# ANDROGEN RECEPTOR POLYMORPHISMS: ASSOCIATION WITH PROSTATE CANCER RISK, RELAPSE AND OVERALL SURVIVAL

Stephen M. Edwards[1], Mike D. Badzioch[2], Ralph Minter[1], Rifat Hamoudi[1], Nadine Collins[1], Audrey Ardern-Jones[3], Anna Dowe[3], Simon Osborne[1], Jo Kelly[1], Robert Shearer[3], Doug F. Easton[4], Grady F. Saunders[5], David P. Dearnaley[6] and Rosalind A. Eeles[1,3]*

[1]CRC Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK
[2]Unit of Genetic Epidemiology, International Agency for Research on Cancer, Lyon, France
[3]Royal Marsden NHS Trust, Sutton, Surrey, UK
[4]CRC Section of Genetic Epidemiology, Strangeways Laboratories, Cambridge, UK
[5]University of Texas, M.D. Anderson Cancer Center, Houston, Texas, USA
[6]Academic Radiotherapy Unit, Royal Marsden NHS Trust, and Institute of Cancer Research, Sutton, Surrey, UK

**Several reports have suggested that one or both of the trinucleotide repeat polymorphisms in the human androgen receptor *(hAR)* gene, $(CAG)_n$ coding for polyglutamine and $(GGC)_n$ coding for polyglycine, may be associated with prostate cancer risk; but no study has investigated their association with disease progression. We present here a study of both *hAR* trinucleotide repeat polymorphisms not only as they relate to the initial diagnosis but also as they are associated with disease progression after therapy. Lymphocyte DNA samples from 178 British Caucasian prostate cancer patients and 195 control individuals were genotyped by PCR for the $(CAG)_n$ and $(GGC)_n$ polymorphisms in *hAR*. Univariate Cox proportional hazard analysis indicated that stage, grade and GGC repeat length were individually significant factors associated with disease-free survival (DFS) and overall survival (OS). The relative risk (RR) of relapse for men with more than 16 GGC repeats was 1.74 (95% CI 1.08–2.79) and of dying from any cause, 1.98 (1.13–3.45). Adjusting for stage and grade, GGC effects remained but were not significant ($RR_{DFS} = 1.60$, $p = 0.052$; $RR_{OS} = 1.65$, $p = 0.088$). The greatest effects were in stage T1-T2 ($RR_{DFS} = 3.56$, 95% CI 1.13–11.21) and grade 1 ($RR_{DFS} = 6.47$, 95% CI 0.57–72.8) tumours. No differences between patient and control allele distributions were found by odds-ratio analysis, nor were trends with stage or grade evident in the proportion of short CAG alleles. Non-significant trends with stage and grade were found in the proportion of short GGC alleles. The $(GGC)_n$ polymorphism in this population is a significant predictor of disease outcome. Since the $(GGC)_n$ effect is strongest in early-stage tumours, this marker may help forecast aggressive behaviour and could be used to identify those patients meriting more radical treatment. *Int. J. Cancer (Pred. Oncol.)* 84:458–465, 1999.**
© 1999 Wiley-Liss, Inc.

Prostate cancer been described as a disease that will affect every man if he lives long enough. Many studies have shown that age is the most important risk factor for prostate cancer in virtually all populations studied. It is also believed that underlying the wide variation among populations in the age-adjusted rates of clinical disease there is a universally high rate of development of histological or latent prostate cancer that also increases with age (Carter *et al.*, 1990). This implies that much of the age-associated risk is likely to be biologically innate, common in and among populations and independent not only of external causes but also of the tumour itself. The growth and development of the human prostate gland together with the maintenance of its physiological integrity in the adult is dependent on the presence of circulating androgens and intact intracellular steroid signalling pathways (Cunha *et al.*, 1987). Androgen stimulation is also required during the initial stages of prostate cancer development (Dearnaley, 1994). During this period, androgen withdrawal induces tumour shrinkage through the release of androgenic inhibition of normal apoptotic pathways present in primary and metastatic prostate cancer (Isaacs, 1984; Kyprianou and Isaacs, 1988).

Whereas early prostate cancer is dependent on an androgenic environment, the disease invariably becomes hormone-unresponsive and advanced stages become unresponsive to steroid regulation by one or more presently unknown mechanisms (Dearnaley, 1994). Several aberrations in the human androgen receptor *(hAR)* gene, including enhanced expression (Sadi *et al.*, 1991), gene amplifications (Visakorpi *et al.*, 1995), point mutations (Veldscholte *et al.*, 1990) and small deletions (Schoenberg *et al.*, 1994), have been suggested, but not proven, to underlie this important clinical characteristic.

The *hAR* gene, located at chromosome Xq11–12, encodes the androgen receptor (AR), a ligand-activated transcription factor which mediates the androgenic response. In addition to stimulating expression of genes associated with the differentiated phenotype of the prostate, such as prostate-specific antigen (PSA) (Riegman *et al.,* 1991), it has been reported that AR may regulate genes involved in cell-cycle control, *e.g.,* CDK2, CDK4 and p16 (Lu *et al.,* 1997). The 8 exons of *hAR* define 3 major protein functional domains: a C-terminal ligand-binding domain, a central DNA-recognition and -binding domain and an N-terminal transcription-activation domain. Within the latter domain are 2 single–amino acid repeat length polymorphisms: the first, a polyglutamine tract, is coded by a repeated CAG (Sleddens *et al.,* 1992); the second, a variable number of glycine residues, is coded by a tandem array of GGC sequences (Sleddens *et al.,* 1993).

When case-control studies have found associations between prostate cancer risk and one or both of the *hAR* trinucleotide repeat polymorphisms, the risks have tended to be strongest for men with shorter $(CAG)_n$ alleles and higher clinical stage and tumour grade, *i.e.,* more aggressive tumours (Irvine *et al.,* 1995; Giovannucci *et al.,* 1997; Hardy *et al.,* 1996; Ingles *et al.,* 1997; Stanford *et al.,* 1997). Hakimi *et al.* (1997) observed that the frequency of shorter $(CAG)_n$ alleles was substantially higher in patients and that a significant proportion of these patients, who were candidates for prostatectomy, had unsuspected metastatic disease identified as lymph node-positive prostate cancer. Hardy *et al.* (1996) extended

S.M. Edwards and M.D. Badzioch are considered equal first authors.

M.D. Badzioch's former address is: University of Texas, M.D. Anderson Cancer Center, Houston, TX, USA.

*Correspondence to: CRC Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, SM2 5NG UK. Fax: +44 (0)181 770 1489. E-mail: ros@icr.ac.uk

the analysis to include possible effects on the clinical course of disease (PSA-related relapse) but found no association with $(CAG)_n$ length. Differences in average $(CAG)_n$ repeat length have been found to exist among ethnic groups and have correlated with the prostate cancer risk among those populations (Irvine *et al.,* 1995). *In vitro* investigations have suggested that variation in $(CAG)_n$ length affects the AR capacity to stimulate transcription at target gene promoters (Kazemi-Esfarjani *et al.,* 1995; Chamberlain *et al.,* 1994). These functional results indicate that fewer CAG triplets result in a more active receptor in controlled conditions and are correlated with ethnic variations in prostate cancer risk.

