Combined Genome-Wide Scan for Prostate Cancer Susceptibility Genes

Elizabeth M. Gillanders, Jianfeng Xu, Bao-li Chang, Ethan M. Lange, Fredrik Wiklund, Joan E. Bailey-Wilson, Agnes Baffoe-Bonnie, MaryPat Jones, Derek Gildea, Erica Riedesel, Julie Albertus, Sarah D. Isaacs, Kathleen E. Wiley, Caroline E. Mohai, Mika P. Matikainen, Teuvo L. J. Tammela, S. Lilly Zheng, W. Mark Brown, Annika Rökman, John D. Carpten, Deborah A. Meyers, Patrick C. Walsh, Johanna Schleutker, Henrik Gronberg, Kathleen A. Cooney, William B. Isaacs, Jeffrey M. Trent

Background: **Prostate cancer represents a substantial public health burden worldwide. It is the second leading cause of cancer death among men in the United States. A family history of the disease is among the most well-established risk factors for prostate cancer. Efforts to localize prostate cancer susceptibility alleles by using genetic linkage analysis methods have been hindered by genetic heterogeneity, incomplete penetrance, disease phenocopies, and the lack of DNA samples from parents of individuals with late-onset prostate cancer.** ***Methods:*** **We performed a combined genome-wide linkage analysis among 426 families from four existing hereditary prostate cancer (HPC) study populations to systematically search for prostate cancer susceptibility genes. To decrease the degree of locus heterogeneity, we analyzed subsets of families with similar clinical and demographic characteristics. Nonparametric multipoint linkage was the primary method of analysis. Results are presented as allele-sharing logarithm of the odds (LOD) scores, and all reported *P* values are two-sided.** ***Results:*** **The strongest evidence for prostate cancer linkage was found at chromosome region 17q22 (nonparametric multipoint Kong and Cox allele-sharing LOD score = 3.16 at marker D17S787; *P* = .00007). Stratified analyses revealed several additional chromosomal regions that are likely to segregate prostate cancer susceptibility genes among specific subsets of HPC families, including 15q11 among families with late-onset disease (allele-sharing LOD = 5.57 at marker D15S128; *P*<.00001) and 4q35 among families with four or more affected family members (allele-sharing LOD = 3.10 at marker D4S1615; *P* = .00008).** ***Conclusion:*** **Fine mapping studies to facilitate identification of prostate cancer susceptibility genes in these linked regions are warranted. [J Natl Cancer Inst 2004;96: 1240–7]**

The genetic predisposition to prostate cancer is well established *(1)* and possibly the strongest among all common cancers *(2)*. Evidence for major prostate cancer susceptibility genes that segregate in families has been obtained from several complex segregation analyses, with the majority supporting a dominant mode of inheritance *(2–7)* and the remainder supporting recessive or X-linked modes of inheritance *(8,9)*. At least five candidate prostate cancer susceptibility genes have been reported thus far, including ELAC2/HPC2 *(10)*, RNASEL *(11)*, SR-A/MSR1 *(12)*, CHEK2 *(13)*, and BRCA2 *(14)*. However, these genes, if confirmed, likely account for only a small fraction of the observed genetic predisposition to prostate cancer.

The use of genetic linkage studies to identify disease susceptibility genes in families with multiple affected members has proven to be a fruitful approach for those rare diseases that show Mendelian inheritance. This approach is most effective when locus heterogeneity and phenocopies of disease (i.e., disease caused by environmental risk factors) are low and when the sample sizes (i.e., the number of families) are large. However, for prostate cancer, there is evidence for a substantial number of susceptibility loci as well as a high rate of sporadic disease in the general population. In addition, most individual prostate cancer linkage studies are based on a relatively small number of families, typically fewer than 100 extended pedigrees. This combination of factors makes it particularly difficult to successfully apply genetic linkage methods to this disease. The current inconclusive status of prostate cancer linkage studies reflects this complexity: multiple prostate cancer loci have been reported for different sets of families, but none of the loci has been consistently observed in replication studies. One effective approach to overcome the problems of small sample size, locus heterogeneity, and disease phenocopies is to combine data from several different study populations. This approach not only increases the sample size but also could potentially decrease the degree of locus heterogeneity by enlarging the number of families that have similar clinical and demographic characteristics. We combined four existing hereditary prostate cancer (HPC) study populations to perform a genome-wide linkage analysis to systematically search for prostate cancer susceptibility genes. With a total of 426 HPC families, this combined analysis is the largest

Affiliations of authors: National Human Genome Research Institute, National Institutes of Health, Bethesda, MD (EMG, JEBW, ABB, MPJ, DG, ER, JA); Center for Human Genomics (JX, BC, LZ, DAM) and Department of Public Health (EML, WMB), Wake Forest University School of Medicine, Winston-Salem, NC; Department of Radiation Sciences, Oncology, University of Umeå, Umeå, Sweden (FW, HG); Division of Population Science, Fox Chase Cancer Center, Philadelphia, PA (ABB); Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD (SDI, KEW, PCW, WBI); Departments of Internal Medicine (CEM) and Urology (KAC), University of Michigan, Ann Arbor, MI; Department of Urology (MPM, TLJT) and Institute of Medical Technology (AR, JS), University of Tampere and Tampere University Hospital, Tampere, Finland; Translational Genomics Research Institute (TGen), Phoenix, AZ (JD, JMT).

Correspondence to: Jianfeng Xu, MD, PhD, Medical Center Blvd., Winston-Salem, NC 27157 (e-mail: jxu@wfubmc.edu) or Jeffrey M. Trent, PhD, TGen, 400 N. Fifth St., Suite 1600, Phoenix, AZ 85004 (e-mail: jtrent@tgen.org).

See "Notes" following "References."

DOI: 10.1093/jnci/djh228

Journal of the National Cancer Institute, Vol. 96, No. 16,

genome-wide scan for prostate cancer susceptibility genes to date.

Subjects and Methods

Study Populations

This study included four HPC family study populations and a total number of 426 families. Written informed consent was obtained from all participants, and study protocols were reviewed and approved by the institutional review boards at each institution.

