

The CAG and GGC Microsatellites of the Androgen Receptor Gene Are in Linkage Disequilibrium in Men with Prostate Cancer¹

Ryan A. Irvine, Mimi C. Yu, Ronald K. Ross, and Gerhard A. Coetzee²

Departments of Urology [R. A. I., G. A. C.] and Preventive Medicine [M. C. Y., R. K. R.], University of Southern California/Norris Comprehensive Cancer Center, University of Southern California School of Medicine, Los Angeles, California 90033-0804

ABSTRACT

The androgen receptor genotype was determined in the white blood cell DNA of 45 African-American, 39 non-Hispanic white, and 39 Asian (Chinese, Japanese) normal subjects and 68 patients with prostate cancer (57 whites), all of whom were residents of Los Angeles County. For each subject, we measured the number of repeats in the polymorphic CAG and GGC microsatellites of exon 1 of the androgen receptor gene. In normal subjects, the distributions of CAG and GGC microsatellites differed significantly among the races (two-sided $P = 0.046$ and < 0.0005 , respectively). The prevalence of short CAG alleles (<22 repeats) was highest (75%) in African-American males with the highest risk for prostate cancer, intermediate (62%) in intermediate-risk non-Hispanic whites, and lowest (49%) in Asians at very low risk for prostate cancer. High-risk African-Americans also had the lowest frequency (20%) of the GGC allele with 16 repeats; the comparable values for intermediate-risk whites and low-risk Asians were 57% and 70%, respectively. Consistent with the interracial variation in CAG and GGC distributions, there was an excess of white patients with <22 CAG and not-16 GGC repeats relative to white controls (relative risk, 2.1; one-sided $P = 0.08$). We observed no association (linkage) between the two microsatellites among normal subjects. On the other hand, there was a statistically significant negative association between the numbers of CAG and GGC repeats among the prostate cancer patients studied (two-sided $P = 0.008$). Among the 47 subjects with short CAG alleles (<22 repeats), 43% had long GGC alleles (>16 repeats) whereas only 15% of the 20 subjects with long CAG alleles (≥ 22 repeats) had long GGC alleles. Overall, our data suggest a possible association between CAG and GGC microsatellites of the androgen receptor gene and prostate cancer development.

INTRODUCTION

In the United States, prostate cancer accounts for about 38,000 deaths per year and approximately 200,000 new cases are expected to be diagnosed in 1995. It is now the most frequently diagnosed non-skin cancer in U.S. men. Prostate cancer incidence differs considerably among the various racial-ethnic groups in Los Angeles County. It is highest in African-Americans (116/100,000 person-years), intermediate in non-Hispanic whites (71/100,000 person-years), and lowest among Asians [Japanese (39/100,000 person-years) and Chinese (28/100,000 person-years)] (1).

Since prostatic epithelium is androgen dependent, it is possible that the AR³ gene may influence an individual's susceptibility to prostate cancer. Although the presence of somatic mutations in the AR has been reported in advanced tumors (2), the growth of most early forms of prostate cancer is likely to require the normal activity of the AR. This receptor is necessary to translate the well-known effects of androgens on prostate cells, both normal and malignant (3). One of the critical functions of the AR is to activate the expression of target

genes. This transactivation activity resides in the N-terminal domain of the protein (4, 5), which is encoded in exon 1 and contains polymorphic CAG (6) and GGC (7) repeats (microsatellites). We have postulated previously that a smaller size of the CAG repeat is associated with a higher level of receptor transactivation function, thereby possibly resulting in a higher risk of prostate cancer (8). Recently, a somatic mutation resulting in a contraction in this microsatellite (CAG₂₄→₁₈) was observed in an adenocarcinoma of the prostate, and the effects of the shorter allele were implicated in the development of that tumor (9). In the present study, we tested the involvement of the AR in prostate cancer by analyzing the possible linkage disequilibrium/equilibrium between the two microsatellites at this locus in men diagnosed with prostate cancer and in healthy normal men.