Length variation of the $(GGC)_n$ sequence has been less widely studied for several reasons, resulting in a possibly biased assessment of risk. For instance, the sequence of GGC triplets is inherently more difficult to efficiently amplify using PCR, is less polymorphic and has not been shown to have significant ethnic variation. The results from $(GGC)_n$ analysis are also conflicting. Whereas Irvine *et al*. (1995) found long $(GGC)_n$ alleles to be associated with increased risk, others (Stanford *et al.,* 1997; Hakimi *et al*., 1997) found that men with short alleles were at increased risk.

Overall, these results implicate the involvement of normal AR polymorphisms in the natural history of prostate cancer, but the mechanism or nature of that involvement remains poorly defined. For example, which of them is involved in prostate cancer pathogenesis, or are they both involved? Irvine *et al*. (1995) have provided evidence that the $(CAG)_n$ and $(GGC)_n$ sequences may be in linkage disequilibrium in prostate cancer patients and not in control individuals. Also, to date, all studies of AR polymorphisms and prostate cancer have been undertaken in North America. Studies of migrating populations have shown that prostate cancer risks may be significantly influenced by place of residence (Swerdlow *et al.,* 1995; Shimizu *et al.,* 1991). Dietary fat intake has been suggested to be the basis for some of the environmentally induced differences in prostate cancer rates between populations (Whittemore *et al.,* 1995). Overall, the extent to which the selection of *hAR* polymorphisms in prostate cancer may be influenced by environmental factors remains to be addressed.

We hypothesised that if an association between *hAR* and prostate cancer risk exists, it would also be seen, and may even be stronger, in relation to the post-diagnosis clinical course of the disease. It was our belief that disease aggressiveness and clinical behaviour would be more closely monitored and less susceptible to confounding by extraneous factors such as socio-economic status than would be disease status at diagnosis. Additionally, although studies of the conservative management of localised disease have determined overall progression rates (Chodak *et al.,* 1994), it remains difficult to predict the time course of progression within each component of the localised stages, and further determinants and markers of disease advancement need to be found. The present study of both *hAR* trinucleotide polymorphisms emphasises disease outcome, rather than characteristics at diagnosis, in a European population. We have genotyped 178 British prostate cancer patients to assess risk by comparison with population-based controls, to correlate their genotypes with several clinical characteristics of the disease and to determine if any differences exist in *hAR*-associated risk patterns between previously published reports and these British patients.

## MATERIAL AND METHODS

### Patients

Prior to August 1993, 201 referred cases of prostate cancer were systematically ascertained at the Royal Marsden NHS Trust (RMH), London, UK, to study the effects of inherited mutations and polymorphisms on prostate cancer development. Of these, 183 were Caucasian, 13 were black (native Africans and Caribbeans) and 5 were Asian. Since the numbers of black and Asian men were not sufficient to study ethnicity-specific effects, they are not included in this analysis. On review, 5 of the Caucasian patients were found to not have primary cancer of the prostate, *i.e.,* stage T0, and so were excluded. The average age at presentation of the 178 Caucasian patients was 68.1 years (SD = 6.82 years), with a range of 48 to 80 years, and the median date of admission to RMH was March 1992. There was no association of age with year of presentation. After receiving informed consent through interview, demographic data, detailed family history and a 10-ml peripheral blood sample were obtained. Tumour specimens were also collected for histological review. Tumours were classified as grade 3 if any poorly differentiated cells/foci were present (Gleason score 8 to 10), grade 2 if only moderate differentiation was seen (Gleason score 4 to 7) and grade 1 if all cells/foci were well-differentiated (Gleason score <4). We were unable to obtain pathology specimens or to verify tumour grade in 16 cases, and these were not considered for further analysis except where indicated. Disease staging was performed by clinical examination, abdomino-pelvic computerised tomography and bone scans and serum PSA measurement. Patients were categorised into 1 of 6 groups based on the TNM classification system: T1, non-palpable tumour; T2, palpable but organ-confined; T3 and T4, penetrating margins but no metastasis identified; N+, pelvic nodal involvement; and M+, widespread metastasis. Patients were treated with hormone therapy alone, radiotherapy or a combination of these. Those presenting with higher stage disease were generally given hormone treatment alone. After therapy, patients were routinely assessed for disease progression at 6-month intervals until April 1997, the end of the study. At this time, 80 (49.4%) were alive and well at last routine clinic visit, 54 (33.3%) had died and 28 (17.3%) had not been seen by the clinic for over 6 months.

### Controls

Control genotypes were obtained from the DNA of Caucasian females who were age- and location-matched to breast cancer patients for a previous breast cancer case-control study conducted by the UK National Case-Control Study Group (1989). Up to 195 women were available for analysis. These DNA samples provided the source of data for allele frequency comparisons, and as samples were from females, data are presented on up to 390 X chromosomes. All controls were cancer-free, randomly selected *via* general practice surgeries and aged up to 45 years.

### DNA extraction

DNA was extracted from blood samples (10 ml) as previously described (Edwards *et al.,* 1997) with the inclusion of a second proteinase K digestion at 50°C for 2 to 3 hr. DNA was dissolved in 0.2 to 0.4 ml of water (BDH, Poole, UK) and stored at –20°C until required.

### Radioactive end-labelling

The appropriate amplification primer (see below) was end-labelled in a 30 μl reaction volume containing 1× reaction buffer (Stratagene, Cambridge, UK), 0.33 units T4-PNK enzyme (Stratagene) and 0.17 to 0.34 mCi $^{32}$P-γ-dATP (ICN, High Wycombe, UK; 7,000 Ci/mmol).

### $(CAG)_n$ amplification

PCR was conducted with primers 5′-tccgcgaagtgatccagaac and 5′-cttggggagaaccatcctca. Reactions (25 μl) were conducted with Perkin-Elmer (Beaconsfield, UK) reaction buffer [50 mM KCl, 10 mM Tris-HCl (pH 8.3)], 0.8 mM $MgCl_2$ (Perkin-Elmer), 0.2 mM each dNTP (0.8 mM total, Stratagene), primers at 0.1 OD/ml (0.5 μM final), 0.75 units Amplitaq Gold (Perkin-Elmer) and 100 ng DNA. Thermocycling consisted of 95°C for 9 min, 4 cycles of 1 min at 94°C, 30 sec at 64°C and 1 min at 72°C, followed by a ''touchdown'' reduction of annealing temperature by 2°C every 4 cycles until 54°C (20 cycles on ''touchdown''), then another 29 cycles at 54°C annealing, with a final polymerase extension for 10 min at 72°C.