Study population 1 consisted of 188 HPC families recruited at the Brady Urology Institute at Johns Hopkins Hospital (Baltimore, MD). A detailed description of this study population has been presented elsewhere *(15,16)*. Each of these families had at least three first-degree relatives affected with prostate cancer. The majority of the families (≈65%) were ascertained through referrals generated as a response to a letter by P. C. Walsh to 8000 urologists throughout the country. Approximately 23% of the families were identified through family history records for patients treated for prostate cancer at Johns Hopkins Hospital. The remaining families (≈12%) were ascertained through individuals who contacted us in response to articles describing our prostate cancer family studies that were published in a variety of lay publications. Prostate cancer diagnosis was verified by reviewing the medical records for each affected male studied. Age at diagnosis of prostate cancer was confirmed either through medical records or through two other independent sources (such as reports from family members). DNA was extracted from whole blood that was obtained from living affected individuals as well as unaffected family members who contributed information to linkage analyses. The mean within-family age at prostate cancer diagnosis (i.e., the mean age at diagnosis for all prostate cancer patients within one family) for this study population was 64.4 years (range = 43.0–76.0 years).

Study population 2 consisted of 175 HPC families recruited at the University of Michigan (Ann Arbor). A detailed description of this study population has been presented elsewhere *(17)*. Briefly, since 1995, the Prostate Cancer Genetics Project (PCGP), a family-based study designed to examine the molecular basis for inherited prostate cancer susceptibility, has recruited men with early-onset prostate cancer and/or a family history of prostate cancer. The PCGP currently has more than 2500 participants from more than 900 families. Most of the families were recruited directly from the University of Michigan Comprehensive Cancer Center; other sources included direct patient or/physician referrals. Prostate cancer diagnoses were routinely verified by review of medical records; occasionally, when medical records were not available, confirmation of the diagnosis was achieved by independent reports from two family members. All of the families from this population included in this study had three or more confirmed cases of prostate cancer or two members who were diagnosed with the disease at or before age 55 years. DNA was extracted from whole blood that was obtained from living affected PCGP participants as well as from unaffected family members who contributed information to linkage analyses. The mean within-family age at prostate cancer diagnosis for this study population was 64.0 years (range = 49.5–78.2 years).

Study population 3 consisted of 50 HPC families recruited at the University of Umeå (Umeå, Sweden). A detailed description of this study population has been presented elsewhere *(18)*. Briefly, since 1995, the Department of Oncology at the University of Umeå, in collaboration with urologists and oncologists throughout Sweden, has recruited families with hereditary prostate cancer. All prostate cancer diagnoses were confirmed through the Swedish National Cancer Registry and by examination of medical records. The majority of patients (79%) had clinical symptoms at diagnosis, and 60% were diagnosed with locally advanced or metastatic disease. Blood for DNA extraction was collected from affected men and their spouses and children so that genotypes could be determined. Each of the 50 families from this population had at least three male members with prostate cancer, and at least two of the affected men had a known or inferred genotype. The mean within-family age at prostate cancer diagnosis for this study population was 69.1 years (range = 56.0–77.5 years).

Study population 4 consisted of 13 HPC families recruited at the University of Tampere and Tampere University Hospital (Tampere, Finland). A detailed description of this study population has been presented elsewhere *(19)*. Briefly, these 13 multiplex families were selected from a total of 57 linkage families previously used for HPC1 and HPCX analyses *(20,21)* solely on the basis of their informativeness for linkage analyses (i.e., they represented the largest number of cases and best-sampled families). None of these families exhibited bilineality of prostate cancer history. Prostate cancer diagnoses were confirmed using the Finnish Cancer Registry or individual patient records obtained from regional hospitals. All living affected cases, as well as the spouse and adult-aged offspring, were contacted to obtain informed consent and to request a blood sample for linkage analyses. The mean within-family age at prostate cancer diagnosis for this study population was 68.6 years (range = 56.5–77.0 years).

Marker Genotyping

Standard techniques were used to isolate genomic DNA from blood samples. All DNA samples were genotyped in a single laboratory at the National Human Genome Research Institute (Cancer Genetics Branch Laboratory, Bethesda, MD) by using 406 short tandem-repeat markers that had an average intermarker spacing of approximately 10 cM and an average heterozygosity of 80%. Polymerase chain reaction (PCR) assays, using fluorescently labeled primers, were carried out using a Genesis 200 robot (Tecan, Research Triangle Park, NC). All PCR assays were done in a 15-μL volume that contained 20 ng of genomic DNA, each of the primers at 0.33 m*M*, each deoxynucleotide triphosphate at 0.25 m*M*, 2.5 m*M* $MgCl_2$, 10 m*M* Tris–HCl, 50 m*M* KCl, and 0.5 U Taq polymerase. PCR amplification was performed using GeneAmp thermocyclers (models 9600 and 9700; Applied Biosystems, Foster City, CA) and the following conditions: 95 °C for 12 minutes; 10 cycles of 94 °C for 15 seconds, 55 °C for 15 seconds, and 72 °C for 30 seconds; 20 cycles of 89 °C for 15 seconds, 55 °C for 15 seconds, and 72 °C for 30 seconds; and a final extension at 72 °C for 10 minutes. We combined

5–15 μL of each PCR product (depending on the yield) with the PCR products of up to 20 other markers of appropriate size and fluorescent label. The combined PCR products were separated using DNA sequencers (model 377 or 3100; Applied Biosystems) that allow multiple fluorescently labeled markers to be run in a single lane. ROX 500 or 400 (Applied Biosystems) was run as an internal size standard. Allele size was calculated with the local southern algorithm available in GeneScan Analysis software (Applied Biosystems). Allele calling and binning was done using Genotyper software (Applied Biosystems). All genotyping included a Centre d'Etude du Polymorphisme Humain reference individual (specifically, family number 1347, individual number 02) for quality control purposes. We evaluated the genotyping error rate by including 1% of samples as blinded duplicates. We detected 11 genotyping errors among the 12 992 duplicated genotypes (0.085% error rate).

