Hum Genet (2007) 121:257–267
DOI 10.1007/s00439-006-0310-2

ORIGINAL INVESTIGATION

Segregation analysis of 1,546 prostate cancer families in Finland shows recessive inheritance

Sanna Pakkanen · Agnes B. Baffoe-Bonnie · Mika P. Matikainen · Pasi A. Koivisto · Teuvo L. J. Tammela · Snehal Deshmukh · Liang Ou · Joan E. Bailey-Wilson · Johanna Schleutker

Received: 7 August 2006 / Accepted: 25 November 2006 / Published online: 3 January 2007

Abstract Prostate cancer (PCa) is the most frequently diagnosed cancer in men worldwide and is likely to be caused by a number of genes with different modes of inheritance, population frequencies and penetrance. The objective of this study was to assess the familial aggregation of PCa in a sample of 1,546 nuclear families ascertained through an affected father and diagnosed during 1988–1993, from the unique, founder population-based resource of the Finnish Cancer Registry. Segregation analysis was performed for two cohorts of 557 early-onset and 989 late-onset families evaluating residual paternal effects and assuming that age at diagnosis followed a logistic distribution after log-transformation. The results did not support an autosomal dominant inheritance as has been reported in many of the hospital-based prostatectomy series. Instead, it confirmed the existence of hereditary PCa in the Finnish population under a complex model that included a major susceptibility locus with Mendelian recessive inheritance and a significant paternal regressive coefficient that is indicative of a polygenic/multifactorial component. The strengths of our study are the homogenous Finnish population, large epidemiological population-based data, histologically confirmed cancer diagnosis done before the PSA-era in Finland and registry based approach. Our results support the evidence that the inheritance of PCa is controlled by major genes and are in line with the previous linkage studies. Moreover, this is the first time a recessive inheritance is suggested to fit PCa in all data even when divided to early and late-onset cohorts.

Sanna Pakkanen and Agnes B Baffoe-Bonnie equally contributed to this work.

S. Pakkanen · P. A. Koivisto · J. Schleutker (✉)
Laboratory of Cancer Genetics,
Institute of Medical Technology,
University of Tampere and Tampere University Hospital,
Biokatu 8, 33014 Tampere, Finland
e-mail: Johanna.Schleutker@uta.fi

A. B. Baffoe-Bonnie · S. Deshmukh · L. Ou
Division of Population Science, Fox Chase Cancer Center,
Philadelphia, PA 19111, USA

A. B. Baffoe-Bonnie · J. E. Bailey-Wilson
National Human Genome Research Institute,
National Institutes of Health, Baltimore, MD 21131, USA

M. P. Matikainen · T. L. J. Tammela
Department of Urology,
Tampere University Hospital and Medical School,
University of Tampere, Tampere, Finland

Introduction

Prostate cancer (PCa) is the most frequently diagnosed cancer among men in the western world (World Health Organization 2003). In Finland, where the incidence of PCa has been rising in the last decade, it is estimated that in the year 2006, there will be 5,485 newly diagnosed PCa cases with the age-adjusted incidence rate of 115.4/100,000 inhabitants (Finnish Cancer Registry 2006).

Familial clustering of PCa was observed as early as the 1950s (Gianferrari et al. 1956), and in about 10% of all cases there is a clear positive family history of the disease. Carter et al. (1992) reported that for 40–50% of PCa cases, familial clustering was associated with multiple affected relatives, especially in families of

early-onset probands. Hereditary prostate cancer (HPC), which accounts for 5–10% of all PCa, is an etiologically complex disease with several genes implicated in determining risk (Grönberg et al. 1997). In a large Nordic consortium study of twins, Lichtenstein et al. (2000) reported an unusually high heritability of 42% for PCa. The clinical phenotype of PCa is complex and heterogeneous, and the arrival of the prostate specific antigen (PSA) era has further complicated the genetic analysis of PCa by allowing the early diagnosis of disease that might remain latent or clinically unimportant. The International Consortium for Prostate Cancer Genetics (ICPCG), which seeks to improve the mapping of PCa genes, has emphasized that one of the major difficulties in studying PCa is genetic heterogeneity, possibly due to multiple, incompletely penetrant PCa-susceptibility genes (Xu et al. (2005). Using parametric (dominant and recessive) and nonparametric analyses on 1,233 families, Xu et al. (2005) identified five distinct chromosomal regions with ''suggestive'' linkage (LOD score > 1.86) to PCa, namely 5q12, 8p21, 15q11, 17q21, and 22q12. Subsets of the analyzed group of families characterized by large numbers of early-onset (≤65 years) PCa, which are more likely to segregate highly penetrant mutations, provided stronger evidence of linkage in several regions (including the 22q12 locus, with a LOD score of 3.57). Additional PCa susceptibility loci reported to date (Schaid 2004) also include the three cloned genes: *HPC1/RNASEL, HPC2/ELAC2* and *MSR1* (Rebbeck et al. 2000; Tavtigian et al. 2001; Carpten et al. 2002; Xu et al. 2002).

The Finnish population of 5 million inhabitants represents a genetically isolated population with a unique gene pool useful for the study of genetic susceptibility to cancer and other complex diseases (de la Chapelle 1993; Peltonen 1997). Reliable population data are obtainable from various linked registries and the population-based Finnish Cancer Registry (FCR) covers virtually all histologically confirmed cancer diagnoses over almost 50 years. In addition, church and parish records enable the identification of familial relationships for individuals over several centuries. In Finland, *HPC1/RNASEL, HPC2/ELAC2* and *MSR1* loci explain only a small fraction of PCa cases (Rökman et al. 2001, 2002; Seppälä et al. 2003). Instead, three additional major susceptibility loci have been mapped in Finnish families including the *HPCX* (Xq27–28), 3p25–26 and 11q14 regions (Xu et al. 1998; Schleutker et al. 2003). Even so, a large proportion of Finnish HPC remains unexplained.