## *(GGC)$_n$ amplification*

PCR was conducted with primers 5′-tcctggcacactctcttcac and 5'-gccagggtaccacacatcaggt. Reaction conditions were as described above with the following modifications: 1.3 mM MgCl$_2$, 0.25 mM each dNTP [1.00 mM total with 3:1 mixture of 7-deaza-dGTP: dGTP (Pharmacia, Milton Keynes, UK)], 2.5% v/v DMSO (Sigma, Poole, UK; molecular grade), primers at 0.15 OD/ml (0.7μM final concentration) and 1.0 unit Amplitaq Gold. Thermocycling consisted of 12 min at 95°C, 21 cycles of 1 min at 96°C, 1 min at an initial annealing temperature and 1 min at 72°C, using a "computed touchdown" algorithm starting with an annealing temperature of 64°C and ending at 54°C. This was equivalent to reducing the annealing temperature by 0.5°C per cycle. A further 25 cycles were conducted at the annealing temperature of 54°C, with a final polymerase extension for 10 min at 72°C.

## *Polyacrylamide gels*

PCR products were denatured at 95°C for 6 min with 23 μl of formamide (containing bromophenol blue and xylene CFF), placed on ice for 3 min, then loaded onto a 6% polyacrylamide gel (40%; 19:1 bis-acrylamide; National Diagnostics, Hull, UK) containing urea (50% w/v). Gels were loaded with 3 μl of sample and run at 80 W for 3 to 4 hr in 1× GTB buffer (glycerol tolerant buffer, Tris-base, taurine, sodium-EDTA). Gels were moved to Whatman 3MM paper, dried and exposed to autoradiography film (Amersham, Aylesbury, UK).

## *Repeat length determination*

Representative samples were chosen for repeat length determination using ABI 377 and 310 automatic fluorescent DNA sequencers (ABI, Perkin-Elmer) in "Sequencing" and "Genescan" modes of the ABI software. Aliquots of these samples were then run routinely on the gels to aid in the identification of sample genotypes. For (CAG)$_n$ genotyping, a ladder was constructed from samples with 15, 17, 19, 21, 23, 25, 27, 28 and 32 repeats and was run in every 8th lane. For (GGC)$_n$ genotyping, a sample with 16 repeats was run at regular intervals.

## *Statistical analysis*

Odds ratios (ORs) were calculated without age or other adjustments, and all $\chi^2$ calculations were done with correction for continuity. To test for trends in ORs, a $\chi^2$ with 1 degree of freedom was generated according to equation 4.31 in Breslow and Day (1980). Cox proportional hazard analysis, *t*-tests and Kaplan-Meier survival estimates were generated by Stata (College Station, TX) software, version 5.0, which also compared survival curves by log-rank tests.

## RESULTS

The patient and control allele distributions for the (CAG)$_n$ polymorphism are shown in Figure 1*a* and those for the (GGC)$_n$ polymorphism in Figure 1*b*. The allele frequencies for controls were obtained from Caucasian females (see "Material and Methods"). Data are presented as number of chromosomes analysed (390 and 284, respectively). For each region, the same number of repeated triplets was most common in both patients and controls: 21 CAG repeats and 16 GGC repeats. Overall, there was no statistically significant difference between patient and control distributions for either polymorphism. However, there were distinct differences between the 2 polymorphisms. In the (CAG)$_n$ system, 5 of the 19 observed alleles occurred at a frequency of over 10%. In the (GGC)$_n$ system, only 2 alleles, those with 16 and 17 repeats, accounted for over 80% of all 13 alleles genotyped in patient and control populations. A single additional (GGC)$_n$ allele, of 12 repeats, occurred at a frequency of greater than 5%.

The ORs for the risk of prostate cancer by type and length of repeat are shown in Table I. Calculations were performed using the entire series of patients (n = 178) and separately using only those patients for whom complete histological data were available (n =
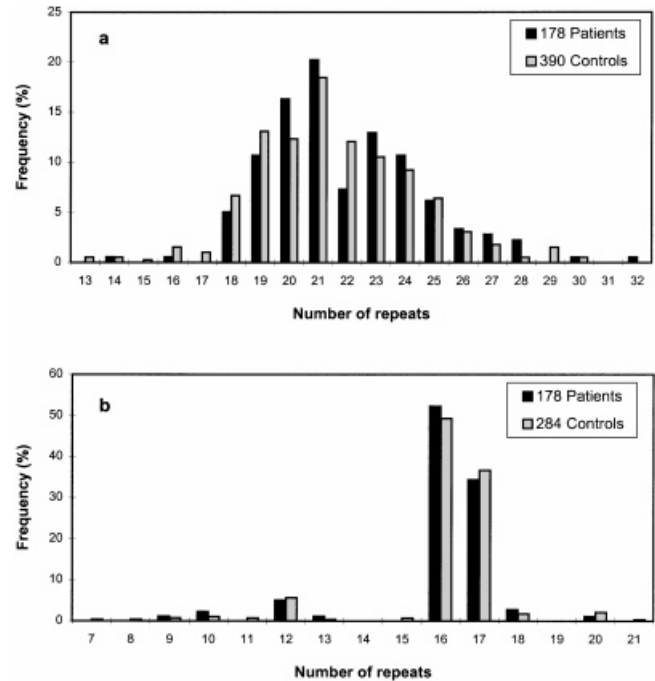


**FIGURE 1** – Distributions of alleles in prostate cancer patients (filled bars) and controls (unfilled bars). *(a)* CAG polymorphic repeat, *(b)* GGC polymorphic repeat. Number of X chromosomes analysed displayed in inset.

**TABLE I** – ORs AND 95% CIs COMPARING LONG *VS.* SHORT *hAR* TRINUCLEOTIDE REPEAT ALLELES IN PATIENTS AND CONTROLS

| Population | CAG repeats | | | GGC repeats | | |
|---|---|---|---|---|---|---|
| | ≤21 | >21 | Total | ≤16 | >16 | Total |
| Patients | 88 (54%) | 74 (46%) | 162 | 98 (60%) | 64 (40%) | 162 |
| Controls | 212 (54%) | 178 (46%) | 390 | 168 (59%) | 116 (41%) | 284 |
| | OR = 1.00 (0.96–1.03) | | | OR = 1.06 (0.57–1.96) | | |
| Patients[1] | 95 (53%) | 83 (47%) | 178 | 110 (62%) | 68 (38%) | 178 |
| | OR = 0.96 (0.52–1.76) | | | OR = 1.12 (0.70–1.76) | | |

[1]Includes 16 patients with unreviewed tumour histology.