Statistical Analyses

Because of uncertainty about the mode(s) of inheritance of prostate cancer, we used nonparametric multipoint linkage analyses as the primary method of analysis, although parametric multipoint analyses were also performed. We used the Merlin computer program, version 0.9.8 *(22)*, to analyze linkage among affected pairs of relatives. We compared the estimated marker identity-by-descent (IBD) allele sharing among various affected pairs of relatives with the values expected under the null hypothesis of no linkage. The scoring function *all* was used as the primary choice for measuring IBD sharing; we also used the scoring function *pairs* in some analyses, as previously described *(23)*. Allele-sharing logarithm of the odds (LOD) scores were then calculated by using equal weights for all families as described by Kong and Cox *(24)*. *P* values for LOD scores were calculated by assuming that the statistic was normally distributed and were not adjusted for multiple tests. Marker allele frequencies were estimated from pedigree founders and were calculated for each of the four study populations. Haldane's mapping function was used in the multipoint analyses. For the parametric analyses, we used the autosomal dominant model described by Smith et al. *(25)*. In the recessive model, the disease allele frequency was assumed to be 0.15 and the penetrances for men with risk and non-risk genotypes were assumed to be 1.0 and 0.001, respectively. These assumptions were based on segregation analysis results of Schaid et al. *(4)*. A maximum likelihood approach was used to estimate the proportion of linked families (α), by maximizing the heterogeneity LOD score (HLOD), as implemented in the computer program GENEHUNTER, version 2.0 *(26)*. All statistical tests were two-sided.

Results

A total of 426 HPC families were included in this combined analysis. Table 1 shows the numbers of families from each study population grouped according to the mean age at prostate cancer diagnosis, the number of affected family members, ethnicity, and whether there was evidence of male-to-male disease transmission. Each of these subsets contained a large number of HPC families. For example, 285 HPC families had at least four men affected with prostate cancer and 201 HPC families had a mean age at diagnosis of less than 65 years.

Table 1. Characteristics of families in the combined analysis

Characteristic	Study population No. of families (No. of genotyped individuals)* 1 188 (1033)	2 175 (640)	3 50 (190)	4 13 (87)	Total 426 (1950)
Mean age at diagnosis, y					
<65	96	91	10	4	201
≥65	92	84	40	9	225
Number of affected family members					
≤3	28	96	13	4	141
4	47	45	17	5	114
≥5	113	34	20	4	171
Race/ethnicity					
White	154	157	50	13	374
Black	15	16	0	0	31
Jewish	17	0†	0†	0†	17
Other	2	2	0	0	4
Male-to-male transmission					
Yes	123	126	37	6	292
No	65	49	13	7	134

*Study population 1: Johns Hopkins Hospital (Baltimore, MD); study population 2: University of Michigan (Ann Arbor, MI); study population 3: University of Umeå (Umeå, Sweden); study population 4: University of Tampere and Tampere University Hospital (Tampere, Finland).

†Information about Jewish heritage was not specifically requested in the study.

Results of nonparametric multipoint linkage analyses for prostate cancer susceptibility genes among the 426 HPC families are plotted in Fig. 1. The strongest evidence for prostate cancer linkage was found at chromosome region 17q22, which gave a nonparametric multipoint Kong and Cox allele-sharing LOD score (henceforth referred to as a LOD score) of 3.16 at marker D17S787 ($P = .00007$) when the IBD scoring function *all* was used (Fig. 2). Slightly stronger evidence for linkage at this marker was obtained when we used the IBD scoring function *pairs* (LOD score = 3.40; $P = .00004$). The statistically significant linkage at this region could be partially influenced by marker allele frequencies because parental genotyping information was incomplete in most of the HPC families. However, the influence of the marker allele frequencies appears to be limited because the LOD scores in this chromosomal region continued to suggest a linkage when we used conservative marker allele frequency estimates that were based on all family members within each study; from this analysis, the LOD score at D17S787 was 2.32 ($P = .0005$). One hundred ninety-seven HPC families provided positive LOD scores at marker D17S787. Among them, 62 HPC families had LOD scores greater than 0.30, and 15 families had LOD scores greater than 0.5.

To evaluate the statistical significance of the linkage at 17q22, we used the simulation function of the computer program Merlin to randomly simulate 200 replicates of the genome-wide scan data, assuming no linkage. Each of these replicates consisted of 406 markers that were based on the exact pedigree structure (including missing data patterns) of our 426 HPC families and the marker information reported in our study. We then analyzed each replicate using nonparametric multipoint analysis with the IBD scoring function *all* and marker allele frequencies estimated from pedigree founders. Only nine of the 200 replicates produced a LOD score greater than 3, suggesting

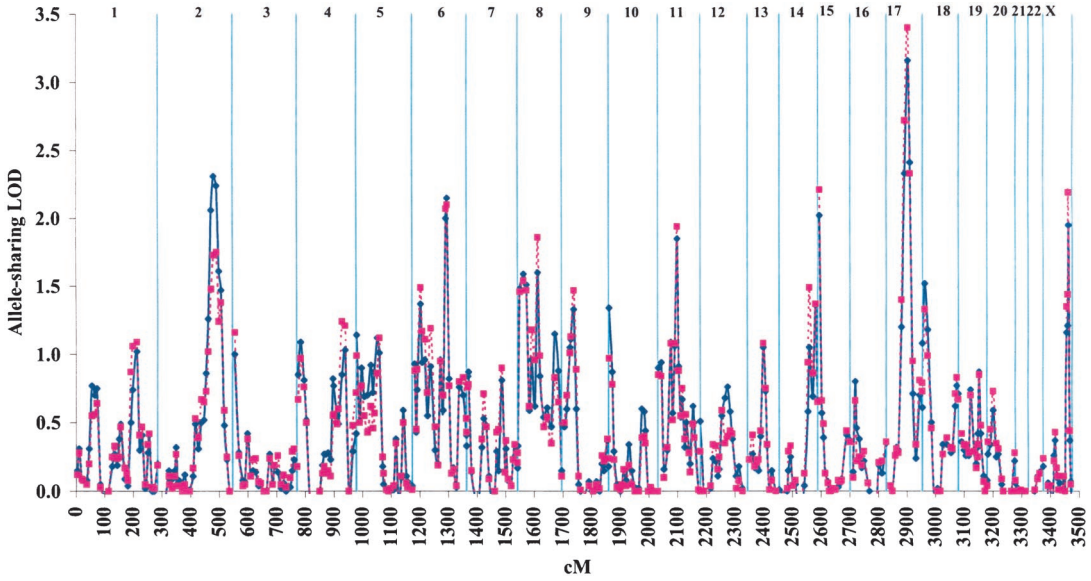

Fig. 1. Combined genome-wide screen for prostate cancer susceptibility genes using nonparametric multipoint linkage analysis among 426 hereditary prostate cancer families. Allele-sharing logarithm of the odds (LOD) scores based on the scoring functions *all* (**solid blue line**) or *pairs* (**dotted pink line**) are plotted by individual chromosomes for the whole genome.