MATERIALS AND METHODS

Prostate Cancer Patients

In late 1992 a pilot study of familial prostate cancer was initiated among prostate cancer patients identified in Los Angeles County in 1991-1992. Subjects were identified by the Cancer Surveillance Program, the population-based Surveillance, Epidemiology, and End Results Cancer Registry of Los Angeles County. This registry has been in existence since 1972, and is estimated to be 99% complete in its goal of identifying all incident cancer patients among the 9 million current residents of Los Angeles County (1). Five hundred fifty-nine (53%) of the 1062 eligible patients responded to a mailed questionnaire which requested detailed information on family history of cancer. From these 559 patients, we selected those with one or more brothers currently residing in Los Angeles County; the goal was to enlist approximately equal numbers of these patients either negative or positive for a family history of prostate cancer. A 10-ml tube of venous blood in EDTA was collected on all consenting patients and their brothers. The data presented in this report comprise the first 68 incident prostate cancer patients diagnosed in 1991-1992, who were recruited for this pilot study. Fifty-seven patients were non-Hispanic whites (whites), nine were African-Americans, one was an Egyptian national, and one was Chinese. Twelve patients had one or more first-degree relatives (father, brother, son) with prostate cancer. The mean age of the patients at diagnosis was 57.7 (range, 51-68) years. Thirty-five patients had disease localized to the prostate gland, 32 had advanced disease (defined as having tumor invading and extending beyond the prostatic capsule), and in 1 patient the disease stage was unknown. Buffy coats of lymphocytes prepared from the blood specimens were stored continuously at -70°C until analysis.

Control Subjects

In 1991, a cross-sectional survey involving apparently healthy adult males from several racial-ethnic groups in Los Angeles was initiated (10). Subjects were identified from a comprehensive list of registered drivers residing in the Los Angeles County, who were over the age of 35 years. One hundred thirty-three such men were recruited into that study, and buffy coats of lymphocytes were collected from 123 of them [45 African Americans, 39 non-Hispanic whites, and 39 Asians (Chinese and Japanese)]. This group comprised our control subjects. For purposes of the original study (10), 53% ($n = 65$) of our control subjects were lifelong nonsmokers, and the remaining 58 subjects were current smokers of varying intensity.

Microsatellite Size Determination

DNA was isolated from a suspension of buffy coats using a rapid extraction protocol (11).

Received 12/9/94; accepted 3/2/95.

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.

¹ This work was supported in part by Grants P01 CA 17054 and R35 CA 53890 from the National Cancer Institute.

² To whom requests for reprints should be addressed, at University of Southern California/Norris Comprehensive Cancer Center, Room 630A, 1441 Eastlake Avenue, P.O. Box 33804, Los Angeles, CA 90033.

³ The abbreviation used is: AR, androgen receptor.

CAG Repeat. About 10 ng undigested DNA were amplified by two rounds of PCR in a series using nested primers surrounding the CAG repeat in exon 1 of the AR (12): outside, 5'-GTGCGGAAGTGATCCAGAA-3' and 5'-TCTGGGACGCAACCTCTCTC-3'; inside, 5'-AGAGGCCGCGAGCGCAG-CACCTC-3' and 5'-GCTGTGAAGGTTGCTGTTCTCAT-3'. The first round consisted of 17 cycles (94°C for 1 min, 55°C for 1 min, and 72°C for 1.5 min). The second round, performed using 1 µl of the first and the addition of 2 µCi [α -³²P]dCTP, consisted of 28 cycles (94°C for 1 min, 66°C for 1 min, and 72°C for 1.5 min). The final products were analyzed by electrophoresis on 5% denaturing polyacrylamide gels. These were subjected to autoradiography and the number of CAG repeats were calculated from the size of the predominant PCR product (middle main band) in relation to a series of previously determined (by direct sequencing of PCR products) CAG size standards. All samples were then ranked according to size and rerun a second time with equal-sized alleles next to each other, to check and modify, if necessary, the accuracy of the original determination.