**TABLE II** – ORs AND SIGNIFICANCE OF THE PROPORTION OF LONG ALLELES BY DISEASE STAGE AND TESTS FOR LINEAR TRENDS

| Stage | CAG repeats | | | | GGC repeats | | | |
|---|---|---|---|---|---|---|---|---|
| | ≤21 | >21 | OR | *p* | ≤16 | >16 | OR | *p* |
| T1–T2 | 24 | 16 | 1.00 | | 29 | 11 | 1.00 | |
| T3–T4 | 39 | 25 | 1.04 | 0.91 | 37 | 27 | 0.52 | 0.19 |
| N+–M+ | 25 | 33 | 0.51 | 0.15 | 32 | 26 | 0.47 | 0.13 |
| | $\chi^2_{(trend)} = 3.215, p = 0.07$ | | | | $\chi^2_{(trend)} = 2.696, p = 0.10$ | | | |

162). The results indicate that there is no statistical association between prostate cancer risk and the number of repeated sequences in either the (CAG)$_n$ or (GGC)$_n$ polymorphism of *hAR*. Allele distributions were subdivided at other repeat lengths with no significant association found (data not shown). Table II shows the numbers of patients of each clinical stage who had long and short alleles at each polymorphism. Although the ratios of (CAG)$_n$ repeat length (short ≤21 repeats *vs.* long >21 repeats) for stages T3–T4 did not differ from those for T1–T2, there is evidence of a trend for an increasing proportion of longer (CAG)$_n$ repeat lengths in association with metastatic disease; however, this was not significant ($p = 0.07$). A similar non-significant trend ($p = 0.10$) was observed within (GGC)$_n$ repeat length stage-specific proportions. Analysis by histological grade (Table III) indicated that although

neither polymorphism differed significantly, longer alleles may be associated with higher grade.

Table IV presents the results of univariate Cox proportional hazard analysis of factors potentially related to disease free survival (DFS) and overall survival (OS). During the course of the study, 75 (46.3%) of the patients experienced treatment failure (relapse) and 54 (33.3%) died. Each of these possible risk factors was assessed as both a categorical and a continuous variable, and the results did not depend on how the factor was coded. Overall, age at diagnosis was not significantly associated with either DFS or OS.

Disease stage was significantly associated with both DFS and OS. For stages T3–T4, the risks of relapse and death were, respectively, 76% and 87% greater than those for stages T1–T2. The risk of relapse continued to rise but at a slower rate in men with confirmed metastatic spread [N+ or M+; relative risk (RR) = 2.07, 95% CI 1.06–4.03]. This group, however, had nearly a 6-fold increase (RR = 5.99, 95% CI 2.33–15.40) in risk of death compared to men with T1/T2 disease.

Significant trends in DFS and OS were also found in relation to tumour grade. Patients with tumours of moderately differentiated histology (grade 2) were 2.49 times more likely to relapse than

those with well-differentiated tumours (grade 1). The risk significantly rose to over 4-fold (RR = 4.07, 95% CI 1.55–10.67) in men with poorly differentiated tumours (grade 3). Since none of the patients with grade 1 tumours died during the study period, these men were combined with those with grade 2 tumours to establish the baseline risk. Compared to grades 1/2, men with poorly differentiated tumours (grade 3) had a significantly worse OS (RR = 2.40, 95% CI 1.40–4.13).

The $(CAG)_n$ polymorphism was not a significant factor in either DFS or OS. However, in both cases, longer alleles (>21 repeats) were associated with a mild decrease in disease risk. The $(GGC)_n$ trinucleotide repeat polymorphism was significantly related to both DFS and OS when examined in a univariate model. Long $(GGC)_n$ alleles (>16 repeats) increased the risk of relapse by a factor of 1.74 (95% CI 1.08–2.79). These alleles were also associated with a 1.98-fold (95% CI 1.13–3.45) increased risk of death.

In our study, 50 men had pre-treatment PSA levels available for analysis. For these, non-significant associations between PSA levels and DFS and OS were found, which, as expected, indicated a worse prognosis with increasing PSA levels.

Since both stage and grade have clinical relevance and were also significant factors in predicting DFS and OS in our patient population, we examined their associated RRs by Kaplan-Meier survival estimates, to verify their effects. Figure 2a,b presents these survival estimates for DFS. For both risk factors, the expected patterns were observed, i.e., higher stages and grades were associated with decreased DFS.

The results from multivariate Cox proportional hazard analysis are presented in Table V. This analysis included age in years at diagnosis and the factors stage, grade, $(CAG)_n$ and $(GGC)_n$ coded as grouped variables, as in Table IV. Log(PSA) was included in a

**TABLE III** – ORs AND SIGNIFICANCE OF THE PROPORTION OF LONG ALLELES BY DISEASE GRADE AND TESTS FOR LINEAR TRENDS

| Grade | CAG repeats | | | | GGC repeats | | | |
|---|---|---|---|---|---|---|---|---|
| | ≤21 | >21 | OR | p | ≤16 | >16 | OR | p |
| 1 | 14 | 8 | 1.00 | | 15 | 7 | 1.00 | |
| 2 | 43 | 45 | 0.55 | 0.32 | 54 | 34 | 0.74 | 0.73 |
| 3 | 31 | 21 | 0.84 | 0.95 | 29 | 23 | 0.59 | 0.46 |
| | $\chi^2_{(trend)}$ = 0.029, p = 0.865 | | | | $\chi^2_{(trend)}$ = 1.045, p = 0.307 | | | |

**TABLE IV** – CHARACTERISTICS OF 162 PROSTATE CANCER PATIENTS AND UNIVARIATE COX PROPORTIONAL HAZARDS ANALYSIS: RRs OF FACTORS POTENTIALLY RELATED TO DISEASE-FREE SURVIVAL AND OVERALL SURVIVAL

| Factor | Total number | Disease-free survival | | Overall survival | |
|---|---|---|---|---|---|
| | | Number | RR (95% CI) | Number | RR (95% CI) |
| Total | 162 | 75 | | 54 | |
| Age (years) | | | | | |
| <60 | 17 | 9 | 1.00 | 5 | 1.00 |
| 60–79 | 72 | 37 | 1.06 (0.51–2.21) | 29 | 1.72 (0.66–4.48) |
| 80+ | 73 | 29 | 0.85 (0.40–1.82) | 20 | 1.32 (0.48–3.60) |
| Continuous | | | 0.99 (0.96–1.03) | | 1.02 (0.98–1.07) |
| Stage | | | | | |
| T1–T2 | 40 | 14 | 1.00 | 7 | 1.00 |
| T3–T4 | 64 | 33 | 1.76 (0.94–3.30) | 15 | 1.87 (0.72–4.82) |
| N+–M+ | 58 | 28 | 2.07 (1.06–4.03) | 32 | 5.99 (2.33–15.40) |
| Continuous | | | 1.35 (1.01–1.82) | | 2.61 (1.71–3.99) |
| Grade | | | | | |
| 1 | 22 | 5 | 1.00 | 0 | |
| 2 | 88 | 41 | 2.49 (0.98–6.30) | 28 | 1.00[1] |
| 3 | 52 | 29 | 4.07 (1.55–10.67) | 26 | 2.40 (1.40–4.13) |
| Continuous | | | 1.78 (1.24–2.57) | | 2.53 (1.60–4.00) |
| Unknown[2] | 16 | 5 | 1.74 (0.50–6.03) | 9 | 2.34 (1.09–5.01) |
| CAG (repeats) | | | | | |
| ≤21 | 88 | 45 | 1.00 | 30 | 1.00 |
| >21 | 74 | 30 | 0.73 (0.45–1.16) | 24 | 0.92 (0.53–1.58) |
| Continuous | | | 0.95 (0.87–1.04) | | 1.01 (0.91–1.12) |
| GGC (repeats) | | | | | |
| ≤16 | 98 | 42 | 1.00 | 29 | 1.00 |
| >16 | 64 | 33 | 1.74 (1.08–2.79) | 25 | 1.98 (1.13–3.45) |
| Continuous | | | 1.18 (1.02–1.37) | | 1.23 (1.03–1.46) |
| PSA[3] (ng/ml) | | | | | |
| <10 | 9 | 2 | 1.00 | 3 | 1.00 |
| 10–49 | 20 | 9 | 1.81 (0.38–8.55) | 4 | 0.53 (0.12–2.40) |
| 50+ | 21 | 10 | 2.39 (0.51–11.08) | 8 | 1.17 (0.30–4.52) |
| Continuous | | | 1.00 (1.00–1.00) | | 1.00 (1.00–1.00) |
| Log(continuous) | | | 1.53 (0.84–2.82) | | 1.92 (0.96–3.85) |

