ARTICLE

(CAG)nCAA and GGN repeats in the human androgen receptor gene are not associated with prostate cancer in a French–German population

Lina Correa-Cerro[1,2], Gudrun Wöhr[1], Jürgen Häussler[1], Philippe Berthon[3], Eric Drelon[3], Philippe Mangin[3], Georges Fournier[4], Oliver Cussenot[3], Petra Kraus[1], Walter Just[1], Thomas Paiss[5], José María Cantú[2] and Walther Vogel[1]

[1]Department of Medical Genetics, University of Ulm, Germany
[2]Department of Human Genetics, University of Guadalajara, Mexico
[3]Saint Louis Hospital, Paris, France
[4]Service d'Urologie, CHU de la Cavale Blanche, Brest, France
[5]Department of Urology, University of Ulm, Germany

Alleles of the CAG and the GGC repeat in the first exon of the human androgen receptor (AR) gene have been shown to be associated with the risk of (advanced) prostate cancer. These studies had been carried out in the United States. We have analysed these polymorphisms in a French–German collection of 105 controls, 132 sporadic cases, and a sample of prostate cancer families comprising 85 affected and 46 not affected family members. The allele distributions were very similar in all four groups and chi square statistics on contingency tables did not detect any significant differences. The relative risk (odds ratio, OR) were calculated using logistic regression and did not reach significance despite sufficient numbers of patients and controls. Typical results were OR = 1.007; 95% Confidence Interval (CI) 0.97–1.1, $P = 0.87$ for CAG as continuous variable and OR = 1.2 (95% CI 0.7–2.0), $P = 0.47$ for CAG classes < 22 and > = 22 repeats. Similar results were obtained for subgroups defined by age or Gleason score. We conclude that these polymorphisms can not be used as predictive parameters for prostate cancer in the French or German population.

Keywords: human androgen receptor gene; CAG and GGC microsatellites; prostate cancer

Introduction

Androgens are required for the development and maintenance of normal prostate and act via the

androgen receptor (AR).[1,2] The *AR* gene, located in the Xq11, spans more than 90 kb of genomic DNA, and the resulting transcription codes a protein with three major functional domains. The N-terminal domain of the protein which serves a transactivation function is encoded by the 1586 bp exon 1.[3] This exon contains two polymorphic repeats; a CAG stretch in the 5' portion and a GGC stretch in the 3' portion which encode polyglutamine and polyglycine tracts, respectively.[4,5]

The CAG repeat has a normal length from 8 to 31 repeats and averages about 20, whilst the normal length of the GGC repeat is about 16.[4,6,7] Since the number of CAG repeats has been shown to modulate AR expression[8] and the AR activity may influence the development of prostate cancer, an association between CAG repeat length and prostate cancer could have a functional basis. Several studies in American populations revealed associations of shorter CAG alleles (< 22 or < 18) to incidence, or age of onset, or advanced stage prostate cancer.[4,6,9–11] Only one Canadian study on sibs did not give results compatible with this association.[12] The GGC repeat has been studied and found longer (> 16 repeats) or shorter (< 16 repeats) in prostate cancer patients but no linkage disequilibrium between these two repeats was observed.[4,11] The risk of prostate cancer is well known to vary according to the ethnic origin of the population studied, with Africans having the highest and Asians the lowest risk. The same risk distribution has been observed in North America and it has been shown that the length of the AR triplet repeats varies accordingly.[4]

We investigated the number of GAC and GGC repeats in controls, sporadic cases with prostate cancer, familial prostate cancer and their non-affected relatives in a French–German collection.

Materials and Methods

Sporadic Cases and Controls
Our study was carried out on 132 patients with clinical diagnosis of prostate cancer. They were recruited as consecutive cases of prostate cancer in two hospitals (Nancy and Paris). The diagnosis has always been confirmed histologically and the patients gave informed consent to participate in a molecular genetic study on prostate cancer (consent from more than 90%). The cases were considered sporadic if they did not have an affected first degree relative and if there was no more than one affected relative more distantly related. For these patients, age of onset (mean 68.2, range 46–90 years) and Gleason score (mean 5.8, ± 1.7 SD) were available.