The purpose of this study was to assess the nature of familial aggregation of PCa in a sample of 1,546 Finnish nuclear families using regressive models as employed in complex segregation analysis. Segregation analysis is a statistical method for testing compatibility with Mendelian expectations by estimating the parameters of a given model of inheritance from family data. Previous segregation analyses in diverse populations have suggested that familial aggregation of PCa follows autosomal dominance, multifactorial, recessive or X-linked inheritance, but remain inconclusive. Six reports suggest a dominant inheritance mode (Carter et al. 1992; Grönberg et al. 1997; Schaid et al. 1998; Verhage et al. 2001; Conlon et al. 2003; Valeri et al. 2003). Cui (2001) reported a mixture of models including autosomal dominant inheritance in younger onset families with recessive or X-linked inheritance in older-onset families. A multifactorial model has been suggested by Gong et al. (2002). Families of Icelandic breast cancer probands with PCa-affected men yielded a codominant model (Baffoe-Bonnie et al. 2002). To account for the possibility of different modes of inheritance in families of early-onset probands (<61 years) versus late-onset families (≥61 years), we performed segregation analyses on these two separate cohorts and also analyzed the complete, combined dataset to determine the most parsimonious model for explanation of the familial aggregation of the disease in Finland.

Subjects and methods

Data sources

The nation-wide population based Finnish Cancer Registry (FCR) was founded in 1952 and reporting of cancer to the FCR was made obligatory in 1961. Currently physicians, hospitals and pathology laboratories send their reports to the registry independently. In addition, the FCR receives information from every death certificate in which cancer is mentioned, registering over 99% of all solid tumors diagnosed in Finland (Teppo et al. 1994). The FCR files can be linked to the registry of deaths and of immigrants issued by the Population Register Center in Finland. Population registration in Finland has traditions dating back to the sixteenth century and is considered to be of excellent quality. Since 1964 a centralized, nation-wide, computer-based population registry has been maintained by the National Population Registry Center and is based on unique personal identifiers, which are now used as main keys in every major person registry including the Finnish Cancer Registry.

Probands and relatives

We chose the pre-PSA time period between 1 January 1988 and 31 December 1993 and identified 9,142 men with newly diagnosed PCa nationwide from the Finnish Cancer Registry. Two non-overlapping cohorts were identified with 557 early-onset probands (diagnosed at <61 years of age) and 989 late-onset probands (diagnosed at ≥61 years of age). The cut-off of 61 years was selected so that the cohort of early-onset PCa would be informative, i.e., has enough cases. Although all cut-offs are arbitrary, ours is in line with the one used by previous PSA screening trial in Finland (Mäkinen et al. 2002). Details of the collection of population-based PCa families and the analyses for other cancers among first-degree relatives have been published elsewhere (Matikainen et al. 2001). Briefly, information on the birthplaces of probands was obtained from the Central Population Registry. The local registries (parishes and local authorities) of the communities where probands were born were contacted to obtain the names and birth dates of their parents, siblings, spouses and children. Family members were successfully traced for 94% of the probands, giving a total of 10,650 first-degree relatives out of the 11,427 identified. Descriptive statistics of these two non-overlapping cohorts and after combining them are shown in Table 1.

Segregation analysis

To test specifically for Mendelian inheritance of PCa in these Finnish pedigrees, maximum likelihood segregation analyses were performed on the age at diagnosis expressed as a censored trait using the REGTL module of the Statistical Analysis for Genetic Epidemiology program (SAGE 3.1. 1997). Under model 1 of this program, employing class A regressive models (Bonney 1986), the ''type'' or ''ousiotype'' (Cannings et al. 1978) influences age at diagnosis of PCa through the location and scale parameters of the logistic distribution, but does not influence susceptibility. Specifically, some constant proportion (γ) of the male population is assumed to be at a risk of PCa. The PCa phenotype is defined as a dichotomous variable (Y), where $Y = 1$ if affected and $Y = 0$ if unaffected (censored). Parameters estimated in the analysis include: $q_{A,}$ the frequency of the putative high-risk allele 'A', β_i baseline parameters, where i represents an individual's type (AA, AB, BB); α_i the age coefficients and γ_i the susceptibilities (Elston and George 1989). The logistic function describing the probability that an individual is affected by age ''a'' is given as $\gamma_i[1/(1 + e^{-\Phi}]$, where

$$\Phi = \beta_{\mathrm{i}} + \alpha_i(\mathrm{a}) + \delta_{\mathrm{F}}(Y_{\mathrm{F}}) \quad (1)$$

The coefficient δ_{F} reflects familial influence on risk corresponding to having an affected father.

Table 1 Descriptive statistics for prostate cancer cohorts in Finland

Description	Cohort-1	Cohort-2	Combined cohorts
Probands	557	989	1,546
Non-probands	3,631	7,019	10,650
Affected non-probands	51	109	160
Number of affected	608	1,098	1,706
Individuals in cohort	4,188	8,008	12,196
Number of males	2,418 (57.7%)	4,664 (58.2%)	7,082 (58.1%)
Number of females	1,770 (42.3%)	3,344 (46.9%)	5,114 (41.9%)
Number of fathers	390	673	1,063
Number of brothers	846	1,840	2,686
Number of sons	625	1,162	1,787
Number of mothers	388	681	1,069
Number of sisters	794	1,642	2,436
Number of daughters	588	1,021	1,609
Mean age of probands (years)	56.6 ± 3.4	74.4 ± 7.4	68.2 ± 10.7
Range of age at diagnosis of probands (years)	41.8 – 60.9	61 – 96	41.8 – 96
Mean age of affected non-probands (years)	69.6 ± 8.0	71.9 ± 7.8	71.2 ± 7.9
Range of age at diagnosis of affected non-probands (years)	48.3–85.5	48.1–88.7	48.1–88.7
Unaffected men aged ≥48 (years)	909 (18.3%)	2,141 (23.0%)	3,050 (21.4%)
Mean age of unaffected men (years)	65.2 ± 9.8	67.5 ± 12.3	66.6 ± 11.4
Range of ages of unaffected men (years)	48–85.5	48–104.3	48–104.3
Pedigree sizes (average, range)	8.9 (3–23)	9.4 (3–25)	9.2 (3–25)
Percent of pedigrees with ≥10 and ≥20 persons	28% (0.7%)	37.6% (0.5%)	34.2% (0.6%)

Positive values of δ_F mean that the individual with an affected father is more likely to have an earlier age at diagnosis, while negative values mean that the individual with an affected father is more likely to have a later age at diagnosis. Nonzero values of δ_F indicate the effects of polygenic and/or unmeasured shared-familial environmental risk factors on PCa risk.