that, given our marker map and missing data pattern, the chance of obtaining a LOD score of 3.0 from a genome-wide scan when there is no major susceptibility gene in the genome (i.e., the false-positive rate) was 4.5%. Therefore, results of this genome-wide scan of 426 HPC families provided statistically significant evidence that 17q22 harbors a prostate cancer susceptibility gene. Specifically, the estimated genome-wide empirical *P* value for the LOD score (3.16) observed at marker D17S787 was .03.

Nonparametric multipoint linkage analysis produced suggestive evidence (i.e., LOD scores greater than 2) for a prostate cancer linkage in four chromosomal regions in addition to 17q22 (Fig. 1). At the 2q32 region, a maximum LOD score was found at marker D2S117 (LOD = 2.31; $P = .0006$). At the 15q11 region, a maximum LOD score was found at marker D15S128 (LOD = 2.21; $P = .0007$). At the Xq27 region, a maximum LOD score was found at marker DXS1227 (LOD = 2.19; $P = .0007$). At the 6q22 region, a maximum LOD score was found at marker D6S287 (LOD = 2.15; $P = .0008$). We obtained similar linkage results in these regions regardless of which IBD scoring function (i.e., *all* or *pairs*) was used. On the basis of the genome-wide simulation described above, the chance of observing at least one region in the genome with a LOD score of 2.0 or greater is 76.5%, and the chance of observing four regions in the genome with LOD scores of 2 or greater is 2.0% under the null hypothesis of no linkage. Therefore, our observation of four regions in the genome with LOD scores greater than 2 provides statistical evidence that at least one of these four regions is also likely to harbor a prostate cancer susceptibility gene.

We examined the potential impact of marker allele frequencies on these linkage results and found that three (2q32, Xq25, and 6q22) of the four regions remained suggestive for linkage when conservative marker allele frequency estimates based on all family members were used (data not shown). However the evidence for linkage in the 15q11 region decreased considerably

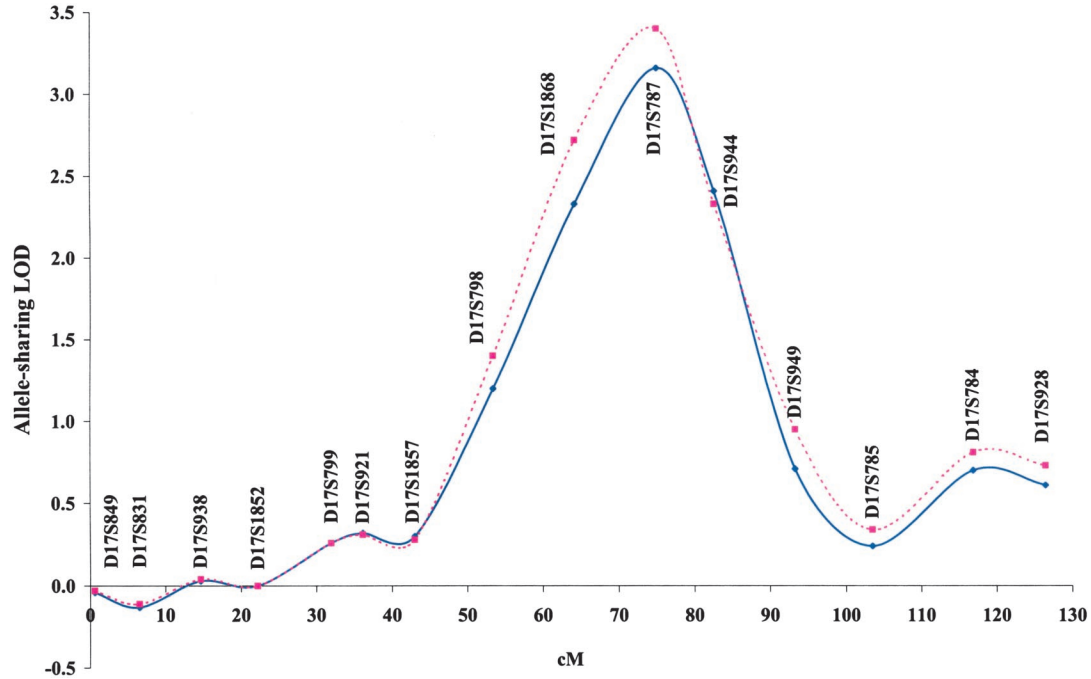

Fig. 2. Nonparametric linkage analysis for chromosome 17 among 426 hereditary prostate cancer families. Allele-sharing logarithm of the odds (LOD) scores based on the scoring functions *all* (**solid blue line**) or *pairs* (**dotted pink line**) are plotted by map position for chromosome 17.

when we used the allele frequencies from all family members (LOD = 0.99; $P = .02$).

Results of parametric multipoint linkage analyses provided similar, albeit weaker, evidence for linkage compared with the results of nonparametric linkage analyses. Using parametric analyses, we found that the 17q22 region was also the region with the strongest evidence for linkage in the genome. The maximum parametric multipoint LOD score, assuming heterogeneity (HLOD), was 2.45 at D17S787 ($P = .0008$), assuming a recessive mode of inheritance for the disease.

To potentially lower the degree of locus heterogeneity, we performed nonparametric linkage analysis among subsets of HPC families that had similar clinical and demographic characteristics. The linkage results using marker allele frequencies estimated from pedigree founders, stratified by the mean within-family age at prostate cancer diagnosis, are shown in Fig. 3. Among the 201 HPC families for which the mean age at diagnosis was less than 65 years, we obtained statistically significant evidence for linkage at chromosome region 17q22 (LOD = 3.00 at marker D17S944; $P = .0001$). Based on 100 replicates of genome-wide scan data, assuming no linkage in this subset of families, there is a 6% chance that this finding is false. D17S944 is located approximately 7 cM away from D17S787, the marker with the highest LOD score in the complete set of families. We also obtained statistically suggestive evidence for linkage at Xq27 in this subset of families (LOD = 2.30 at marker DXS1227; $P = .0006$). The linkage results were similar when conservative marker allele frequencies that were based on all family members were used in the analyses. For example, the LOD score at D17S944 was 2.20 ($P = .0007$), and the LOD score at DXS1227 was 2.12 ($P = .0009$).