The 105 controls (age distribution: mean 71.2, range 64–86 years) had originally been recruited in companies in Paris in a sample comprising 700 individuals in 1989. They were followed by digital rectal examination (DRE) and prostate specific antigene (PSA) level for 7 years and found free of any signs of prostate cancer. In the case of ambiguous results these were verified by ultrasound and biopsy.

We used these highly stringent controls because prostate cancer is a late onset disorder and other types of controls (eg matched pairs) are bound to be contaminated by men who are going to develop prostate cancer later.

Familial Cases
From the 47 families (criteria according to Carter *et al*[13] used for linkage analysis and described in detail by Berthon *et al*[14]

we selected 28 potentially informative for X chromosomal inheritance because DNA of the mother was available or there were four or more brothers. Thus, we could determine if the affecteds did preferentially inherit the shorter or longer allele from the mother. In addition, we used ten families from Ulm, Germany, selected according to the same criteria which were not included in the earlier linkage study.

Genetic Analysis
About 50 to 150 ng of genomic DNA was used in polymerase chain reaction (PCR) to amplify the CAG or GGC region in exon 1 of the *AR* gene. DNA was extracted by standard methods from whole blood. The forward primer AR1 (5'-TCCAGAATCTGTTCCAGAGCGTGC-3') was taken from Giovannucci *et al*.[6] The reversed primer AR2 (5'-CAA-GAGGGGTTCGGGTAGCATCTC-3') was designed on the basis of GenBank sequences of the *AR* gene and used to amplify the CAG region. AR1 was labelled with Cy5, allowing fluorescent detection. PCR was started by initial denaturation 95°C for 5 min, followed by 32 cycles, each 95°C for 45 s, 61°C for 30 s and 72°C for 30 s. Final extension was carried out at 72°C for 5 min. We used Amplitaq DNA polymerase (Perkin-Elmer, Stuttgart, Germany).

The region of the GGC stretch was amplified by the PCR primers AR5 (5'-CTATGGGGACCTGGCGAGC-3') and AR6 (5'-GGTGCGGTGAAGTCGGCTTTC-3'). The corresponding PCR program consisted of 35 cycles with 1 min at 98°C and 5 min at 70°C as described by Irvine *et al*.[4] The amplification of the GGC region was carried out with the proof reading *pfu* DNA polymerase (Stratagene, Heidelberg, Germany). Due to the high GC content of the latter region some samples could not be amplified under these conditions. In these cases we used the nested primer described by Irvine *et al*[4] with a fluorescent label and the same PCR conditions as described above.

The resulting DNA fragments were run on a 6% denaturing polyacrylamide gel with automated fluorescence detection (ALFexpress; Pharmacia Biotech, Freiburg, Germany). The length of the PCR products and the size of the trinucleotide repeat were calculated with the software Fragment Manager (Pharmacia Biotech). To calculate the number of repeats, we cloned and sequenced one CAG and one GGC stretch of a reference. These were used as references to determine the number of repeats in the test samples.

Statistical Analysis
The frequencies of the different CAG and GGC alleles were determined for the different groups (controls, sporadic cases, familial not affected, and familial cases) using the single alleles as classes. These frequencies as well as classes of grouped alleles (see below) were used in contingency tables and compared between the groups by the χ^2 test. In the families, we determined whether the shorter or the longer allele had been inherited and compared these frequencies between affecteds and non-affecteds by the χ^2 test.

The repeat numbers of the CAG and GGC repeats were tested for correlation by the Kruskal-Wallis test and the Spearman rank correlation test statistics. After these primary tests, the analysis of association between AR alleles and prostate cancer risk was restricted to the CAG repeat because there was obviously no effect of the GGC repeat and there was no correlation between the two repeats.

Since the logistic regression model with the CAG number as a continuous variable did not reveal any significant result

the analysis was repeated using the repeat number as a nominal variable. To this end, the alleles had to be grouped. The following categories were used with sporadic cases and controls: (A) single alleles, (B) < 22 and > = 22 according to Ingles *et al*[p] and (C) < = 19, 20/21, 22/23, > = 24, which gave similar frequencies of cases and controls in these four classes of our sample.

In addition, we used the categories < 22 and > = 22 repeats to analyse sub-samples of the sporadic cases with age of diagnosis < = 60, < 65, < 70 and > = 70 years old. Similarly the Gleason score was used to construct and analyse sub-samples defined by tumour progression (Gleason score 2–5, 2–6, 4–10 and 7–10).