Age at diagnosis for prostate cancer phenotype is assumed to follow a logistic distribution described by two parameters α and β, with the probability distribution function according to Elston and George (1989)

$$f(age) = [\alpha e^{\beta i+\alpha(age)}]/(1+e^{\beta i+\alpha(age)})^2 \quad (2)$$

This symmetric distribution is similar to a normal distribution and has a mean $-\beta/\alpha$, and variance, $\pi^2/3\alpha^2$, where π has a value of 3.1416. Based on the logistic distribution, the cumulative distribution function (CDF) is given by

$$F(age) = [\alpha e^{\beta i+\alpha(age)}]/[1+e^{\beta i+\alpha(age)}]. \quad (3)$$

The CDF represents the probability that a susceptible person will be affected by a given age. Age-specific penetrances were calculated for each genotype as

$$P(Y|\text{genotype } i,\ \text{age}) = [e^{\beta i+a(age)}]/[1+e^{\beta i+a(age)}]. \quad (4)$$

If the observed sex-specific ages at diagnosis do not follow a logistic distribution, this model may still be appropriate after transformation. A transformation equation equivalent to: aG1 × ln (age) was considered here, where aG1 is the geometric mean age at diagnosis for prostate cancer, computed from the observed ages at diagnosis among the 160 affected non-probands with PCa, 51 for the early-onset and 109 for the late-onset cohorts (Table 1).

Tests for genetic contribution to disease risk were implemented by postulating three types of individuals (AA, AB, BB) with three corresponding transmission parameters (τ_{AA}, τ_{AB}, τ_{BB}) describing the probability that a parent of a given type transmits the disease producing factor 'A' to his/her offspring (Elston and Stewart 1971; Elston and Yelverton 1975; Elston 1981). Under the hypothesis of genetic transmission, these τ parameters are constrained to the Mendelian values of $\tau_{AA} = 1.0$, $\tau_{AB} = 0.5$, $\tau_{BB} = 0.0$. Five sub-models of disease transmission were tested against a general model, where the transmission probabilities are estimated but with the restriction of homogeneity of trait distribution across generations to identify the best model for these data (Elston 1981). The ''no major gene'' model assumes that baseline risk is not influenced by ''type'' therefore all persons would come from a single distribution of age-specific risk for PCa. Single-locus Mendelian models assume that a major locus with two alleles should act in codominant, dominant or recessive fashion. The dominant and recessive models are special cases of the codominant model, where each genotype has a distinct age at diagnosis distribution. An environmental model with potentially distinct types of individuals was also tested, but here the transmission probability was held constant for all individuals.

We present results from the maximum likelihood segregation analyses performed on the log-transformed age at diagnosis of PCa expressed as a censored trait using the REGTL program (SAGE 3.1. 1997). Log-transformation of ages at diagnosis and ages at examination for all individuals with non-zero ages led to a final model that estimated genotypic baseline parameters (β_i) and age coefficient (α_i) and lifetime susceptibility for PCa for males along with the frequency (qA) of the high-risk allele A.

Hypothesis testing

The likelihood ratio test (LRT) was used to test each sub-model against the general model, and was computed as minus twice the natural log likelihood [–2ln(L)] of the general model subtracted from that for a restricted sub-model. This difference is asymptotically distributed as a χ^2 distribution with degrees of freedom equal to the difference in the number of independent parameters estimated in the two models.

Another method to compare models uses Akaike's information criteria (AIC), defined as: AIC = –2ln(L) + 2(number of parameters estimated). The most parsimonious model has the minimum AIC value (Akaike 1974). To correct for ascertainment bias, the likelihood of each pedigree was conditioned on the proband's affection status, using his age at diagnosis as recorded in the Cancer Registry (Cannings and Thompson 1977; Elston and Sobel 1979).

Results

Cohort-1 with 557 early-onset PCa families

As shown in Table 2, the no major gene model gave a very poor fit to the data in the early-onset PCa cohort and was thus rejected against the general unrestricted model in which all parameters were estimated based on the likelihood ratio test (LRT)

Table 2 Parameter estimates from segregation analysis of prostate cancer in 557 early-onset Finnish families ascertained through a single prostate cancer proband aged <61 years

Hypothesis	–2ln L	AIC	*df*	χ^2	*P*	Value of parameter									
						q_A	τ_{AA}	τ_{AB}	τ_{BB}	β_{AA}	β_{AB}	β_{BB}	α	γ	*F*(aff)
No major gene	479.02	487.02	6	18.36	0.005	[1.0]	–	–	–	–53.73 (5.31)	$=\beta_{AA}$	$=\beta_{AA}$	0.20 (0.014)	1.0 (0.0)	3.4 (0.25)
Dominant	473.08	485.08	4	12.42	0.014	0.9421 (0.597)	[1.0]	[0.5]	[0.0]	–62.27 (5.56)	$=\beta_{AA}$	–57.97 (4.32)	0.24 (0.019)	1.0 (0.0)	3.4 (0.24)
Codominant	474.18	488.18	3	13.52	0.036	0.0000007 (0.0)	[1.0]	[0.5]	[0.0]	–41.41 (3.67)	–63.40 (4.48)	–85.39 (7.01)	0.25 (0.017)	1.0 (0.0)	3.4 (0.33)
Recessive	462.20	474.20	4	1.54	0.8266	0.0536 (0.01)	[1.0]	[0.5]	[0.0]	–60.11 (4.11)	–64.63 (5.55)	$=\beta_{AB}$	0.25 (0.021)	1.0 (0.0)	3.7 (0.41)
Environmental	465.76	481.76	2	5.1	0.082	0.4183 (0.24)	1.0 (0.0)	$=\tau_{AA}$	$=\tau_{AA}$	–77.25 (7.01)	–79.69 (8.34)	–68.55 (6.98)	0.27 (0.036)	1.0 (0.0)	3.3 (0.31)
General	460.66	480.66	–	–	–	0.7328 (0.45)	1.0 (0.0)	0 (0.0)	0 (0.0)	–65.08 (6.93)	–67.98 (7.01)	–76.79 (7.86)	0.26 (0.019)	1.0 (0.0)	3.0 (0.30)

χ^2 is defined as (–2ln *L*) of the data under the hypothesis minus (–2ln *L*) of the data under the general model. Numbers in brackets are fixed at the indicated value; q_A the frequency of the putative high-risk allele, τ transmission parameter denoting the probability that a parent of a given type transmits the disease-producing factor A to his or her offspring, β baseline parameter, α age adjustment parameter, γ susceptibility parameter describing the cumulative probability of prostate cancer (assuming infinite lifespan)