GGC Repeat. This repeat, due to its high G + C content, was amplified using the more thermostable *Pfu* DNA polymerase (Stratagene). About 10 ng undigested DNA were amplified by two rounds of PCR in a series using nested primers surrounding the GGC repeat in exon 1 of the AR: outside (7), 5'-TCCTGGCACACTCTCTTTCAC-3' and 5'-GCCAGGGTACCACATCAGGT-3'; inside,⁴ 5'-ACTCTTTCACAG-CCGAAGAAGGC-3' and 5'-ATCAGGTGCGGTGAAGTCGCTTTC-3'. The first round consisted of 17 cycles (98°C for 1 min and 70°C for 5 min). The second round, performed using 1 µl of the first and the addition of 2 µCi [α -³²P]dCTP, consisted of 34 cycles (98°C for 1 min and 70°C for 5 min). The final products were analyzed by electrophoresis on 5% denaturing polyacrylamide gels. These were subjected to autoradiography and the number of GGC repeats were calculated from the size of the predominant PCR product. The relative size determinations were determined in exactly the same manner as described for the CAG repeat above.

Statistical Analysis

The χ^2 test of association was used to examine the relationship of CAG and GGC repeats by race, as well as the relationship between CAG and GGC repeats among prostate cancer subjects and controls separately. The binomial test was used to examine the difference in prevalence of specific CAG/GGC alleles between white prostate cancer subjects and white control subjects. All *P* values quoted are two-sided unless otherwise stated.

RESULTS

Table 1 presents the distribution of CAG repeats by race among control subjects. Using the median repeat size in Asians (*i.e.*, 22 repeats) as an arbitrary cutoff, 33 (75%) of 44 African-American versus 24 (62%) of 39 whites and 19 (49%) of 39 Asians possessed short alleles (<22 repeats). These differences were statistically significant (*P* = 0.046).

The distribution of GGC repeats was also significantly different among African-American, white, and Asian men (*P* < 0.0005, Table 2). Twenty-six (70%) of 37 Asian controls had 16 GGC repeats compared to 21 (57%) white and 8 (20%) African-American controls. The prevalence of short alleles (<16 repeats) was especially high in African-Americans: 25 (61%) had the short alleles relative to 10 (27%) Asians and 4 (11%) whites. Very few Asians (1/37) had alleles longer than 16. The comparable proportions of these long alleles in African-Americans and whites were 8 of 41 and 12 of 37, respectively.

Table 3 presents the prevalence of "high-risk" alleles (based on results of interracial comparisons described above) of the CAG and GGC microsatellites between white prostate cancer patients and white control subjects. There were modest, nonsignificant excesses of <22 CAG repeats and not-16 GGC repeats among patients relative to controls (relative risk, 1.25 and 1.18, respectively). The excess in

Table 1 Distributions of allele frequencies (%) of the CAG microsatellite at the AR locus among control subjects according to racial-ethnic grouping^a

No. of CAG	African-American (n = 44)	White (n = 39)	Asian (n = 39)
9	2.3	0	0
10	0	0	0
11	0	0	0
12	2.3	0	0
13	6.8	0	0
14	2.3	0	0
15	6.8	0	2.6
16	9.1	0	0
17	13.6	5.1	0
18	4.5	12.8	7.7
19	6.8	12.8	10.3
20	11.4	23.1	12.7
21	9.1	7.7	15.4
22	13.6	5.1	15.4
23	6.8	12.8	5.1
24	0	12.8	10.3
25	2.3	2.6	5.1
26	0	2.6	7.7
27	2.3	2.6	5.1
28	0	0	0
29	0	0	2.6
Total	100	100	100

^a One specimen was noninformative, therefore *n* = 122; two-sided *P* value (difference in distribution by the three racial groups. <22 versus ≥22 repeats) = 0.046.

patients became more substantial when high-risk alleles of both microsatellites were considered simultaneously. There was a 2.1-fold excess risk for prostate cancer associated with <22 CAG repeats coupled with not-16 GGC repeats relative to other allelotypes, but the association did not reach statistical significance (one-sided *P* = 0.08). Among the control subjects, African-Americans had the highest prevalence of the putative highest risk allelotype (25/39 = 64%) while whites and Asians had intermediate (9/37 = 24%) and low (3/34 = 9%) prevalence, respectively, and the differences were statistically significant (*P* < 0.0005).