The statistical analyses were performed using the StatView program (SAS, USA).

Results

The number of CAG repeats in sporadic and familial prostate cancer, in controls, and in familial non-affected individuals are shown in Figure 1. The CAG repeat length ranges from 14 to 31 in controls and from 13 to 30 in sporadic prostate cancers. In the familial cases the range was from 13 to 26 in the non-affected and from 9 to 32 in the familial-affected subgroup. We observed a bimodal distribution of the CAG repeat with peaks at 21 and 23 repeats in all of the populations studied.

The number of GGC repeats ranges from 9 to 29 in controls and from 11 to 28 in sporadic prostate cancer.
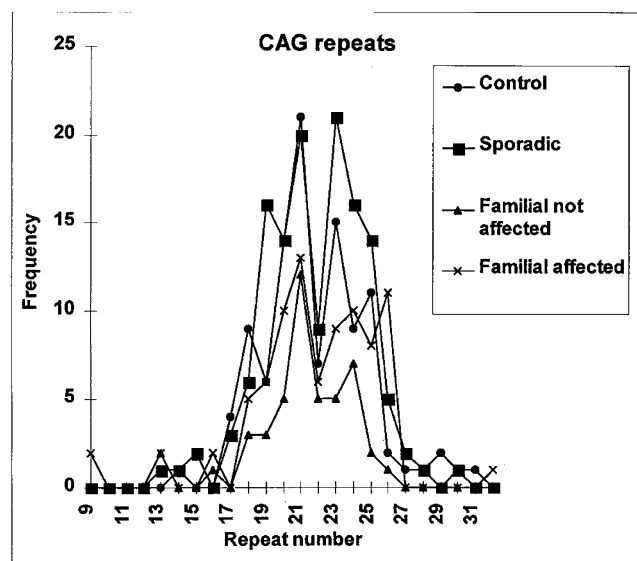
In familial cancer cases we observed a range from 13 to 20, and in the familial non-affected group from 13 to 22 repeats. The predominant allele had 17 repeats in all of these populations (Figure 2).

The statistical analysis comparing repeat distribution between the groups did not reveal any significant differences, eg the Kruskal-Wallis test on all non-affecteds *vs* all affecteds yielded $\chi^2 = 0.587$ at 1 DF with $P = 0.443$ similar to the Wilcoxon two sample test.

Although the allele distribution of the GGC repeat appeared to be almost identical in all four groups, the combination of CAG and GGC repeats was evaluated for correlation. The analysis revealed that the correlation between these two repeats was –0.003 by the Pearson coefficient and –0.117 by the Spearman coefficient which is independent of distribution.

In logistic regression, the number of CAG repeats may be used either as a continuous variable as suggested by its effect on transcriptional activity or as a nominal variable identifying only the different AR alleles. The results of both approaches are presented in Tables 1 and 2. In both cases, no significant association between the number of CAG repeats and the risk of prostate cancer was detected. This was also true when the different subgroups of the sporadic cases (age, Gleason score) were analysed separately (Table 3).

Thirty-nine patients and 22 non-affecteds were informative for CAG in 19 out of the 38 families. Fourteen

Figure 1 *Distribution of CAG repeat numbers in the four different groups of the study population. All four distributions are very similar and appear to be bimodal. The two peaks are separated by only one low frequency allele and therefore cannot formally be proven to be separate. Since this distribution is seen in all four subgroups of which three are really independent as well in one study on Caucasian Americans it is most probably real.*
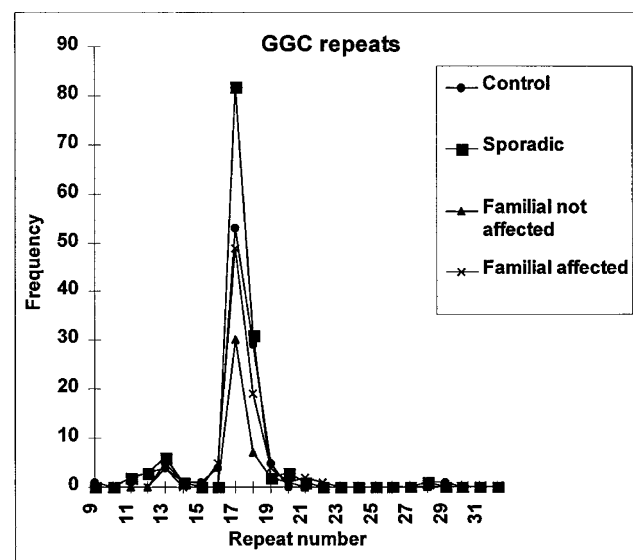
Figure 2 *Distribution of the GGC repeat numbers in the four different groups of our study population. There are no differences between the four groups and there is no hint to specific occurrence or lack of alleles which might be associated with prostate cancer.*