Among the 226 HPC families for which the mean age at diagnosis was 65 years or older, multiple regions with extremely strong evidence for linkage were found. When marker allele frequencies estimated from pedigree founders were used, we observed a maximum LOD score of 5.57 ($P<.00001$) at the 15q11 region. Among 100 replicates of genome-wide scan data that assumed no linkage in this subset of families, no replicate reached a LOD score of this magnitude. The linkage peak was observed at marker D15S128, one of the four markers that were positive in the overall nonparametric multipoint linkage analysis. At the 2q32 region, a maximum LOD of 3.41 ($P = .00004$) was found at marker D2S117, which was also one of the four markers that were positive in the overall nonparametric multipoint linkage analysis. On chromosome 8, an approximately 50-cM region that extended from 8p21 to 8q12 had statistically significantly elevated LOD scores and a maximum LOD of 3.28 at marker D8S1771 ($P = .00005$). The evidence for linkage at each of these regions in this subset of families was considerably weaker when conservative marker allele frequencies based on all family members were used (i.e., the LOD scores were 2.68, 1.86, and 1.72 at 15q11, 2q32, and 8p21–q12, respectively).

Figure 4 shows the linkage results, stratified by the number of affected family members, obtained using marker allele frequencies estimated from pedigree founders. Among the 285 HPC families that had four or more affected family members, there was statistically significant evidence for linkage at the 17q22 region (LOD = 3.87 at marker D17S787; $P = .00001$). Among 100 replicates of genome-wide scan data that assumed no linkage in this subset of families, only one replicate had a LOD score of this magnitude. In addition, we identified a new prostate cancer linkage region at 4q35 (LOD = 3.1 at D4S1615; $P = .00008$). The chance that this finding was a false-positive was 4%, based on the simulation results. The LOD scores were lower (LOD = 2.56 at D17S787; $P = .0001$ and LOD = 2.24 at D4S1615; $P = .0007$) when we used conservative marker allele frequencies that were based on all family members.

Discussion

To date, this is the largest set of prostate cancer families analyzed by genome-wide scanning techniques. The strongest evidence for linkage was obtained at 17q22, which had a LOD score of 3.16 among 426 HPC families. On the basis of simulation results, this linkage was statistically significant at a genome-wide level. Furthermore, we obtained suggestive evidence for linkage (i.e., LOD scores greater than 2) at four other chromosomal regions, 2q32, 15q11, Xq25, and 6q22. Stratified linkage analyses among subsets of families that had similar clinical and demographic characteristics revealed two additional chromosome regions that are likely to contain prostate cancer susceptibility genes that segregate in those HPC families, including the 15q11 region among HPC families with late-onset prostate cancer and the 4q35 region among HPC families with four or more affected family members.

Prostate cancer is the most common cancer and the second leading cause of cancer death among men in the United States. In 2004, an estimated 230 110 new prostate cancer cases will be diagnosed, accounting for more than 33% of all cancers affecting men. More than 29 000 deaths are expected to result annually from this disease *(27).* Among the many suggested risk factors for prostate cancer, age, race, and family history are the three strongest and most consistently observed. Results from many family studies that have used either retrospective or cohort study designs provide consistent evidence for the aggregation of prostate cancer in families *(1).* Results of studies among twins suggest that a substantial portion of this familial aggregation is due to genetic factors because concordance rates for prostate cancer are higher among monozygotic twins than among dizygotic twins *(28,29).* More recently, Lichtenstein et al. *(2)* found concordance rates of 21% and 6% in monozygotic twins and dizygotic twins, respectively, and concluded that 42% (95% confidence interval = 29% to 50%) of the variation in prostate cancer risk may be accounted for by heritable factors. Results of segregation studies provide further evidence that genetic susceptibility to prostate cancer is most consistent with a major gene(s) on a polygenic background. Specifically, these segregation studies have suggested that 1) the familial aggregation of prostate cancer is best explained by the autosomal dominant inheritance of a rare and high-risk allele *(3–7);* 2) this inherited form of prostate cancer accounts for a substantial proportion of early-onset disease and is responsible for approximately 9% of all prostate cancers *(3);* 3) a dominant inheritance model is the best fit for families with early-onset prostate cancer, whereas a recessive or X-linked model is the best fit for families with late-onset prostate cancer *(8);* and 4) two-gene models fit the data better than single-gene models *(8,9).*

During the last decade, tremendous efforts have been made to identify major prostate cancer susceptibility genes. Results of genetic linkage studies have suggested that multiple chromosomal regions harbor prostate cancer susceptibility genes, in-