[1]Grades 1 and 2 were combined since no deaths occurred among grade 1 patients.–[2]These patients with unknown grade were excluded from further analysis unless otherwise stated.–[3]PSA measurements were available for only 50 men, of whom 21 experienced relapse and 15 died during the study.
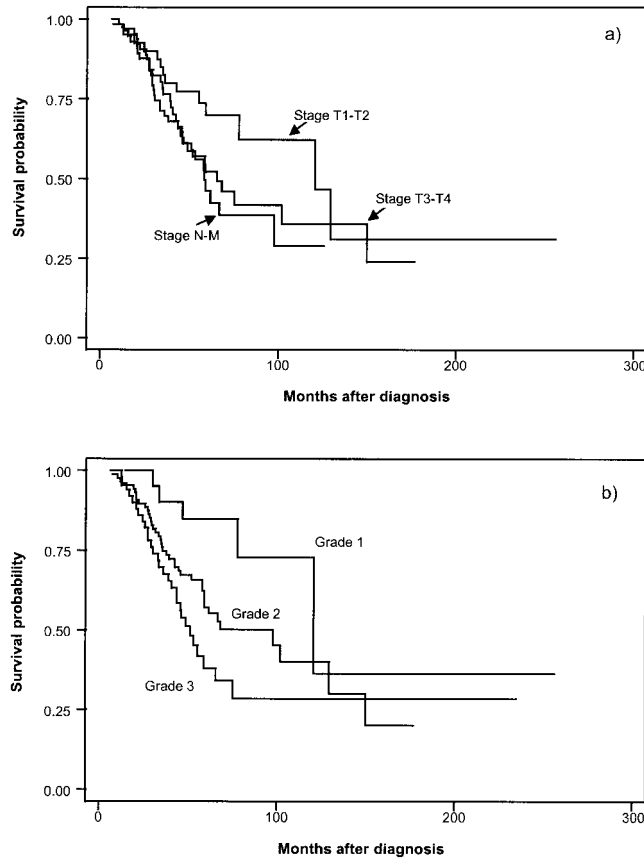
**FIGURE 2** – Kaplan-Meier estimates of disease-free survival by *(a)* stage and *(b)* grade. Stage and grade indicated.

2nd multivariate analysis on the reduced set of men for whom measurements were available. For the larger data set of 162 men, only grade remains significantly associated with DFS and OS. Disease stage loses significance in DFS but remains a consistently significant predictor of OS. Both of the trinucleotide polymorphisms become non-significant variables in DFS and OS, with a range of $0.10 > p > 0.05$.

Figure 3*a,b* presents the Kaplan-Meier estimates of DFS and OS, respectively, partitioned according to the $(GGC)_n$ genotype groups, $\leq 16$ repeats and $> 16$ repeats. Log-rank tests indicate that this $(GGC)_n$ classification results in overall significant differences in DFS ($p = 0.020$) and OS ($p = 0.014$). However, in the case of DFS (Fig. 3*a*), it is evident that the RR or proportional hazard associated with long alleles was not constant throughout the follow-up period; *i.e.,* it was time-dependent, changing from being associated with a slightly better to an increasingly poor survival. By examining the detailed survivorship functions, it was found that the increased risk associated with long $(GGC)_n$ alleles became effective only at 33 months after diagnosis. From this time through 72 months, the association of long alleles with worse survival was progressively more severe. In both genotype groups, the survival probability did not fall below approximately 0.25 during the course of the study. A similar delayed effect of 33 months before long $(GGC)_n$ alleles became prominently associated with worse OS was also observed (Fig. 3*b*).

$(GGC)_n$ genotype was significantly associated with DFS and OS when tested as a univariate predictor but became non-significant, $p = 0.052$ (DFS) and $p = 0.088$ (OS), when stage and grade were simultaneously included. To explain this finding, we thought that there may be some interaction between $(GGC)_n$ and stage and/or $(GGC)_n$ and grade in terms of their effects on DFS and OS. This

possibility was also suggested by the previous finding of long $(GGC)_n$ alleles being mildly correlated with stage (Table II) and grade (Table III). We tested each of the other possible risk factors, age, grade, stage and $(CAG)_n$ genotype, in a 2-factor analysis of DFS and OS and found that the inclusion of grade information was most responsible for the decrease in $(GGC)_n$ effect, with stage having the second largest effect on $(GGC)_n$ (data not shown). We therefore analysed $(GGC)_n$ genotype separately for 2 levels each of grade and stage. These results are presented in Table VI. In both measures of disease outcome, DFS and OS, long $(GGC)_n$ alleles were associated with a worse prognosis at all levels of stage and grade. Graded effects of $(GGC)_n$ genotype on DFS were associated with both risk factors, with the largest effects in early, T1–T2, stages (RR = 3.56, 95% CI 1.13–11.21) and grade 1 (RR = 6.47, 95% CI 0.57–72.8) tumours. Long $(GGC)_n$ alleles were also associated with decreased OS at all levels of stage and grade, though the RRs were smaller and the variation in the effects among the clinical/pathological classes was less than for DFS. In this case, the overall trend by stage was not evident; the highest risk, RR = 2.22 (95% CI 0.76–6.44), occurred in stages T3–T4. The effect of $(GGC)_n$ length was significant in the combined group, grades 1 and 2 (RR = 2.21, 95% CI 1.01–4.85). A similar analysis of the $(CAG)_n$ sequence showed that this polymorphism was not a significant factor in DFS or OS in any of the stage/grade categories (data not shown).