sib pairs out of 25 were concordant for the CAG-repeat, seven had inherited the shorter and seven the longer allele. Similarly, three out of nine sib trios were concordant, two carrying the longer allele. The expected values are 50% concordance rate for sib pairs and 25% for sib trios, with equal probability for both alleles. Coding the patients and healthy relatives according to the inheritance of the shorter or longer repeat did not reveal significant differences (Table 4).

Discussion

Prostate cancer is one of the most common malignancies in men. The minority of cases are familial and it has been shown that this familiarity can best be explained by autosomal inheritance of a predisposing gene.[13,15] Two predisposing genes have been localised on 1q, HPC1[16] and PCaP.[14] HPC1 has been detected mostly in American families but did not show linkage in the French–German collection of families in which PCaP was detected and estimated to be implicated in the disease of 30–50% of the families.[14] This may reflect a difference between the American and European population as that reported here for the *AR* gene. Beside these predisposing genes, a number of so called susceptibility genes has been considered to be involved in the occurrence of sporadic cases. These are the androgen receptor gene, the vitamin D receptor gene,

the 5α-steroid reductase type 2 gene, and a homeobox gene, *NKX3.1*. Some polymorphisms of these genes have indeed been found to be associated with an increased risk of prostate cancer.[4,6,9–11,17–20]

The association of the microsatellites in the first exon of the human androgen receptor gene, the CAG and the GGC repeat, with prostate cancer have been studied extensively.[4,6,9–11] The observed effects (odds ratio) were not large (usually < 2.5) and sometimes reached significance only in subgroups of the patients with higher staging of the tumour[4,6,9] or when the longest CAG repeats were compared with the smallest ones.[6]

Our results are based on a French–German collection of prostate cancer patients comprising sporadic cases, well selected controls, familial cases and their healthy male relatives. In contrast to the earlier studies, all comparisons (contingency tables) were far from any significance. In fact, the observed *P* values are almost all above 0.5 which makes a type 2 error (missing of a significant difference) unlikely in our collection, although it cannot be formally excluded.

We included almost all variants in our analysis which had been used in the earlier studies. For example, we used various subdivisions of the CAG alleles as used by Ingles *et al*[9] and Giovannucci *et al*.[6] A bimodal distribution of CAG alleles had been reported by Ingles *et al*[9] but had not been mentioned in later publications. We saw the same peaks at 21 and 23 repeats with a

Table 1 Odds ratios and 95% confidence intervals from logistic regression analyses for the CAG repeat in the androgen receptor gene (105 controls, 132 sporadic cases). A) repeat number treated as a continuous variable giving the incremental increase for adding one repeat. B) CAG repeats arranged into classes

Controls = constant n=*105*		*Odds ratio*	*CI 95% lower*	*CI 95% upper*	P-*value*
A) CAG continuous n=132	increment per repeat	1.007	0.92	1.10	0.87
B) CAG nominal	Patients/controls				
>=24	39/28	1	–	–	–
>=22 – <=23	30/22	1.02	0.49	2.13	0.95
>19 – <=21	34/35	1.43	0.73	2.82	0.29
<=19	29/20	0.96	0.45	2.03	0.91

Table 2 Odds ratios and 95% confidence intervals from logistic regression analyses for the GGC repeat in the androgen receptor gene (105 controls, 132 sporadic cases). A similar result is obtained when the cases are selected for age of onset or Gleason score or classes of allele length are analysed. The repeat number has been treated as a continuous variable giving the incremental increase for adding one repeat