(model 1 vs. model 6, $\chi^2 = 18.36$, $P < 0.005$ for 6*df*). The dominant ($P < 0.0014$), codominant ($P < 0.036$) and environmental ($P < 0.0082$) models were all rejected compared to the general model. The final general model reported is almost identical to the recessive Mendelian model (model 4 vs. model 6, $\chi^2 = 1.54$, $P < 0.83$ for 4*df*). The AIC, which takes into account the number of parameters estimated also, confirmed that the recessive model was the most parsimonious model. The estimated frequency ±SE (standard error) for the high-risk allele q_A was 0.054 (±0.01) for the recessive model. Figure 1 presents the predicted cumulative distribution function curves for log-transformed ages at diagnosis for the early-onset families under the recessive model. High-risk homozygous carriers of the putative risk allele AA have predicted age-specific cumulative probabilities greater than the heterozygous AB and BB non-carriers. The predicted mean age at diagnosis (i.e., 50% cumulative risk in Fig. 1) for the AA individuals is 60 and 64.6 years for the non-carriers. The susceptibility parameter γ was estimated at 1.0 for all male carriers of the risk allele, suggesting that 100% of the male population if they lived to infinity and did not die of competing causes, would express PCa if they were homozygous carriers of the allele A. Under this Mendelian recessive model, the cumulative probability that a male in Finland would be affected by PCa by age 70 was 0.92 for carriers and 0.79 for non-carriers, thus implying that if carriers and non-carriers did not die from competing causes, the estimated risk of being diagnosed with PCa at age 70 years for the homozygote carriers of the deleterious allele ($q = 0.054$), would be 2.7 per 1,000 among a hypothetical cohort of 100,000 men.

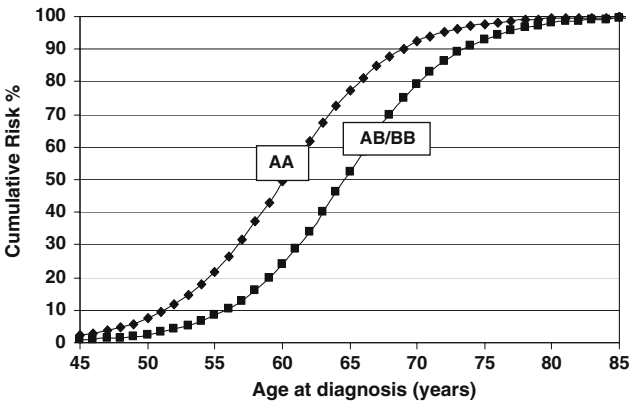

Fig. 1 Predicted cumulative risks for recessive *AA* (carriers) and *AB*/*BB* (non-carriers) with an affected father: 557 early-onset Finnish prostate cancer families

Cohort-2 with 989 late-onset PCa families

Table 3 shows the parameter estimates from the segregation analysis for the 989 families ascertained through late-onset probands diagnosed at ≥61 years of age. Compared to the unrestricted general model, the no major gene, the Mendelian dominant and environmental models did not fit the data and were rejected at $P < 0.001$. The Mendelian codominant model was also rejected by the LRT with a χ^2 of 12.64 and a P value of 0.006. The recessive model was the most parsimonious model according to the LRT (model 4 vs. model 6, $\chi^2 = 6.82$, $P < 0.15$ for $4df$), and it also had the lowest AIC = 1,331.10. Under this recessive model, inheritance of a putative high-risk allele A with an allele frequency (±SE) of 0.086 (±0.006) had predicted mean ages of onset of 65.6 years for men with the AA genotype and 72.2 years for AB/BB males, respectively. The lifetime risk of being diagnosed with PCa under this model was 5.0 per 1,000 among a hypothetical cohort of 100,000 men.

Figure 2 shows the predicted cumulative distribution function curves for log-transformed ages at diagnosis for this late-onset cohort in which the AA genotype has a distinctly different mean age at diagnosis of PCa.

Combined Cohort-1 and Cohort-2 with 1,546 PCa families

From Table 4, the combined cohorts with 1,546 probands gave parameter estimates very similar to those obtained from Cohort-2 with 989 probands. All other models except the Mendelian recessive model were significantly rejected when compared with the unrestricted general model. The recessive model was the most parsimonious model according to the LRT (model 4 vs. model 6, $\chi^2 = 8.78$, $P < 0.07$ for $4df$), and it also had the lowest AIC = 1,795.12. Under this recessive model, inheritance of a putative high-risk allele A with an allele frequency (±SE) of 0.0903 (±0.005) had predicted mean ages of onset of 63.6 years for men with the AA genotype and 71.0 years for AB/BB genotype males, respectively. Figure 3 shows that under the recessive model, the predicted cumulative risks for PCa are distinctly different for the AA compared to the AB/BB genotypes. The estimated mean age at diagnosis for the male homozygous carriers of the putative, high-risk allele A is 63.6 and it is 71.0 years for AB/BB genotype males. With a cumulative risk of 0.80 for homozygote carriers of the A allele at age 70 years, the estimated risk of being diagnosed with PCa in the absence of competing

Table 3 Parameter estimates from segregation analysis of prostate cancer in 989 late-onset Finnish families ascertained through a single prostate cancer proband

Hypothesis	−2ln L	AIC	*df*	χ^2	*P*	Value of parameter									
						q_A	τ_{AA}	τ_{AB}	τ_{BB}	β_{AA}	β_{AB}	β_{BB}	α	γ	*F*(aff)
No major gene	1371.08	1379.08	6	58.8	<0.001	[1]		[0.5]	[0]	−59.92 (4.16)	$=\beta_{AA}$	$=\beta_{AA}$	0.18 (0.013)	0.26 (0.06)	2.02 (0.18)
Dominant	1342.72	1354.72	4	30.44	<0.001	0.00148 (0.001)	[1]	[0.5]	[0]	−57.81 (4.53)	$=\beta_{AA}$	−62.53 (6.17)	0.18 (0.012)	1.0 (0.0)	3.18 (0.29)
Codominant	1324.92	1338.92	3	12.64	0.006	0.03137 (0.005)	[1]	[0.5]	[0]	−70.61 (6.88)	−75.26 (7.46)	−79.91 (8.01)	0.22 (0.016)	1.0 (0.0)	4.17 (0.51)
Recessive	1319.10	1331.10	4	6.82	0.15	0.0855 (0.006)	[1]	[0.5]	[0]	−64.71 (6.12)	−71.30 (6.95)	$=\beta_{AB}$	0.21 (0.015)	1.0 (0.0)	3.99 (0.63)
Environmental	1339.66	1355.66	2	27.38	<0.001	0.04486 (0.003)	0 (0.0)	$=\beta_{AA}$	$=\beta_{AA}$	−64.20 (5.01)	−62.32 (4.99)	−84.39 (8.15)	0.19 (0.012)	1.0 (0.0)	1.91 (0.09)
General	1312.28	1332.28	–	–	–	0.9974 (0.014)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	−66.93 (5.71)	−70.93 (6.34)	−72.21 (6.94)	0.21 (0.014)	0.44 (0.11)	2.85 (0.45)

