52). Therefore, glutamine and/or glycine repeat length might affect the strength of interactions with AR-interacting proteins (coac-
tivators and corepressors) and thereby affect the efficiency of transcription of target genes. Glutamine repeats form β-pleated sheets held together by hydrogen bonds between their amide groups, thereby potentially functioning as polar zippers by bringing transcription factors together (53, 54). This suggests a potential role for glutamine repeats in AR dimerization, interaction between the NH2-terminal and COOH-terminal domains, or in the assembly and stabilization of the AR with the multiple factors that constitute an active transcription complex. Because glutamine repeats in other transcription factors can modulate protein-protein interactions (55) and the length of the repeat can affect transcriptional activity (30, 31, 35), we propose that polymorphisms in the length of the AR glutamine and/or glycine repeat could affect the strength of these AR-protein interactions and thereby affect androgen signaling and the role of androgen in prostate carcinogenesis. In addition, polymorphisms might affect the androgenicity of endogenous androgen levels by affecting AR mRNA or protein stability (32).

ACKNOWLEDGMENTS

We thank our colleagues Drs. G. Steven Bova, Jonathan I. Epstein, and William B. Isaacs for providing DNA specimens; Dr. Terri L. H. Beaty for help with the analysis of linkage equilibrium; Jennifer P. Macke and Dr. Jeremy Nathans for advice about PCR conditions to amplify the AR GGC repeat and for clarifying that glycine and glutamine repeat lengths reported by them (22) actually refer to GGC and CAG repeat lengths, respectively; Laura Kasch for advice about primers and PCR conditions to amplify the HD CAG repeat; and Stacey A. May for technical assistance.

REFERENCES

repressive function of the long polyglutamine tract in the human andro-
gen receptor: possible pathogenetic relevance for the (CAG)6-expanded
and location of CAG trinucleotide repeats in the androgen receptor
N-terminal domain affect transactivation function. Nucleic Acids Res.,
Reduced androgen receptor gene expression with first exon CAG repeat
33. Brooks, B. P., and Fischbeck, K. H. Spinal and bulbar muscular
atrophy: a trinucleotide-repeat expansion neurodegenerative disease.
34. Sobue, G., Doyu, M., Morishima, T., Mukai, E., Yasuda, T., Kachi,
T., and Mitsuma, T. aberrant androgen action and increased size of
35. Gerber, H-P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M.,
Rusconi, S., and Schaffner, W. Transcriptional activation modulated by
homopolymeric glutamine and proline stretches. Science (Washington DC),
36. Morgan, T. O., McCarthy, W. F., McLeod, D. G., Jacobsen, S. J.,
and Moul, J. W. Age-specific reference ranges for prostat-specific
37. Morton, R. A., Jr. Racial differences in adenocarcinoma of the
38. Powell, I. J., Schwartz, K., and Hussain, M. Removal of the finan-
tial barrier to health care: does it impact on prostate cancer at presen-
tation and survival? A comparative study between black and white men
39. Sambrook, J., Riess, O., Epplen, J. T., Fenger, K., Hasholt, L.,
and Sorensen, S. A. Trinucleotide repeat elongation in the Huntingtin
gene in Huntington disease patients from 71 Danish families. Hum. Mol.
40. Alexander, R. B., Maguire, M. G., Epstein, J. I., and Walsh, P. C.
Pathological stage is higher in older men with clinical stage B1 aden-
41. Loeb, L. A. Microsatellite instability: marker of a mutator pheno-
42. Bussemakers, M. J. G., Bova, G. S., Schoenberg, M. P., Hakimi,
J. M., Barrack, E. R., and Isaacs, W. B. Microsatellite instability in
43. Cunningham, J. M., Shan, A., Wick, M. J., McDonnell, S. K.,
Schaid, D. J., Tester, D. J., Qian, J., Takahashi, S., Jenkins, R. B.,
Bostwick, D. G., and Thibodeau, S. N. Allelic imbalance and microsat-
44. Sgrignoli, A. R., Walsh, P. C., Steinberg, G. D., Steiner, M. S., and
Epstein, J. I. Prognostic factors in men with stage DI prostate cancer:
identification of patients less likely to have prolonged survival after
45. Partin, A. W., Yoo, J., Carter, H. B., Pearson, J. D., Chan, D. W.,
Epstein, J. I., and Walsh, P. C. The use of prostate-specific antigen,
clinical stage, and Gleason score to predict pathological stage in men
46. MacLean, H. E., Choi, W-T., Rekaris, G., Warne, G. L., and Zajac,
J. D. Abnormal androgen receptor binding affinity in subjects with
Kennedy's disease (spinal and bulbar muscular atrophy). J. Clin. Endo-
47. McPhaul, M. J., Marcelli, M., Tilley, W. D., Griffin, J. E., Isidro-
Gutierrez, R. F., and Wilson, J. D. Molecular basis of androgen resis-
tance in a family with a qualitative abnormality of the androgen receptor
and responsive to high-dose androgen therapy. J. Clin. Invest., 87:
48. Jenster, G., de Ruiter, P. E., van der Korput, H. A. G. M., Kuiper,
G. J. M., Trapman, J., and Brinkmann, A. O. Changes in the
abundance of androgen receptor isotypes: effects of ligand treatment,
glutamine-stretch variation, and mutation of putative phosphorylation
49. Beato, M., Chavez, S., and Truss, M. Transcriptional regulation by
50. Katzenellenbogen, J. A., O'Malley, B. W., and Katzenellenbogen,
B. S. Tripartite steroid hormone receptor pharmacology: interaction
with multiple effector sites as a basis for the cell- and promoter-specific
51. Goodrich, J. A., Cutler, G., and Tjian, R. Contacts in context:
promoter specificity and macromolecular interactions in transcription.
Cell, 84: 825–830, 1996.
52. Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. Glutamine
repeats as polar zippers: their possible role in inherited neurodegenera-
53. Stott, K., Blackburn, J. M., Butler, P. J. G., and Perutz, M. Incor-
dation of glutamine repeats makes protein oligomerize: implications
6513, 1995.
54. Emili, A., Greenblatt, J., and Ingles, C. J. Species-specific interac-
tion of the glutamine-rich activation domains of Spl with the TATA
distribution of the androgen receptor (AR) CAG microsatellite in
human prostate cancer. Abstracts of the 8th Annual Meeting of the
Society for Basic Urologic Research, San Francisco, California, May
Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer.

J M Hakimi, M P Schoenberg, R H Rondinelli, et al.


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/3/9/1599

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.