As Irvine *et al.* (1995) reported linkage disequilibrium between the *hAR* polymorphisms in prostate cancer patients, we tested the independence of 2 classifications of $(CAG)_n$ alleles with 3 classes of $(GGC)_n$ alleles by Pearson's $\chi^2$ test for independence of rows and columns, and the results were non-significant ($\chi^2_{(df=2)} = 1.31$, $p = 0.511$). We also performed *t*-tests to estimate the significance of the difference in mean $(CAG)_n$ repeat number when partitioned by $(GGC)_n$ length. The difference of 0.23 repeats was also not significant ($t = 0.563$, $p = 0.574$).

### DISCUSSION

In the British population, we have found that compared to controls overall prostate cancer risk was not significantly associated with either *hAR* $(CAG)_n$ or $(GGC)_n$ variation. This result is in accord with several other studies in which no overall significant genotype-specific prostate cancer risk was found with either the $(CAG)_n$ or $(GGC)_n$ polymorphism compared with control groups (Irvine *et al.,* 1995; Giovannucci *et al.,* 1997; Stanford *et al.,* 1997). When we looked for stage and grade associations, we found similar effects in the 2 polymorphisms; *i.e.,* long alleles in both polymorphisms appeared at higher frequency in higher stages and higher grades. Although the association was stronger for the $(GGC)_n$ polymorphism, the suggested association for the $(CAG)_n$ system is opposite to what other investigators have generally found. One concern was that the clinical staging schemes are different in the United Kingdom (TNM system) than in the United States (ABC system). However, since both systems reflect the natural growth and metastatic potential of prostate cancer and score tumour progression on the same continuum, from intracapsular to capsular penetration to documented metastasis, we do not believe this would have affected our results.

We next examined the 2 polymorphisms as well as 4 other potential risk factors for their involvement in DFS and OS. Disease stage and grade were individually significant factors for each of these end-point measures. These correlations were not unexpected since both stage and grade are used as prognostic indicators in the clinic. Also in the univariate analysis, the $(CAG)_n$ system behaved in a manner similar to that hypothesised by other workers in that short $(CAG)_n$ alleles were associated with more aggressive disease. In our patient population, men with long $(CAG)_n$ alleles were at less risk of relapse and death than those with short alleles. The risk differentials were not significant however. $(GGC)_n$ variation was significantly related to DFS and OS in that longer alleles (*i.e.,* those with 17 or more repeats) were associated with shorter time to

**TABLE V** – MULTIVARIATE COX PROPORTIONAL HAZARDS ANALYSIS: RRs AND $p$ VALUES FOR RISK FACTORS ON DISEASE-FREE SURVIVAL AND OVERALL SURVIVAL

| Factor (n groups) | Disease-free survival | | | | Overall survival | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Grade reviewed [n = 162 (75)][1] | | PSA available [n = 50 (21)][1] | | Grade reviewed [n = 162 (54)][1] | | PSA available [n = 50 (15)][1] | |
| | RR | $p$ | RR | $p$ | RR | $p$ | RR | $p$ |
| Age (years) | 0.99 | 0.414 | 0.91 | 0.005 | 1.01 | 0.517 | 0.98 | 0.525 |
| Stage (3) | 1.30 | 0.103 | 0.86 | 0.660 | 2.81 | 0.001 | 1.46 | 0.438 |
| Grade (3) | 1.71 | 0.005 | 2.56 | 0.024 | 2.39 | 0.001 | 4.17 | 0.006 |
| CAGL (2) | 0.65 | 0.080 | 0.49 | 0.175 | 0.61 | 0.097 | 0.37 | 0.125 |
| GGCL (2) | 1.60 | 0.052 | 2.06 | 0.173 | 1.65 | 0.088 | 4.33 | 0.021 |
| Log(PSA) (ng/ml) | | | 1.70 | 0.203 | | | 1.63 | 0.239 |

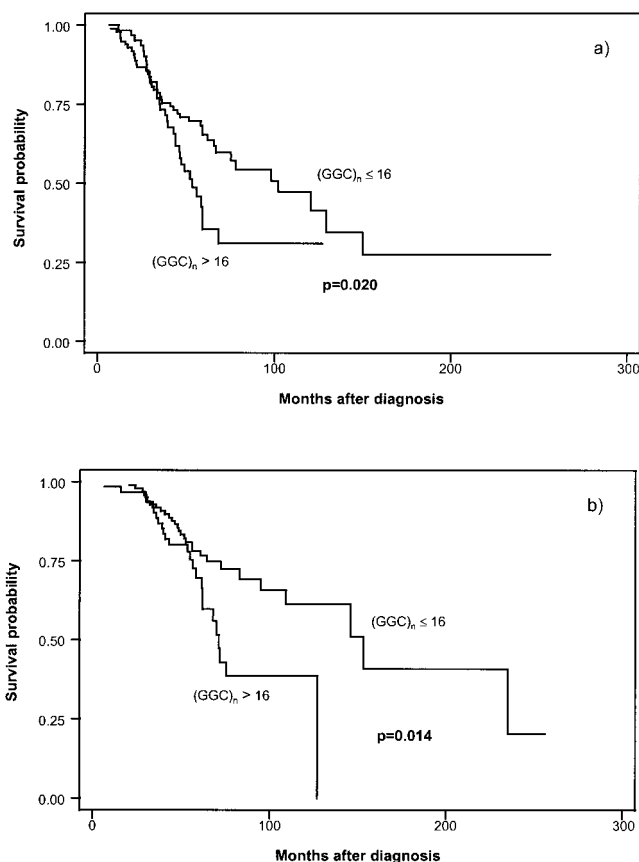[1]Number at risk (number affected).–CAGL and GGCL, alleles grouped as long (L), see Table IV.

**FIGURE 3** – Kaplan-Meier survival estimates by $(GGC)_n$ length *(a)* for disease-free survival and *(b)* for overall survival. Patients catagorised by repeat length as indicated. $p$ values indicate log-rank significance tests.

relapse and worse OS. This $(GGC)_n$ correlation is a new finding and much stronger than, and independent of, DFS and OS association with $(CAG)_n$ variation.