Control = constant n=*105*	*Odds ratio*	*CI 95% Lower*	*CI 95% Upper*	P-*value*
increment per GGC repeat	1.004	0.89	1.13	0.95

short depression at 22 repeats in all subgroups of our patients. Therefore, we think this is not a chance observation and it lends itself to the formation of two allele classes. However, no difference in the risk of prostate cancer could be attributed to these classes in the samples studied here. Subdivision of the classes according to age of onset or Gleason score had no effect on our results (Table 3). Another important point may be that we did not use matched pairs as controls because they might be 'contaminated' by men who may

Table 3 Logistic regression analysis with the CAG repeat in the AR gene split into two classes (<22 and >=22) and tested for various subgroups of patients defined by age or Gleason score

Control = constant				
CAG nominal >=22 – <22	No. of patients	No. of controls	OR (CI 95%)	P-value
All sporadic cases	132	105	1.2 (0.7–2.0)	0.47
Age (yr)				
<=60	15	50	1 (Reference)	
>=22 <22	12	55	1.38 (0.59–3.22)	0.46
<65	24	50	1 (Reference)	
>=22 <22	23	55	1.15 (0.58–2.29)	0.69
<70	40	50	1 (Reference)	
>=22 <22	35	55	1.26 (0.69–2.28)	0.45
>=70	29	50	1 (Reference)	
>=22 <22	28	55	1.14 (0.60–2.17)	0.69
Gleason score				
2–5				
>=22	22	50	1 (Reference)	
<22	14	55	1.73 (0.80–3.74)	0.16
2–6				
>=22	32	50	1 (Reference)	
<22	27	55	1.30 (0.69–2.47)	0.41
4–10				
>=22	41	50	1 (Reference)	
<22	42	55	1.07 (0.60–1.91)	0.80
7–10				
>=22	16	50	1 (Reference)	
<22	19	55	1.93 (0.43–1.99)	0.84

Table 4 Inheritance of CAG alleles in prostate cancer families. Contingency table showing the frequency of affecteds and non-affecteds who inherited the shorter or longer allele

Maternal CAG allele	Affected	Non-affected	
short	18	10	x^2=0.003
long	21	12	P=0.958
Total	39	22	

later develop the disease. Instead, we used only old men (mean age above 70 years) who had been carefully watched for at least 7 years to exclude any signs of prostate cancer. If their distribution of CAG repeat differs from that of the general population this difference is expected to be contrary to that in prostate cancer cases and thus enhance statistical power. Furthermore, we separated strictly familial and sporadic cases because other genes could be responsible in the familial cases and thus dilute an association.

In the families, we looked for the segregation pattern of the maternal alleles in sibs in order to assess any distortion, if present. However, the segregation of the alleles was very close to chance expectations, and there was no indication of any influence of AR alleles in the occurrence of prostate cancer in the families. In fact, we did not expect such an influence since familial cases are either due to a predisposing gene or chance coincidence, both independent of the AR gene.

There is no correlation between the number of the triplets in both repeats in the sporadic cases of the French–German sample studied here like that found by Irvine *et al*[4] and Stanford *et al*.[11] This was also true for other subgroups and the whole sample. Therefore haplotypes cannot be used to indicate the risk of prostate cancer in the sample studied here.

All five studies[4,6,9–11] reporting an association between the number of CAG repeats and prostate cancer were carried out in the US. Both parameters linked in these studies, CAG repeat length and incidence of prostate cancer, are also known to vary according to ethnicity of the study population. African-Americans carry short CAG repeats and have a high risk of prostate cancer; men of Asian origin tend to have long CAG repeats and a low risk of prostate cancer, whereas Caucasians are intermediate with respect to both. We found a difference from these observations in the European population studied here in that there is no prostate cancer risk associated with the CAG repeat. We think the same would be true for Caucasian Americans when they emigrated from Europe. It is difficult to imagine that short alleles of these Europeans somehow acquired the risk after emigration. It seems more likely that the association seen in the US-American studies reflects the mixed population. This interpretation would also be compatible with the lack of significance in the Canadian study[12] in which sibs were compared to determine the influence of repeat lengths. The fact, that significant associations were observed in Caucasian and African-American

subgroups[4] may be due to admixtures which cannot be excluded without extended genealogical studies.