When we tested a possible association (linkage disequilibrium) between CAG and GGC repeats among control subjects, the result was negative (*P* = 0.53). Furthermore, when we studied the relationship between the two repeats within the African-American, white, and Asian controls separately, we observed no association between the two microsatellites in any one of the racial subgroups. In contrast, the distribution of CAG repeats was highly correlated with that of GGC repeats among our prostate cancer patients (*P* = 0.008, Table 4). There was a close association between short CAG alleles and long GGC alleles. Of the 47 patients possessing short CAG alleles (<22 repeats), 20 (43%) had GGC alleles greater than 16 repeats. The comparable proportion among the 20 patients with long CAG alleles (≥22 repeats) was 15% (*n* = 3). Comparable results were obtained when we limited the analysis to white prostate cancer patients, although due to the smaller sample size (*n* = 57), the *P* value for the χ^2 test of association between CAG and GGC repeats was slightly higher at 0.015.

DISCUSSION

More than double the average number of CAG repeats were found in the AR genes of patients with Kennedy's disease, and this caused a blunted transactivation activity of the receptor in *in vitro* assays (13). This finding was confirmed and further extended by Chamberlain *et al.* (14) who demonstrated that the elimination of the CAG repeats in both human and rat AR resulted in a marked elevation of transcriptional activation activity, suggesting that the presence of this repeat is inhibitory to transactivation. The investigators also established a negative linear relationship between CAG lengths of 25–77 repeats and

⁴ M. Patterson, personal communication.

transactivation activity. It is possible, therefore, that shorter *CAG* alleles cause a more active growth of prostate cells (due to greater transactivation activity of the receptor), and thereby resulting in a higher level of prostate cancer risk since target cell proliferation is known to promote tumorigenesis.

In the present study, we observed a significant difference in the prevalence of short *CAG* alleles among the three races studied. Short alleles were most common among African-American males who are at relatively high risk of prostate cancer, and least common among Asians who are at relatively low risk of prostate cancer. In white males, whose prostate cancer risk is intermediate between those of the other two groups, the prevalence of short *CAG* alleles was intermediate. Furthermore, we found a modest, nonsignificant excess of short *CAG* alleles among white prostate cancer patients relative to white control subjects. Due to the nature and size of our case/control set, no definitive conclusion with respect to this result can be drawn at present. The data are compatible with the null hypothesis, but they also can be interpreted as suggesting a weak to moderate association between short *CAG* alleles and prostate cancer risk.

GGC allele distributions also differed significantly among normal men of the three races studied. The lowest risk Asians had the highest prevalence of 16 *GGC* repeats, whereas the highest risk African-Americans had the least prevalence. There was a modest, nonsignificant excess of not-16 *GGC* repeats in white patients compared to controls. At present, nothing is known about the functional roles of this repeat. We are in the process of initiating laboratory-based research to assess the possible role of this microsatellite in *AR* transactivation activity and to determine the biological basis of the putative protective role of 16 *GGC* repeats in the context of prostatic epithelial cell proliferation.