These $(GGC)_n$ results indicate that the AR polyglycine tract may have a functional role, at least in regulating tumour growth, if not in normal physiology. Previous evidence of $(GGC)_n$ involvement in prostate cancer has pertained mainly to population risk estimates and has been ambiguous. Irvine *et al*. (1995) studied 68 prostate cancer patients from Los Angeles County, CA, and found that neither the $(GGC)_n$ nor the $(CAG)_n$ polymorphism alone was significantly related to risk compared to controls. In the 57 Caucasian patients, however, the combination of short $(CAG)_n$ alleles and alleles with other than 16 repeats [*i.e.,* essentially alleles with 17 repeats, similar to our $(GGC)_n$ higher-risk alleles] produced

**TABLE VI** – UNIVARIATE COX PROPORTIONAL HAZARDS ANALYSIS: RRs OF GGC GENOTYPE ON DISEASE-FREE SURVIVAL AND OVERALL SURVIVAL BY STAGE AND GRADE

| Factor | Number at risk | Disease-free survival | | Overall survival | |
| --- | --- | --- | --- | --- | --- |
| | | Number | RR (95% CI) | Number | RR (95% CI) |
| Stage | | | | | |
| T1–T2 | 40 | 14 | 3.56 (1.13–11.21) | 7 | 2.16 (0.36–12.9) |
| T3–T4 | 64 | 33 | 1.76 (0.87–3.54) | 15 | 2.22 (0.76–6.44) |
| N+–M+ | 58 | 28 | 1.03 (0.47–2.25) | 32 | 1.65 (0.80–3.45) |
| Grade | | | | | |
| 1 | 22 | 5 | 6.47 (0.57–72.8) | | |
| 2 | 88 | 41 | 1.68 (0.88–3.18) | 28[1] | 2.21 (1.01–4.85) |
| 3 | 52 | 29 | 1.23 (0.59–2.59) | 26 | 1.44 (0.66–3.16) |

[1]Grades 1 and 2 were combined since no deaths occurred among grade 1 patients.

an RR of 2.10 ($p$ = 0.08). Although the patient sample was approximately split between organ-confined and advanced disease classes, analysis by stage was not presented. Stanford *et al*. (1997) studied 301 patients and 277 controls for effects of both $(CAG)_n$ and $(GGC)_n$ on prostate cancer risk in the King County (Seattle, WA) area. In this study, $(CAG)_n$ was not a significant factor in risk but the estimate (RR = 1.23) suggested that men with short alleles were at higher risk. However, in the same report, a significant effect with $(GGC)_n$ alleles was found (RR = 1.60, 95% CI 1.07–2.41), and this was also in the direction of shorter alleles carrying increased risk, *i.e.,* similar to our overall population-based risk result. Additionally, the combined effect of the 2 polymorphisms remained higher at RR = 2.05, which is congruent with independent effects. Further analysis revealed significant interactions between the $(GGC)_n$ system and disease aggressiveness, Quetelet index, family history and age. The association with disease aggressiveness was also suggested by our results. Data presented by Hakimi *et al*. (1997) suggest that short $(GGC)_n$ repeats, in addition to increasing the risk of prostate cancer, identify a subpopulation of patients with clinically localised disease. The report suggests that short glycine repeats may identify men who develop prostate cancer but with decreased aggressiveness. This is an intriguing suggestion and supports our observations on OS and DFS.

Irvine *et al*. (1995) found a significant risk association between short $(CAG)_n$ alleles and long $(GGC)_n$ alleles in all patients ($p$ = 0.008) as well as when white patients only were examined ($p$ = 0.015). In contrast and in agreement with Stanford *et al*. (1997) and Lumbroso *et al*. (1997), we found no evidence for allelic association between the 2 polymorphisms.

To our knowledge, only 2 studies have reported a functional role of the polyglycine tract *in vitro*: Jenster *et al*. (1994) found that complete deletion of the $(GGC)_n$ sequence had "no substantial effect on the activity of the hormone regulated AR", whereas Gao *et al*. (1996) found that this same mutation resulted in a diminished capacity to activate the MMTV luciferase gene. Complex protein interactions involving steroid receptors and positive and negative

acting co-factors are beginning to be elucidated (Shibata *et al.,* 1997). Since the $(GGC)_n$ sequence is in the transactivation domain, it is possible that the single–amino acid difference disrupts the binding affinity of the domain enough to up- or down-regulate the formation of a critical regulatory complex.

A graded relationship or interaction between the effects of $(GGC)_n$ genotype and tumour stage and grade in DFS was also found in our study and is important for 2 reasons. First, it suggests that the role of $(GGC)_n$ in regulating tumour growth is likely to be of heightened relevance during a limited time period out of the total multiple-step progression required for the transition from normal cell growth regulation to fully hormone-independent metastatic disease. Additionally, since the greatest RRs occurred in the earliest-stage tumours, these effects of $(GGC)_n$ genotype are likely to occur relatively early in the pathway. Second, it suggests potential clinical application, and this aspect is discussed later.

We have specifically examined the role of both *hAR* trinucleotide repeat polymorphisms as they relate to disease outcome. We found no evidence of an association between the $(CAG)_n$ polymorphism and prostate cancer risk but a significant association between the $(GGC)_n$ polymorphism and disease progression and outcome. We examined *hAR* genotypes in a population outside the United States. Whether the present findings are unique to British patients is unknown. However, dietary and other environmental factors have been associated with the clinical incidence of prostate cancer (Rose *et al.,* 1986), and it has been suggested that the metabolic link between fat intake and prostate cancer risk may include androgenic pathways (Ross and Henderson, 1994). Also, it has often been found that the fat–prostate cancer risk correlation may be strongest in men with high-grade/advanced cancers (Whittemore *et al.,* 1995; Ross and Henderson, 1994). It is therefore possible that gene–environment interactions may affect the relationship between the *hAR* polymorphisms and prostate cancer progression

With regard to clinical considerations, treatment option decisions for early-stage tumours, particularly those detected through PSA screening, are becoming frequent and have both economic as well as quality-of-life implications. Although further research is needed to validate and assess our present findings in the clinical setting, use of the $(GGC)_n$ genotype may offer relevant information to help predict the growth behaviour of early-stage tumours and determine who should be offered radical treatment.

## REFERENCES

BRESLOW, N.E. and DAY, N.E., *The analysis of case-control studies,* IARC, Lyon (1980).

CARTER, H.B., PIANTADOSI, S. and ISAACS, J.T., Clinical evidence for and implications of the multistep development of prostate cancer. *J. Urol.,* **143,** 742–746 (1990).

CHAMBERLAIN, N.L., DRIVER, E.D. and MIESFELD, R.L., The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res.,* **22,** 3181–3186 (1994).

CHODAK, G.W., THISTED, R.A., GERBER, G.S., JOHANSSON, J.E., ADOLFSSON, J., JONES, G.W., CHISHOLM, G.D., MOSKOVITZ, B., LIVNE, P.M. and WARNER, J., Results of conservative management of clinically localized prostate cancer. *N. Engl. J. Med.,* **330,** 242–248 (1994).

CUNHA, G.R., DONJACOUR, A.A., COOKE, P.S., MEE, S., BIGSBY, R.M., HIGGINS, S.J. and SUGIMURA, Y., The endocrinology and developmental biology of the prostate. *Endocrinol. Rev.,* **8,** 338–362 (1987).

DEARNALEY, D.P., Cancer of the prostate. *Brit. med. J.,* **308,** 780–784 (1994).

EDWARDS, S.M., DEARNALEY, D.P., ARDERN-JONES, A., HAMOUDI, R.A., EASTON, D.F., FORD, D., SHEARER, R., DOWE, A., THE CANCER RESEARCH CAMPAIGN/BRITISH PROSTATE GROUP FAMILIAL PROSTATE CANCER STUDY COLLABORATORS and EELES, R.A., No germline mutations in the dimerization domain of MXI1 in prostate cancer clusters. *Brit. J. Cancer,* **76,** 992–1000 (1997).