Our results cast some doubt on the assumption that the CAG repeats and/or the GGC repeats in the androgen receptor gene have direct influence on the *AR* gene as a susceptibility gene for prostate cancer. Despite of the effect of repeat length on transcriptional activity *in vitro*[8] it may be that repeat length only marks AR alleles associated with different risks. This association may be more systematic in the mixed populations of North America than in the French–German sample studied here. In consequence, there may also be differences in Europe, for example, between the Scandinavian and the Mediterranean populations. In addition to the search for the biological basis of risk differences between AR alleles it may be worthwhile analysing the association between CAG alleles and prostate cancer risk in different but well defined populations.

Acknowledgements

References

1 Wang C, Uchida T: Androgen receptor gene mutations in prostate cancer. *Nippon Hinyokika Gakkai Zasshi* 1997; **88**: 550–556.

2 Ross RK, Henderson BE: Do diet and androgens alter prostate cancer risk via a common etiological pathway? *J Natl Cancer Inst* 1994; **86**: 252–254.

3 Online Mendelian Inheritance in Man, OMIM (TM): John Hopkins University, Baltimore, MD. MIM Number: 313700: 1/27/1998: World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

4 Irvine RA, Yu MC, Ross RK, Coetzee GA: The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res* 1995; **55**: 1937–1940.

5 Lumbroso Rose, Beitel LK, Vasiliou DM, Trifiro MA, Pinsky L: Codon-usage variants in the polymorphic (GGN)n trinucleotide repeat of the human androgen receptor gene. *Hum Genet* 1997; **101**: 43–46.

6 Giovannucci E, Stampfer MJ, Krithivas K *et al*: The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc Natl Acad Sci USA* 1997; **94**: 3320–3323.

7 Edwards A, Hammond HA, Jun L, Caskey T, Chakraborty R: Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 1992; **12**: 241–253.

8 Kazemi-Esfarjani P, Trifiro MA, 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 neuropathies. *Hum Mol Genet* 1995; **4**: 523–527.

9 Ingles SA, Ross RK, Yu MC *et al*: Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. *J Natl Cancer Inst* 1997; **89**: 166–170.

10 Hardy DO, Scher HI, Bogenreider T *et al*: Androgen receptor CAG repeats lengths in prostate cancer: correlation with age of onset. *J Clin Endocrinol Metab* 1996; **81**: 4400–4405.

11 Stanford JL, Just JJ, Gibbs M *et al*: Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Res* 1997; **57**: 1194–1198.

12 Sun S, Narod SA, Aprikian A, Ghadirian P, Labrie F: Androgen receptor and familial prostate cancer. *Nat Med* 1995; **1**: 848–849.

13 Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC: Mendelian inheritance of familial prostate cancer. *Proc Natl Acad Sci USA* 1992; **89**: 3367–3371.

14 Berthon P, Valeri A, Cohen-Akenine A *et al*: Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. *Am J Hum Genet* 1998; **62**: 1416–1424.

15 Schaid D, McDonell SK, Blute ML, Thibodeau SN: Evidence for autosomal dominant inheritance of prostate cancer. *Am J Hum Genet* 1998; **62**: 1425–1438.

16 Smith JR, Freije D, Carpten JD *et al*: Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genomewide search. *Science* 1996; **274**: 1371–1374.

17 Taylor JA, Hirvonen A, Watson M, Pittman G, Mohler JL, Bell DA: Association of prostate cancer with vitamin D receptor gene polymorphism. *Cancer Res* 1996; **56**: 4108–4110.

18 Ingles SA, Coetzee GA, Ross RK *et al*: Association of prostate cancer with vitamin D receptor haplotypes in Africans-Americans. *Cancer Res* 1998; **58**: 1620–1623.

19 Reichardt JKV, Makridakis N, Henderson BE, Yu MC, Pike MC, Ross RK: Genetic variability of the human *SRD5A2* gene: implications for prostate cancer risk. *Cancer Res* 1995; **55**: 3973–3975.

20 Abbaszadegan MR, Brown KM, Voeller JH, Sciavolire P, Abate-Shen C, Gelmann EP: Polymorphic allele of NKX3.1 a prostate specific homeobox gene is a possible risk factor for prostate cancer. *Proc Am Assoc Cancer Res* 1998; **39**: 365.