χ^2 is defined as (−2ln *L*) of the data under the hypothesis minus (−2ln *L*) of the data under the general model. Numbers in brackets are fixed at the indicated value; q_A the frequency of the putative high-risk allele, τ transmission parameter denoting the probability that a parent of a given type transmits the disease-producing factor A to his or her offspring, β baseline parameter, α age adjustment parameter, γ susceptibility parameter describing the cumulative probability of prostate cancer (assuming infinite lifespan)

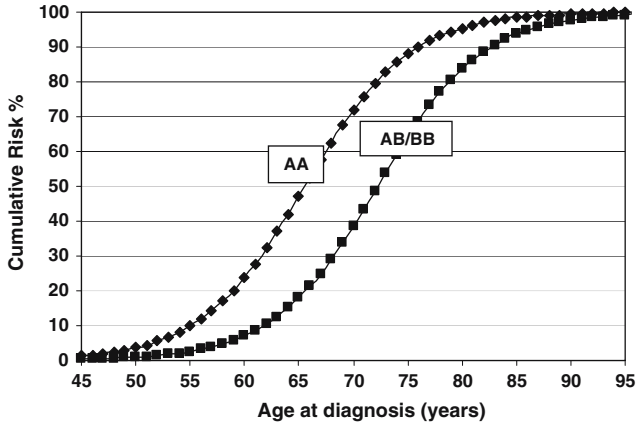

Fig. 2 Predicted cumulative risks for recessive *AA* (carriers) and *AB*/*BB* (non-carriers) of high-risk allele A with an affected father for 989 late-onset Finnish prostate cancer families

causes of death was 6.5 per 1,000 among a hypothetical cohort of 100,000 men.

All of these models (in each cohort and in the combined cohorts) included a residual paternal regressive coefficient, since inclusion of this coefficient significantly improved the fit of these models. The impact of genotype alone versus residual effect of having an affected father can be measured by computing the log odds of various combinations.

In effect, among men of the same age, born in the same cohort, and having the same affected father status, the log odds for being a homozygous carrier of the high risk allele A is computed as the difference between the genotypic baseline coefficients of the homozygous carriers and of the heterozygote and homozygous non-carriers. Using the parameters of the recessive model in the combined cohort as an example:

$$\begin{aligned}&(\beta_{\mathrm{AA}} + \alpha a' + \delta_F(\mathrm{YF})) - (\beta_{\mathrm{AB/BB}} + \alpha a' + \delta_F(Y_F))\\ &= \beta_{\mathrm{AA}} - \beta_{\mathrm{AB/BB}}\\ &= -63.34 - (-71.23) = 7.89.\end{aligned}$$

The odds of PCa in homozygous carriers of the A allele compared to the AB/BB non-carriers is 2,670, i.e., the exp (7.89). The log-odds due to an affected father between two individuals with the same genotype is 3.90 and the corresponding odds ratio is 49.40. The increase in log odds for a homozygote for the high-risk allele A with an affected father compared to the heterozygote carrier of the same age born in the same cohort but having an affected father would therefore be $(\beta_{\mathrm{AA}}-\beta_{\mathrm{AB/BB}}) + \delta_{\mathrm{F}}(Y_{\mathrm{F}})) = 7.89 + 3.90 = 11.79$ leading to high odds ratio.

Table 4 Parameter estimates from segregation analysis of prostate cancer in combined analysis of two cohorts of 1,546 Finnish families ascertained through a single prostate cancer proband

Hypothesis	−2ln L	AIC	*df*	χ^2	*P*	Value of parameter									
						q_A	τ_{AA}	τ_{AB}	τ_{BB}	β_{AA}	β_{AB}	β_{BB}	α	γ	F(aff)
1. No major gene	1884.38	1892.38	6	110.04	<0.001	[1]	[1]	[0.5]	[0]	−54.26 (5.21)	$=\beta_{AA}$	$=\beta_{AA}$	0.17 (0.014)	1.0 (0.0)	3.47 (0.35)
2. Dominant	1830.36	1842.36	4	56.02	<0.001	0.0022 (0.001)	[1]	[0.5]	[0]	−61.57 (5.98)	$=\beta_{AA}$	−65.77 (5.92)	0.21 (0.018)	1.0 (0.0)	3.50 (0.47)
3. Codominant	1806.42	1820.42	3	32.08	<0.001	0.1012 (0.007)	[1]	[0.5]	[0]	−70.16 (6.23)	−74.68 (7.26)	−86.40 (8.11)	0.23 (0.010)	1.0 (0.0)	3.20 (0.45)
4. Recessive	1783.12	1795.12	4	8.78	0.07	0.0903 (0.005)	[1]	[0.5]	[0]	−63.34 (6.01)	−71.23 (6.75)	$=\beta_{AB}$	0.22 (0.013)	1.0 (0.0)	3.90 (0.53)
5. Environmental	1866.34	1882.34	2	92.0	<0.001	0.4868 (0.061)	0 (0.0)	$=\beta_{AA}$	$=\beta_{AA}$	−55.74 (5.64)	−79.41 (8.04)	−56.10 (4.34)	0.18 (0.011)	1.0 (0.0)	3.32 (0.33)
6. General	1774.34	1794.34	–	–	–	0.9542 (0.004)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	−49.77 (4.7)	−53.49 (5.62)	−57.31 (5.89)	0.16 (0.014)	1.0 (0.0)	3.20 (0.29)