Since the *AR* gene is located on the X chromosome (each male has only one copy), the two microsatellites at this locus can simply be associated with each other. Because of the close proximity of the two microsatellites, one would expect to find linkage disequilibrium between them. However, we observed no association between the *CAG*

Table 4 Frequency distribution of the *CAG* and *GGC* microsatellites at the *AR* locus in prostate cancer patients^a

No. of <i>CAG</i>	No. of <i>GGC</i>			Total
	<16	16	>16	
<22	11 (23%)	16 (34%)	20 (43%)	47 (100%)
≥22	2 (10%)	15 (75%)	3 (15%)	20 (100%)

^a One specimen was noninformative for the *GGC* assay, therefore $n = 67$; two-sided P (test of association between no. of *CAG* and no. of *GGC* repeats) = 0.008.

and *GGC* alleles among our control subjects, consistent with the results of a recent study involving 197 homosexual males and 213 unselected subjects (15). These unexpected findings likely indicate that either one or both of the repeats mutate at a relatively high rate and independently of each other. Indeed, when the rate of mutation at the *CAG* site was measured using single-cell assays of sperm, an exceptionally high rate of 1–4% was obtained (16).

In contrast to the situation in normal men, in our prostate cancer patients there was a clear indication of linkage disequilibrium between the *CAG* and *GGC* alleles at this locus. Our data, therefore, suggest that a nonrandom subpopulation of *AR* alleles occurs in men with prostate cancer. In addition, our white prostate cancer patients showed a higher prevalence of a putative high-risk *CAG/GGC* allelotype relative to white controls (relative risk, 2.1), although the difference was not statistically significant.

We propose that the *AR* gene might be one determinant of prostate cancer risk possibly due to a direct influence of the number of the *CAG* and/or the *GGC* repeats on *AR* function. Alternatively, the microsatellites may be in linkage disequilibrium with other disease-causing mutations in the *AR* gene or with other unknown adjacent genes that affect prostate cancer risk. The present results and conclusions must be regarded as tentative due to the relatively small sample size and the fact that the study was not specifically designed to address the hypothesis under consideration, although we can think of no obvious bias in the data set that could result in a spurious association. Well-designed case/control studies of sufficient sample size are required to test directly the hypothesis that the numbers of *CAG* and *GGC* repeats predict the likelihood of prostate cancer development.

ACKNOWLEDGMENTS

We thank Kazuko Arakawa for computing assistance.

REFERENCES

- Bernstein, L., and Ross, R. K. Cancer in Los Angeles County: A Portrait of Incidence and Mortality 1972–1987. Los Angeles: University of Southern California Press, 1991.
- Gaddipati, J. P., McLeod, D. G., Heidenberg, H. B., Sesterhenn, I. A., Finger, M. J., Moul, J. W., and Srivastava, S. Frequent detection of codon 877 mutation in the androgen receptor gene in advanced prostate cancers. *Cancer Res.*, 54: 2861–2864, 1994.
- Ross, R. K., and Henderson, B. E. Do diet and androgens alter prostate cancer risk via a common etiological pathway? *J. Natl. Cancer Inst.*, 86: 252–254, 1994.
- Simental, J. A., Sar, M., Lane, M. V., French, F. S., and Wilson, E. M. Transcriptional activation and nuclear targeting signals of the human androgen receptor. *J. Biol. Chem.*, 266: 510–518, 1991.
- Jenster, G., Van der Korput, H. A. G. M., Van Vroonhoven, C., Van der Kwast, T. H., Trapman, J., and Brinkmann, A. O. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. *Mol. Endocrinol.*, 5: 1396–1404, 1991.
- Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., and Chakraborty, R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, 12: 241–253, 1992.
- Sleddens, H. F. B. M., Oostra, B. A., Brinkmann, A. O., et al. Trinucleotide (*GGN*) repeat polymorphism in the androgen receptor gene. *Hum. Mol. Genet.*, 2: 273, 1993.
- Coetzee, G. A., and Ross, R. K. Prostate cancer and the androgen receptor. *J. Natl. Cancer Inst.*, 86: 872–873, 1994.
- Schoenberg, M. P., Hakimi, J. M., Wang, S., Bova, G. S., Epstein, J. I., Fischbeck, K. H., Isaacs, W. B., Walsh, P. C., and Barrack, E. R. Microsatellite mutation