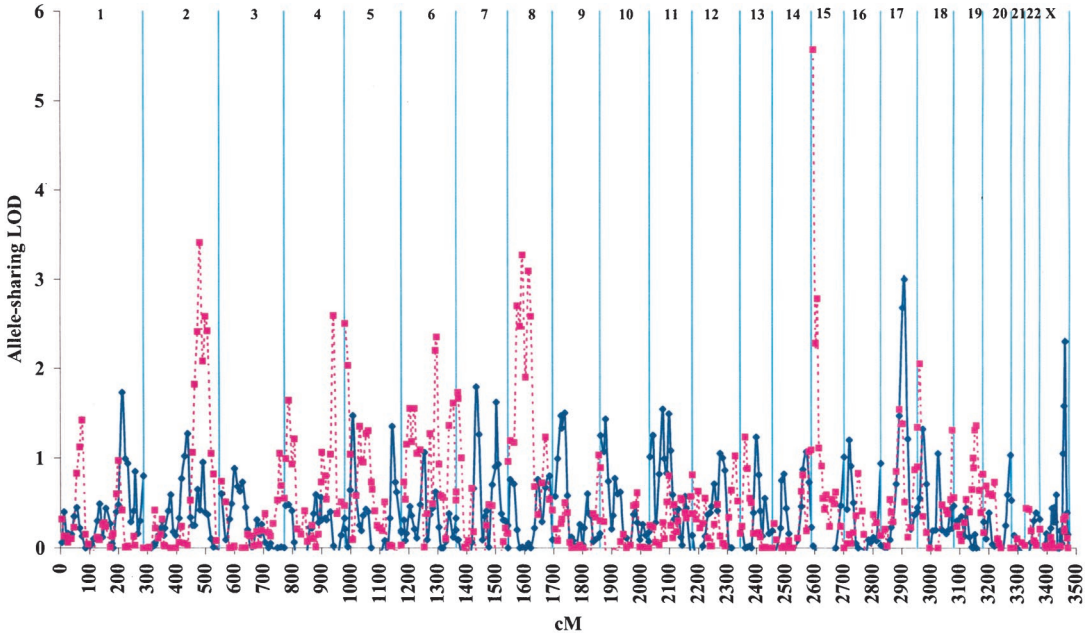

Fig. 3. Combined genome-wide screen for prostate cancer susceptibility genes using nonparametric multipoint linkage analysis in hereditary prostate cancer (HPC) families stratified by the mean within-family age at prostate cancer diagnosis. The allele-sharing logarithm of the odds (LOD) scores for 201 HPC families among which the mean age at prostate cancer diagnosis was younger than 65 years (**solid blue line**) and 226 HPC families among which the mean age at prostate cancer diagnosis was 65 years or older (**dotted pink line**) are plotted by individual chromosomes for the whole genome.

cluding HPC1 at 1q24–25 *(25,30)*, PCAP at 1q42–43 *(31)*, HPCX at Xq27–28 *(20)*, CAPB at 1p36 *(32)*, HPC20 at 20q13 *(33)*, a locus at 16q *(34)*, and a locus at 8p22–23 *(35)*. Attempts to replicate these results in other series of HPC families have yielded mixed results, however, leading to a pessimistic view regarding the utility of prostate cancer linkage studies. Nevertheless, these seemingly conflicting results are not unexpected, given the difficulties in performing linkage studies of genetically complex diseases and the small number of families involved in these published studies. Linkage analysis for prostate cancer is a difficult undertaking for the following reasons. First, there appears to be a substantial degree of genetic heterogeneity in prostate cancer, with potentially multiple modes of inheritance (dominant, recessive, and X-linked) and multiple disease-causing genes (i.e., locus heterogeneity). Second, because of the high prevalence of this disease and the multiple environmental risk factors, phenocopies of disease are likely to be common, even among HPC families. In fact, some of the clustering of prostate cancer observed among HPC families is probably due to chance rather than to shared genes. Third, incomplete and age-dependent penetrance of prostate cancer genes poses additional difficulties because of the potential for misclassification of gene carriers.

One effective approach to overcoming these difficulties is to increase the number of families in genetic linkage studies. As long as mutations in one specific major gene are responsible for some fraction of hereditary prostate cancer, linkage information will be present in some of these "linked" families. The power to detect a major prostate cancer gene ultimately depends on how many linked families are included in the study population. The larger the study population, the more likely it is that a major gene can be detected. Furthermore, when the study population includes a large number of families, it is possible to stratify

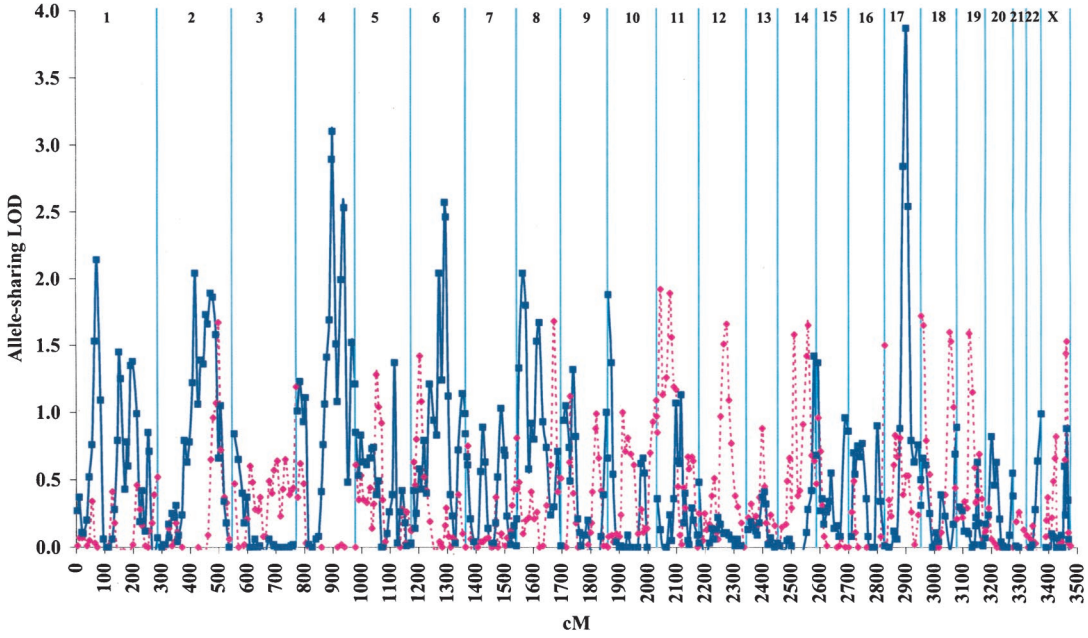

Fig. 4. Combined genome-wide screen for prostate cancer susceptibility genes using nonparametric multipoint linkage analysis among hereditary prostate cancer (HPC) families stratified by the number of affected family members. The allele-sharing logarithm of the odds (LOD) scores for 285 HPC families with four or more affected family members (**solid blue line**) and 142 HPC families with three or fewer affected family members (**dotted pink line**) are plotted by individual chromosomes for the whole genome.

families by clinical and demographic characteristics, such as the mean age at diagnosis, the number of affected members, and race/ethnicity. This approach is likely to reduce the degree of genetic heterogeneity. Our combined linkage study represents a step toward this effort, and our finding of a statistically significant linkage at 17q22 demonstrates the advantages of studying large numbers of families.