χ^2 is defined as (−2ln *L*) of the data under the hypothesis minus (−2ln *L*) of the data under the general model. Numbers in brackets are fixed at the indicated value; q_A the frequency of the putative high-risk allele, τ transmission parameter denoting the probability that a parent of a given type transmits the disease-producing factor A to his or her offspring, β baseline parameter, α age adjustment parameter, γ susceptibility parameter describing the cumulative probability of prostate cancer (assuming infinite lifespan)

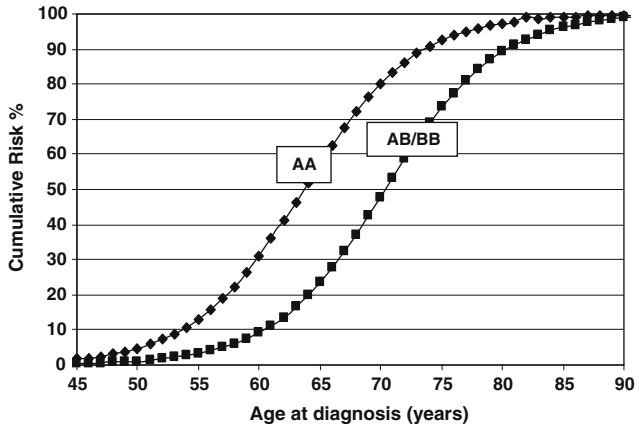

Fig. 3 Predicted cumulative risks for recessive *AA* (carriers) and *AB*/*BB* (non-carriers) of the putative high-risk allele A with an affected father for the 1,546 combined Finnish prostate cancer families

Discussion

In the present segregation analysis of our 1,546 population-based PCa families in Finland with ages at diagnosis of probands ranging between 41.8 and 96 years, a Mendelian recessive model with significant paternal regressive coefficient was shown to fit the homogeneous Finnish population best when the families were not separated into early- and late-onset cohorts. Recessive models with significant paternal regressive coefficients were also the most parsimonious models in both the 557 families of the early-onset cohort and in the 989 late-onset families. Under Hardy–Weinberg equilibrium, the estimated allele frequency of 0.09 for the combined cohort implies that 0.81% of the population in Finland would carry this rare putative high-risk allele. However, being a gender limited disease the susceptibility parameter γ of 1.0 obtained in the analysis suggested that 100% of the homozygous carrier male population at risk would develop PCa if they lived to infinity and did not die of competing causes.

Our results suggest that individuals carrying the risk allele get PCa at younger ages (<66 years) compared to non-carriers, whether they belong to the 557 early-onset or 989 late-onset groups or when the two cohorts are combined. Homozygous carriers of the risk allele (AA genotype) in the three groups above have a mean age at diagnosis of 60.1, 65.6 and 63.6 years, respectively, given that only about 6 years separate the ages at diagnosis of the homozygotes for the risk allele in the early-onset from the late-onset cohorts. Since the residual paternal effect was positive, those with affected fathers were at a higher risk for earlier onset PCa with polygenic and/or unmeasured shared-familial environmental risk factors compared to those with unaffected fathers. These findings are quite different from the previously reported evidence for the segregation of a rare autosomal dominant gene with high penetrance among different populations that included some series of prostatectomy patient families (Carter et al. 1992; Grönberg et al. 1997; Schaid et al. 1998; Verhage et al. 2001; Conlon et al. 2003; Valeri et al. 2003). Likewise a Swedish study by Grönberg et al. (1997) that was carried out on a population-based sample of 2,857 families selected through an affected father diagnosed with PCa in 1959–1963 and identified from the nationwide Cancer Registry, revealed that the observed clustering of PCa was best explained by a high risk allele inherited in a dominant mode, with a high population frequency (1.67%) and a moderate lifetime penetrance (63%). Similarly, the segregation analysis by Valeri et al. (2003) of families identified through 691 PCa patients recruited from three hospitals reported evidence for autosomal dominant gene inheritance of a high risk allele (frequency of 0.03%) with brother–brother dependence.

Our analyses suggest that there are likely to be multiple loci behind PCa and similar results have also been reported previously. A segregation study performed in Australia on 1,476 population-based pedigrees whose probands were diagnosed with PCa before the age of 70 in 1994–1997 suggested that a two-locus model fitted better than single-locus models and included a dominantly inherited risk that was greater at younger ages and a recessively inherited or X-linked increased risk which was greater at older ages (Cui et al. 2001). In a study of 1,719 first degree relatives in American and Canadian families by Gong et al. (2002), it was also observed that the good fit of the multifactorial model suggests that multiple genes, each having low penetrance, may be responsible for most inherited PCa susceptibility, and that the contribution of rare highly penetrant mutations might be small. In a segregation analysis of 389 Icelandic pedigrees that included both breast and prostate cancer, Baffoe-Bonnie et al. (2002), reported that the most parsimonious model was a Mendelian codominant model.

Previously, a recessive mode of inheritance has been reported in only a few studies. Cui et al. (2001) suggested that recessively inherited or X-linked inheritance increased risk at older ages, which was also seen in our linkage analyses of the HPCX locus (Xu et al. 1998; Schleutker et al. 2003). For an adult onset, sex-limited cancer such as PCa, recessive inheritance with incomplete penetrance and sporadic cases is consistent with X-linked PCa, which we previously mapped to the

Xq27–28 region using Finnish families characterized by ''no-male-to-male transmission'' (NMM). A follow-up linkage disequilibrium study utilizing familial/sporadic PCa cases and appropriate healthy controls identified an associated haplotype in the *HPCX* region (Baffoe-Bonnie et al. 2005). The results of this segregation analysis study are therefore consistent with the X-linked PCa transmission described previously (Schleutker et al. 2000). It has also been shown that there is a presence of residual brother–brother dependence (Valeri et al. 2003). In the study, by Narod et al. (1995) the prevalence of PCa was increased in those men with any first-degree relative affected. Most of the increase in relative risk was contributed by affected brothers, thus alluding to recessive or X-linked inheritance of the disease. In the study, by Monroe et al. (1995) an excess risk of PCa in men with affected brothers compared to those with affected fathers was also observed, consistent with the hypothesis of an X-linked, or recessive model of inheritance. Moreover, the prostate cancer risk was higher in probands' brothers than in probands' fathers in the Mayo Clinic Study (Schaid et al. 1998). These observations can be interpreted as evidence of recessive or X-linked effects in the risk of PCa.