Table 2 Distributions of allele frequencies (%) of the *GGC* microsatellite at the *AR* locus among control subjects according to racial-ethnic grouping^a

No. of <i>GGC</i>	African-American ($n = 41$)	White ($n = 37$)	Asian ($n = 37$)
8	2.4	0	0
9	0	0	0
10	4.9	2.7	2.7
11	4.9	0	2.7
12	9.8	0	5.4
13	4.8	0	2.7
14	12.2	0	2.7
15	22.0	8.1	10.8
16	19.5	56.8	70.3
17	19.5	32.4	2.7
Total	100	100	100

^a Eight specimens were noninformative, therefore $n = 115$; two-sided P (difference in distribution by the three racial groups, 16 versus not-16 repeats) < 0.0005.

Table 3 Allelic distributions of the *CAG* and *GGC* microsatellites in white prostate cancer patients and control subjects

Alleles	No. of cases	No. of controls	Relative risk
No. of <i>CAG</i> repeats			
<22	38	24	1.25
≥22	19	15	1.00
No. of <i>GGC</i> repeats			
not-16	27	16	1.18
16	30	21	1.00
No. of <i>CAG</i> /no. of <i>GGC</i>			
<22/not-16	23	9	2.10 ^a
Other	34	28	1.00

^a One-sided $P = 0.08$.

- (CAG₂₄₋₁₈) in the androgen receptor gene in human prostate cancer. *Biochem. Biophys. Res. Commun.*, 198: 74-80, 1994.
10. Yu, M. C., Skipper, P. L., Taghizadek, K., Tannenbaum, S. R., Chen, K. K., Henderson B. E., and Ross, R. K. Acetylator phenotype, aminobiphenyl-hemoglobin adduct levels, and bladder cancer risk in white, black and Asian men in Los Angeles, California. *J. Natl. Cancer Inst.*, 86: 712-716, 1994.
 11. Talmud, P., Tybjaerg-Hansen, A., Bhatnagar, D., Mbewu, A., Miller, J. P., Durrington, P., and Humphries, S. Rapid screening for specific mutations in patients with a clinical diagnosis of familial hypercholesterolaemia. *Atherosclerosis*, 89: 137-141, 1991.
 12. Tsai, Y. C., Simoneau, A. R., Spruck, C. H., Nichols, P. W., Sreven, K., Buckley, J. D., and Jones, P. A. Mosaicism in human epithelium: macroscopic monoclonal patches cover the urothelium. *J. Urol.*, in press, 1995.
 13. Mhatre, A. N., Trifiro, M. A., Kaufman, M., Kazemi-Esfarjani, P., Figlewicz, D., Rouleau, G., and Pinsky, P. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nature Genet.*, 5: 184-188, 1993.
 14. Chamberlain, N. C., Driver, E. D., and Miesfeld, R. L. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect trans-activation function. *Nucleic Acids Res.*, 22: 3181-3186, 1994.
 15. Macke, J. P., Hu, N., Hu S., Bailey, M., King, V. L., Brown, T., Hamer, D., and Nathans, J. Sequence variation in the androgen receptor gene is not a common determinant of male sexual orientation. *Am. J. Genet.*, 53: 844-852, 1993.
 16. Zhang, L., Leeflang, E. P., Yu, J., and Arnheim, N. Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. *Nature Genet.*, 7: 531-535, 1994.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

The CAG and GGC Microsatellites of the Androgen Receptor Gene Are in Linkage Disequilibrium in Men with Prostate Cancer

Ryan A. Irvine, Mimi C. Yu, Ronald K. Ross, et al.

Cancer Res 1995;55:1937-1940.

Updated version Access the most recent version of this article at:
<http://cancerres.aacrjournals.org/content/55/9/1937>

- 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, use this link <http://cancerres.aacrjournals.org/content/55/9/1937>. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.