Our results provide considerable evidence that strongly suggests the presence of a prostate cancer susceptibility gene or genes in the 17q22 region. First, our simulation study, which was based on the exact pedigree structure and marker information from our dataset, provided an estimated genome-wide empirical *P* value of .03 for the LOD score (3.16) observed at 17q22. In other words, when there is no major prostate cancer susceptibility gene in the genome, the chance of obtaining a LOD score of this magnitude in a genome-wide scan is only approximately 3%. This empirical genome-wide *P* value is similar to the *P* values for genome-wide statistical significance proposed by Lander and Kruglyak *(36).* Second, the observed LOD scores at 17q22 were robust to marker allele-frequency estimates because the linkage results obtained using conservative allele frequency estimates from all family members were not substantially lower than those obtained using allele-frequency estimates from founders only. Third, evidence for a prostate cancer linkage at 17q22 came from three of the four study populations; the 175 HPC families from study population 2 had a LOD of 2.17 at D17S787, the 188 HPC families from study population 1 had a LOD of 1.38 at D17S787, and the 13 HPC families from study population 4 had a LOD of 0.68 at D17S787. Fourth, evidence for linkage at this chromosomal region was observed across different races and ethnicities. Finally, evidence for a prostate cancer linkage at 17q22 was strongest among the 201 families with early-onset disease and among the 285 families with four or more affected family members. Both of these subsets of families may be enriched for an inherited form of prostate cancer.

One of the potential candidate genes in the 17q22 region is BRCA1. However, D17S787, the marker that provided the strongest evidence for linkage at this region, is almost 12 Mbp distal to the BRCA1 gene. Nonetheless, although the role of BRCA1 in prostate cancer is not fully understood, results of several studies have suggested that men who carry BRCA1 mutations have an elevated risk of prostate cancer *(37).* Among 11 487 individuals from 699 families segregating a BRCA1 mutation that were ascertained in 30 centers across Europe and North America by the Breast Cancer Linkage Consortium, male BRCA1 mutation carriers younger than 65 years had an elevated risk for prostate cancer (relative risk = 1.82, 95% confidence interval = 1.01 to 3.29) whereas male BRCA1 mutation carriers who were age 65 years or older did not *(38).* So far, most of the published studies on BRCA1 and prostate cancer have focused on deleterious mutations (i.e., nonsense or frameshift mutations). The role of other sequence variants (i.e., missense mutations) of the BRCA1 gene in prostate cancer susceptibility may also be important and warrants further systematic investigation.

We also obtained suggestive evidence for prostate cancer linkage at several other chromosomal regions (i.e., 2q32, 15q11, Xq27, and 6q22) in the complete dataset of 426 HPC families. On the basis of results from a genome-wide simulation of 200 replicates, the chance of observing four chromosomal regions with LOD scores of 2 or greater in a single replicate is 2%. Evidence for linkage at 2q32 was previously reported among 564 men from 254 families with prostate cancer by Goddard et al. *(39).* A LOD score of 2.48 was found for marker D2S434, which is 20 cM telomeric to our peak marker, D2S117. That study also provided evidence for a linkage at 15q11, based on a LOD score of 2.42 at marker D15S165, which maps approximately 6 cM centromeric to our peak marker, D15S128. Evidence for a prostate cancer linkage at Xq was previously reported by our group in a combined analysis of 360 HPC families *(20).* The linked region extended from Xq26 to Xq28, with the highest linkage observed at marker DXS297. The peak marker identified in the current study, DXS1227, is in this region and is approximately 5 cM closer to the centromere than DXS297. It is important to note that 193 HPC families in the current study were included in that previous study *(20);* therefore, the results of this study are not an independent replication of the previous study, but rather, they provide additional support for a prostate cancer susceptibility gene or genes at Xq.

One potential limitation of a combined linkage analysis such as this one is the possibility of introducing clinical heterogeneity due to differences in study subject ascertainment, as well as differences in screening and diagnostic practices among the different research centers, especially when the centers are located in different countries. Although we tried to minimize sources of heterogeneity (e.g., by using standardized diagnostic criteria for prostate cancer cases), some clinical heterogeneity remained. As one example, the mean age at diagnosis tended to be earlier among the two U.S. study populations than among the two European study populations, probably because of the widespread use of prostate-specific antigen screening in the United States.

The large number of HPC families in our study will make it possible to test for gene–gene interactions in future linkage analyses. It is likely that multiple genes acting independently (i.e., locus heterogeneity) and/or simultaneously (i.e., epistasis) will increase the susceptibility to a genetically complex disease such as prostate cancer. Linkage analyses that model gene–gene interactions will most likely improve the statistical power to detect disease susceptibility genes. Combined linkage analysis using a large number of families is a good first step toward achieving appropriate statistical power.

A novel aspect of this study—its size—is the combination of data from four different familial prostate cancer study populations. The International Consortium for Prostate Cancer Genetics has an even larger combined analysis of linkage data from at least 10 different study populations currently underway. It will be interesting to see how the results of these multi-institutional studies compare. In the meantime, results of this large genome-wide scan for prostate cancer susceptibility genes provide a basis for renewed interest, excitement, and confidence in genetic linkage studies of prostate cancer. Fine-mapping studies and identification of prostate cancer susceptibility genes in these linked regions are warranted and are currently underway.

REFERENCES

(1) Isaacs WB, Xu J, Walsh PC. Hereditary prostate cancer. In: Chung LW, Isaacs WB, Simons JW, editors. Prostate cancer, biology, genetics, and the new therapeutics. Totowa (NJ): Humana Press; 2001. p. 13–28.