Probands in the four American studies (Carter et al. 1992; Schaid et al. 1998; Verhage et al. 2001; Conlon et al. 2003) were part of a radical prostatectomy series for primary clinically localized PCa, and thus corresponded to a subgroup of patients not representative of all prostate cancer cases. This, as conceded by Schaid et al. (1998), could limit the power to assess heterogeneity of transmission across different age groups and represent a selection bias due to phenotypic characteristics. The particular strength of our study is the large population-based data composed of a homogenous Finnish population with registry-based approaches that provide unbiased information of malignancies in families. Prostate cancer was histologically confirmed in all cases in our families, and did not rely on PSA screening. The homogeneity of the Finnish population increases our chances of identifying loci, which may be less, pronounced in ethnically more diverse populations. Thus, linkage and association analyses of HPC conducted on Finnish families have found loci that are different from those reported in studies from other countries and populations (Schleutker et al. 2003; Seppälä et al. 2003a, b). Also, the present study was based on all prostate cancer cases in a population in a certain time window, not just known prostate cancer families, where genetic components may contribute to other cancer types and also be biased by specific family collection criteria.

The International Consortium for Prostate Cancer Genetics (ICPCG) recently announced that even though evidence of the existence of major PCa-susceptibility genes has been provided by multiple segregation analyses, genome-wide screens have not yielded conclusive chromosomal regions due to major difficulties that include genetic heterogeneity (Xu et al. 2005). The ICPCG employed parametric (dominant and recessive) and nonparametric analyses of 1,233 families world wide to identify several regions ''indicative'' of linkage. The main subsets of families likely to segregate highly penetrant mutations include families with large numbers of affected individuals or early age at diagnosis, leading to stronger evidence of linkage in several regions. Linkage and the association analysis of HPC conducted on Finnish families have found loci that are different from those reported in studies from other countries and populations (Schleutker et al. 2003; Seppälä et al. 2003a, b). We therefore believe that the results of these segregation analyses will be applied to defining a new model(s) for improving linkage analyses of the multiplex PCa families collected in Finland.

In conclusion, our findings suggest that the inheritance of PCa in the Finnish population is best explained by a Mendelian recessive model with a significant paternal regressive coefficient that is indicative of a polygenic multifactorial component. The rising incidence of PCa in Finland is possibly due to a combination of factors that include socio-cultural and lifestyle changes, environmental factors and the ongoing PSA screening (Finnish Cancer Registry, Mäkinen et al. 2003).

Acknowledgments The authors thank Riitta Vaalavuo for her excellent assistance. Dr. Erica Golemis of the Fox Chase Cancer Center provided beneficial comments on this manuscript. This study was financially supported by the Medical Research Fund of Tampere University Hospital, the Reino Lahtikari Foundation, the Finnish Cancer Organizations, the Sigrid Juselius Foundation and the Academy of Finland (grant no. 211123). The program package SAGE supported by the US Public Health Service Resource Grant RR03655 from Division of Research Resources, was used for the complex segregation analyses. This work was partially supported by the Intramural Research Program of the National Human Genome Research Institute, the National Institutes of Health (Contract Number N01-HG-55389). A.B.B-B, S.D. and L.O also received support from USPHS grant CA-06927 and an appropriation from the Commonwealth of Pennsylvania.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans 19:716–723

Baffoe-Bonnie AB, Kiemeney LA, Beaty TH, Bailey-Wilson JE, Schnell AH, Sigvaldsson H, Olafsdottir G, Tryggvadottir L,

Tulinius H (2002) Segregation analysis of 389 Icelandic pedigrees with breast and prostate cancer. Genet Epidemiol 23:349–363

Baffoe-Bonnie AB, Smith JR, Stephan DA, Schleutker J, Carpten JD, Kainu T, Gillanders EM, Matikainen M, Teslovich TM, Tammela T, Sood R, Balshem AM, Scarborough SD, Xu J, Isaacs WB, Trent JM, Kallioniemi OP, Bailey-Wilson E (2005). A major locus for hereditary prostate cancer in Finland: localization by linkage disequilibrium of a haplotype in the HPCX region. Hum Genet 117(4):307–316

Bonney GE (1986) Regressive logistic models for familial disease and other binary traits. Biometrics 42:611–625

Cannings C, Thompson EA (1977) Ascertainment in the sequential sampling of pedigrees. Clin Genet 12:208–212

Cannings C, Thompson E, Skolnick M (1978) Probability functions on complex pedigrees. Adv Appl Probl 10:26–26–61

Carpten J, Nupponen N, Isaacs S, Sood R, Robbins C, Xu J, Faruque M et al (2002) Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat Genet 30:181–184

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

Conlon EM, Goode EL, Gibbs M, Stanford JL, Badzioch M, Janer M, Kolb S, Hood L, Ostrander EA, Jarvik GP, Wijsman EM (2003) Oligogenic segregation analysis of hereditary prostate cancer pedigrees: Evidence for multiple loci affecting age at onset. Int J Cancer 105:630–635

Cui J, Staples MP, Hopper JL, English DR, McCredie MR, Giles GG (2001) Segregation analyses of 1,476 population-based Australian families affected by prostate cancer. Am J Hum Genet 68:1207–1218

De la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857–865

Elston RC (1981) Segregation analysis. Adv Hum Genet 11:63–120, 372–373

Elston RC, George VT (1989) Age of onset, age at examination, and other covariates in the analysis of family data. Genet Epidemiol 6:217–220

Elston RC, Sobel E (1979) Sampling considerations in the gathering and analysis of pedigree data. Am J Hum Genet 31:62–69

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Elston RC, Yelverton KC (1975) General models for segregation analysis. Am J Hum Genet 27:31–45

Finnish Cancer Registry (2006) Cancer incidence in Finland 1995 and 2004. Cancer statistics at http://www.cancerregistry.fi last updated on 7 June 2006

Gianferrari L, Arrigoni G, Cresseri A, Lovati G, Morganti G (1956) Genetic and clinico-statistical research on neoplasms of the prostate. Acta Gerontol Milano 5:224–233

Gong G, Oakley-Girvan I, Wu AH, Kolonel LN, John EM, West DW, Felberg A, Gallagher RP, Whittemore AS (2002) Segregation analysis of prostate cancer in 1,719 white, African–American and Asian–American families in the United States and Canada. Cancer Causes Control 13:471–482

Grönberg H, Damber L, Damber JE, Iselius L (1997) Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. Am J Epidemiol 146:552–557