GAO, T., MARCELLI, M. and MCPHAUL, M.J., Transcriptional activation and transient expression of the human androgen receptor. *J. Steroid Biochem. mol. Biol.,* **59,** 9–20 (1996).

GIOVANNUCCI, E., STAMPFER, M.J., KRITHIVAS, K., BROWN, M., DAHL, D., BRUFSKY, A., TALCOTT, J., HENNEKENS, C.H. and KANTOFF, P.W., The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc. nat. Acad. Sci. (Wash.),* **94,** 3320–3323 (1997).

HAKIMI, J.M., SCHOENBERG, M.P., RONDINELLI, R.H., PIANTADOSI, S. and BARRACK, E.R., Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. *Clin. Cancer Res.,* **3,** 1599–1608 (1997).

HARDY, D.O., SCHER, H.I., BOGENREIDER, T., SABBATINI, P., ZHANG, Z.F., NANUS, D.M. and CATTERALL, J.F., Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. *J. clin. Endocrinol. Metab.,* **81,** 4400–4405 (1996).

INGLES, S.A., ROSS, R.K., YU, M.C., IRVINE, R.A., LA PERA, G., HAILE, R.W. and COETZEE, G.A., Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. *J. nat. Cancer Inst.,* **89,** 166–170 (1997).

IRVINE, R.A., YU, M.C., ROSS, R.K. and COETZEE, G.A., The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res.,* **55,** 1937–1940 (1995).

ISAACS, J.T., Antagonistic effect of androgen on prostatic cell death. *Prostate,* **5,** 545–557 (1984).

JENSTER, G., DE RUITER, P.E., VAN DER KORPUT, H.A., KUIPER, G.G., 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 sites. *Biochemistry,* **33,** 14064–14072 (1994).

KAZEMI-ESFARJANI, P., TRIFIRO, M.A. and PINSKY, L., Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n-expanded neuronopathies. *Hum. mol. Genet.,* **4,** 523–527 (1995).

KYPRIANOU, N. and ISAACS, J.T., Activation of programmed cell death in the rat ventral prostate after castration. *Endocrinology,* **122,** 552–562 (1988).

LU, S., TSAI, S.Y. and TSAI, M.J., Regulation of androgen-dependent prostatic cancer cell growth: androgen regulation of CDK2, CDK4, and CKI p16 genes. *Cancer Res.,* **57,** 4511–4516 (1997).

LUMBROSO, R., BEITEL, L.K., VASILIOU, D.M., TRIFIRO, M.A. and PINSKY, L., Codon-usage variants in the polymorphic (GGN)n trinucleotide repeat of the human androgen receptor gene. *Hum. Genet.,* **101,** 43–46 (1997).

RIEGMAN, P.H., VLIETSTRA, R.J., VAN DER KORPUT, J.A., BRINKMANN, A.O. and TRAPMAN, J., The promoter of the prostate-specific antigen gene contains a functional androgen responsive element. *Mol. Endocrinol.,* **5,** 1921–1930 (1991).

ROSE, D.P., BOYAR, A.P. and WYNDER, E.L., International comparisons of mortality rates for cancer of the breast, ovary, prostate, and colon, and per capita food consumption. *Cancer,* **58,** 2363–2371 (1986).

ROSS, R.K. and HENDERSON, B.E., Do diet and androgens alter prostate cancer risk via a common etiologic pathway? *J. nat. Cancer Inst.,* **86,** 252–254 (1994).

SADI, M.V., WALSH, P.C. and BARRACK, E.R., Immunohistochemical study of androgen receptors in metastatic prostate cancer. Comparison of receptor content and response to hormonal therapy. *Cancer,* **67,** 3057–3064 (1991).

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 (CAG24→18) in the androgen receptor gene in human prostate cancer. *Biochem. biophys. Res. Comm.,* **198,** 74–80 (1994).

SHIBATA, H., SPENCER, T.E., ONATE, S.A., JENSTER, G., TSAI, S.Y., TSAI, M.J. and O'MALLEY, B.W., Role of co-activators and co-repressors in the mechanism of steroid/thyroid receptor action. *Rec. Progr. Horm. Res.,* **52,** 141–164 (1997).

SHIMIZU, H., ROSS, R.K., BERNSTEIN, L., YATANI, R., HENDERSON, B.E. and MACK, T.M., Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. *Brit. J. Cancer,* **63,** 963–966 (1991).

SLEDDENS, H.F., OOSTRA, B.A., BRINKMANN, A.O. and TRAPMAN, J., Trinucleotide repeat polymorphism in the androgen receptor gene (AR). *Nucleic Acids Res.,* **20,** 1427 (1992).

SLEDDENS, H.F., OOSTRA, B.A., BRINKMANN, A.O. and TRAPMAN, J., Trinucleotide (GGN) repeat polymorphism in the human androgen receptor (AR) gene. *Hum. mol. Genet.,* **2,** 493 (1993).

STANFORD, J.L., JUST, J.J., GIBBS, M., WICKLUND, K.G., NEAL, C.L., BLUMENSTEIN, B.A. and OSTRANDER, E.A., Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Res.,* **57,** 1194–1198 (1997).

SWERDLOW, A.J., COOKE, K.R., SKEGG, D.C. and WILKINSON, J., Cancer incidence in England and Wales and New Zealand and in migrants between the two countries. *Brit. J. Cancer,* **72,** 236–243 (1995).

UK NATIONAL CASE-CONTROL STUDY GROUP, Oral contraceptive use and breast cancer risk in young women. UK National Case-Control Study Group. *Lancet,* **1,** 973–982 (1989).

VELDSCHOLTE, J., RIS-STALPERS, C., KUIPER, G.G., JENSTER, G., BERREVOETS, C., CLAASSEN, E., VAN ROOIJ, H.C., TRAPMAN, J., BRINKMANN, A.O. and MULDER, E., A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens. *Biochem. biophys. Res. Comm.,* **173,** 534–540 (1990).

VISAKORPI, T., HYYTINEN, E., KOIVISTO, P., TANNER, M., KEINANEN, R., PALMBERG, C., PALOTIE, A., TAMMELA, T., ISOLA, J. and KALLIONIEMI, O.P., *In vivo* amplification of the androgen receptor gene and progression of human prostate cancer. *Nature Genet.,* **9,** 401–406 (1995).

WHITTEMORE, A.S., KOLONEL, L.N., WU, A.H., JOHN, E.M., GALLAGHER, R.P., HOWE, G.R., BURCH, J.D., HANKIN, J., DREON, D.M., WEST, D.W., TEH, C. and PAFFENBARGER, R.S., Prostate cancer in relation to diet, physical activity, and body size in blacks, whites, and Asians in the United States and Canada. *J. nat. Cancer Inst.,* **87,** 652–661 (1995).