(2) Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer—

analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
(3) Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC. Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci U S A 1992; 89:3367–71.
(4) Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN. Evidence for autosomal dominant inheritance of prostate cancer. Am J Hum Genet 1998;62: 1425–38.
(5) Gronberg H, Bergh A, Damber JE, Emanuelsson M. Cancer risk in families with hereditary prostate carcinoma. Cancer 2000;89:1315–21.
(6) Verhage BA, Baffoe-Bonnie AB, Baglietto L, Smith DS, Bailey-Wilson JE, Beaty T, et al. Autosomal dominant inheritance of prostate cancer: a confirmatory study. Urology 2001;57:97–101.
(7) Valeri A, Briollais L, Azzouzi R, Fournier G, Mangin P, Berthon P, et al. Segregation analysis of prostate cancer in France: evidence for autosomal dominant inheritance and residual brother-brother dependence. Ann Hum Genet 2003;67:125–37.
(8) Cui J, Staples MP, Hopper JL, English DR, McCredie MR, Giles GG. Segregation analyses of 1476 population-based Australian families affected by prostate cancer. Am J Hum Genet 2001;68:1207–18.
(9) Gong G, Oakley-Girvan I, Wu AH, Kolonel LN, John EM, West DW, et al. Segregation analysis of prostate cancer in 1,719 white, African-American and Asian-American families in the United States and Canada. Cancer Causes Control 2002;13:471–82.
(10) Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, et al. A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 2001;27:172–80.
(11) Carpten J, Nupponen N, Isaacs S, Sood R, Robbins C, Xu J, et al. Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat Genet 2002;30:181–4.
(12) Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Hu JJ, et al. Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. Nat Genet 2002; 32:321–5.
(13) Dong X, Wang L, Taniguchi K, Wang X, Cunningham JM, McDonnell SK, et al. Mutations in CHEK2 associated with prostate cancer risk. Am J Hum Genet 2003;72:270–80.
(14) Edwards SM, Kote-Jarai Z, Meitz J, Hamoudi R, Hope Q, Osin P, et al. Two percent of men with early-onset prostate cancer harbor germline mutations in the BRCA2 gene. Am J Hum Genet 2003;72:1–12.
(15) Xu J, Zheng SL, Carpten JD, Nupponen NN, Robbins CM, Mestre J, et al. Evaluation of linkage and association of HPC2/ELAC2 in patients with familial or sporadic prostate cancer. Am J Hum Genet 2001;68:901–11.
(16) Xu J, Gillanders EM, Isaacs SD, Chang BL, Wiley KE, Zheng SL, et al. Genome-wide scan for prostate cancer susceptibility genes in the Johns Hopkins hereditary prostate cancer families. Prostate 2003;57:320–5.
(17) Lange EM, Gillanders EM, Davis CC, Brown WM, Campbell JK, Jones M, et al. Genome-wide scan for prostate cancer susceptibility genes using families from the University of Michigan Prostate Cancer Genetics Project finds evidence for linkage on chromosome 17 near BRCA1. Prostate 2003;57:326–34.
(18) Wiklund F, Gillanders EM, Albertus JA, Bergh A, Damber JE, Emanuelsson M, et al. Genome-wide scan of Swedish families with hereditary prostate cancer: suggestive evidence of linkage at 5q11.2 and 19p13.3. Prostate 2003;57:290–7.
(19) Schleutker J, Baffoe-Bonnie AB, Gillanders E, Kainu T, Jones MP, Freas-Lutz D, et al. Genome-wide scan for linkage in Finnish hereditary prostate cancer (HPC) families identifies novel susceptibility loci at 11q14 and 3p25-p26. Prostate 2003;57:280–9.
(20) Xu J, Meyers D, Freije D, Isaacs S, Wiley K, Nusskern D, et al. Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 1998;20:175–9.
(21) Schleutker J, Matikainen M, Smith J, Koivisto P, Baffoe-Bonnie A, Kainu T, et al. A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: frequent HPCX linkage in families with late-onset disease. Clin Cancer Res 2000;6:4810–5.
(22) Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97–101.
(23) Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. Biometrics 1994;50:118–27.
(24) Kong A, Cox, NJ. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179–88.
(25) Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD, et al. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 1996;274:1371–4.
(26) Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996;58:1347–63.
(27) Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ. Cancer statistics, 2003. CA Cancer J Clin 2003;53:5–26.
(28) Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, et al. Cancer statistics, 2004. CA Cancer J Clin 2004;54:8–29.
(29) Page WF, Braun MM, Partin AW, Caporaso N, Walsh P. Heredity and prostate cancer: a study of World War II veteran twins. Prostate 1997;33: 240–5.
(30) Xu J, Zheng SL, Chang B, Smith JR, Carpten JD, Stine OC, et al. Linkage of prostate cancer susceptibility loci to chromosome 1. Hum Genet 2001; 108:335–45.
(31) Berthon P, Valeri A, Cohen-Akenine A, Drelon E, Paiss T, Wohr G, et al. Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. Am J Hum Genet 1998;62:1416–24.
(32) Gibbs M, Stanford JL, McIndoe RA, Jarvik GP, Kolb S, Goode EL, et al. Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. Am J Hum Genet 1999;64:776–87.
(33) Berry R, Schroeder JJ, French AJ, McDonnell SK, Peterson BJ, Cunningham JM, et al. Evidence for a prostate cancer-susceptibility locus on chromosome 20. Am J Hum Genet 2000;67:82–91.
(34) Suarez BK, Lin J, Burmester JK, Broman K, Weber JL, Banerjee TK, et al. A genome screen of multiplex sibships with prostate cancer. Am J Hum Genet 2000;66:933–44.
(35) Xu J, Zheng SL, Hawkins GA, Faith DA, Kelly B, Isaacs SD, et al. Linkage and association studies of prostate cancer susceptibility: evidence for linkage at 8p22–23. Am J Hum Genet 2001;69:341–50.
(36) Lander ES, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–7.
(37) Simard J, Dumont M, Labuda D, Sinnett D, Meloche C, El Alfy M, et al. Prostate cancer susceptibility genes: lessons learned and challenges posed. Endocr Relat Cancer 2003;10:225–59.
(38) Thompson D, Easton DF. Cancer incidence in BRCA1 mutation carriers. J Natl Cancer Inst 2002;94:358–65.
(39) Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM. Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. Am J Hum Genet 2001;68:1197–206.

NOTES

E. M. Gillanders and J. Xu contributed equally to the study.

Partially supported by Public Health Service Specialized Programs of Research Excellence (SPORE) grants CA58236 (to W. B. Isaacs) and CA69568 (to K. A. Cooney), and by grants CA89600-02 (to W. B. Isaacs), CA79596 (to K. A. Cooney), and CA95052-01 (to J. Xu) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

We thank all of the family members who participated in this study.

Manuscript received November 10, 2003; revised June 9, 2004; accepted June 22, 2004.