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343:78–85

Mäkinen T, Tammela TLJ, Stenman U-H, Määttänen L, Rannikko S, Aro J, Juusela H, Hakama M, Auvinen A (2002). Family history and prostate cancer screening with prostate-specific antigen. J Clin Oncol 20(11):2658–2663

Mäkinen T, Tammela TLJ, Hakama M, Stenman U-H, Rannikko S, Aro J, Juusela H, Määttänen L, Auvinen A (2003) Tumor characteristics in a population-based prostate cancer screening trial with prostate-specific antigen. Clin Cancer Res 9:2435–2439

Matikainen MP, Pukkala E, Schleutker J, Tammela TL, Koivisto P, Sankila R, Kallioniemi OP (2001) Relatives of prostate cancer patients have an increased risk of prostate and stomach cancers: a population-based, cancer registry study in Finland. Cancer Causes Control 12:223–230

Monroe KR, Yu MC, Kolonel LN, Coetzee GA, Wilkens LR, Ross RK, Henderson BE (1995) Evidence of an X-linked or recessive genetic component to prostate cancer risk. Nat Med 1:827–829

Narod SA, Dupont A, Cusan L, Diamond P, Gomez JL, Suburu R, Labrie F (1995) The impact of family history on early detection of prostate cancer. Nat Med 1:99–101

Peltonen L (1997) Molecular background of the Finnish disease heritage. Ann Med 29:553–556

Rebbeck TR, Walker AH, Zeigler-Johnson C, Weisburg S, Martin AM, Nathanson KL, Wein AJ, Malkowicz SB (2000) Association of HPC2/ELAC2 genotypes and prostate cancer. Am J Hum Genet 67:1014–1019

Rökman A, Ikonen T, Mononen N, Autio V, Matikainen M.P, Koivisto PA, Tammela TL, Kallioniemi OP, Schleutker J (2001) ELAC2/HPC2 involvement in hereditary and sporadic prostate caner. Cancer Res 61:6038–6041

Rökman A, Ikonen T, Seppälä EH, Nupponen N, Autio V, Mononen N, Bailey-Wilson J, Trent J, Carpten J, Matikainen MP, Koivisto PA, Tammela TL, Kallioniemi OP, Schleutker J (2002) Germline alterations of the RNASEL gene, a candidate HPC1 gene at 1q25, in patients and families with prostate cancer. Am J Hum Genet 70 (5):1299–1304

SAGE 3.1. (1997) Statistical analysis for genetic epidemiology, release 3.1. Computer program package available from the Department of Epidemiology and Biostatistics, Rammelcamp Center for Education and Research, MetroHealth campus, Case Western Reserve University, Cleveland

Schaid DJ (2004) The complex genetic epidemiology of prostate cancer. Hum Mol Genet 13 Spec no. 1, ppR103–R121

Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN (1998) Evidence for autosomal dominant inheritance of prostate cancer. Am J Hum Genet 62:1425–1438

Schleutker J, Matikainen M, Smith J, Koivisto P, Baffoe-Bonnie A, Kainu T, Gillanders E, Sankila R, Pukkala E, Carpten J, Stephan D, Tammela T, Brownstein M, Bailey-Wilson J, Trent J, Kallioniemi OP (2000) A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: frequent HPCX linkage in families with late-onset disease. Clin Cancer Res 6(12):4810–4815

Schleutker J, Baffoe-Bonnie AB, Gillanders E, Kainu T, Jones MP, Freas-Lutz D, Markey C, Gildea D, Riedesel E, Albertus J, Gibbs KD Jr, Matikainen M, Koivisto PA, Tammela T, Bailey-Wilson JE, Trent JM, Kallioniemi OP (2003) Genome-wide scan for linkage in Finnish hereditary prostate cancer (HPC) families identifies novel susceptibility loci at 11q14 and 3p25–26. Prostate 57:280–289

Seppälä EH, Ikonen T, Autio V, Rökman A, Mononen N, Matikainen MP, Tammela TL, Schleutker J (2003a) Germ-

line alterations in MSR1 gene and prostate cancer risk. Clin Cancer Res 9:5252–5256

Seppälä EH, Ikonen T, Mononen N, Autio V, Rökman A, Matikainen MP, Tammela TL, Schleutker J (2003b) CHEK2 variants associate with hereditary prostate cancer. Br J Cancer 89:1966–1970

Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD, Brownstein MJ, Bova GS, Guo H, Bujnovszky P, Nusskern DR, Damber JE, Bergh A, Emanuelsson M, Kallioniemi OP, Walker-Daniels J, Bailey-Wilson JE, Beaty TH, Meyers DA, Walsh PC, Collins FS, Trent JM, Isaacs WB (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371–1374

Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ et al (2001) A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 27:172–180

Teppo L, Pukkala E, Lehtonen M (1994) Data quality and quality control of a population-based cancer registry, experience in Finland. Acta Oncol 33:365–369

Valeri A, Briollais L, Azzouzi R, Fournier G, Mangin P, Berthon P, Cussenot O, Demenais F (2003) Segregation analysis of prostate cancer in France: evidence for autosomal dominant inheritance and residual brother–brother dependence. Ann Hum Genet 67:125–137

Verhage BA, Baffoe-Bonnie AB, Baglietto L, Smith DS, Bailey-Wilson JE, Beaty TH, Catalona WJ, Kiemeney LA (2001) Autosomal dominant inheritance of prostate cancer: a confirmatory study. Urology 57:97–101

World Health Organization (2003) Global cancer rates could increase by 50% to 15 million by 2020 http://www.who.int/mediacentre/news/releases/2003/pr27/en/

Xu J, Meyers D, Freije D, Isaacs S, Wiley K, Nusskern D, Ewing C, Wilkens E, Bujnovszky P, Bova GS, Walsh P, Isaacs W, Schleutker J, Matikainen M, Tammela T, Visakorpi T, Kallioniemi OP, Berry R, Schaid D, French A, McDonnell S, Schroeder J, Blute M, Thibodeau S, Trent J (1998) Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20:175–179

Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Hu JJ, Sterling D et al (2002) Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. Nat Genet 32:321–325

Xu J, Dimitrov L, Chang BL, Adams TS, Turner AR, Meyers DA, Eeles RA et al (2005) A Combined genome-wide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the International Consortium for Prostate Cancer Genetics. Am J Hum Genet 77:219–229