Genetic Linkage of Prostate Cancer Risk to the Chromosome 3 Region Bearing FHIT

Garry P. Larson,1 Yan Ding,2 Li S-C. Cheng,3 Cathryn Lundberg,1 Virgil Galagande,1 Guillermo Rivas,1 Louis Geller,1 Jeffrey Weitzel,1,13 Deborah MacDonald,1 John Archambeau,1 Jerry Slater,1 Donna Neuber,1 Mary B. Daly,1 Irene Angel,1 Al B. Benson III,1 Kimberly Smith,1 John M. Kirkwood,1 Peter J. O’Dwyer,10 Barbara Raskay,10 Rebecca Surphee,11 Rosalind Drew,11 James A. Stewart,12 Jae Wernld,12 David Johnson,11 John C. Ruckdeschel,14 Robert C. Elston,15 and Theodore G. Krontiris1

1Divisions of Molecular Medicine and 1Information Sciences, Beckman Research Institute and 1Department of Cancer Genetics, City of Hope National Medical Center, Duarte, California; 2Department of Radiology Medicine, Loma Linda University Medical Center, Loma Linda, California; 3Division of Biostatistics, Dana-Farber Cancer Institute, Harvard Medical School and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; 4Department of Population Science, Fox Chase Cancer Center, Philadelphia, Pennsylvania; 5Department of Medicine, Division of Hematology/Oncology and 1Clinical Research Office, Robert J. Lurie Comprehensive Cancer Center, Northwestern University School of Medicine, Chicago, Illinois; 6Department of Medicine, Division of Hematology/Oncology, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania; 7Department of Medicine, Hematology-Oncology Division, University of Pennsylvania Cancer Center, Philadelphia, Pennsylvania; 8Interdisciplinary Oncology Program, H. Lee Moffitt Cancer Center and Research Institute, University of South Florida, Tampa, Florida; 9University of Wisconsin Comprehensive Cancer Center and University of Wisconsin School of Medicine, Madison, Wisconsin; 10Division of Hematology & Oncology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee; 11Barbara Ann Karmanos Cancer Institute, Detroit, Michigan; and 12Department of Epidemiology and Biostatistics, Case Western Reserve University, Metro Health Medical Center, Cleveland, Ohio

Abstract

We conducted linkage analysis of 80 candidate genes in 201 brother pairs affected with prostatic adenocarcinoma. Markers representing two adjacent candidate genes on chromosome 3p, CDC25A and FHIT, showed suggestive evidence for linkage with single-point identity-by-descent allele-sharing statistics. Fine-structure multipoint linkage analysis yielded a maximum LOD score of 3.17 (P = 0.00007) at D3S1234 within FHIT intron 5. For a subgroup of 38 families in which three or more affected brothers were reported, the LOD score was 3.83 (P = 0.00001). Further analysis reported herein suggested a recessive mode of inheritance. Association testing of 16 single nucleotide polymorphisms (SNP) spanning a 381-kb interval surrounding D3S1234 in 202 cases of European descent with 143 matched, unrelated controls revealed significant evidence for association between case status and the allele of single nucleotide polymorphism rs760317, located within intron 5 of FHIT (Pearson’s χ² = 8.54, df = 1, P = 0.0035). Our results strongly suggest involvement of germine variations of FHIT in prostate cancer risk. (Cancer Res 2005; 65(3): 805-14)

Introduction

Prostate cancer (CaP, MIN 176807) is expected to result in 32% of all new cancer cases among American males in 2003 (American Cancer Society statistics, 2003). It is the second leading cause of cancer deaths in males, with approximately one male in six likely to develop the disease during his lifetime. Although the disease is multifactorial, deriving from both genetic and environmental components, deciphering the genetic factors that play a role would provide improved opportunities for diagnosis and, possibly, treatment. Large studies of twins in Scandinavian countries suggest that a significant component of risk may be attributable to genetic factors (1). However, large differences in disease prevalence observed in populations of varying ethnic backgrounds, such as the high incidence in African Americans versus the relatively low incidence seen in Asians, support the role of locus heterogeneity and environmental factors in disease risk (2).

Using both multigenerational pedigree and affected sibling pair approaches, putative prostate cancer susceptibility loci have been repeatedly mapped to chromosomes 1q24-q25, 1q42-q43, 1p36, 4q24, 5p13, 8p22-p23, 16q23, 17p11, 20q13, and Xq27-q28 (3–6). So far, three genes—the RNase L gene (RNASEL, 1q24-q25, HPCL1), ELAC2 (17p11, HPCL2), and the macrophage scavenger receptor 1 (MSR1, 8p22)—have been identified via subsequent positional cloning approaches (7–9). Mutations in these genes have been reported to be significantly associated with prostate cancer risk. However, in many instances both linkage and association results have been difficult to reproduce consistently, possibly because of locus and/or allele heterogeneity. Segregation of mutations was often found in only a small number of pedigrees originally showing linkage to these regions. A meta-analysis of associations of variants in ELAC2 and prostate cancer risk also concluded that the original maximal risk estimates were inflated, suggesting a limited role for this locus (10). The complex epidemiology of prostate cancer has been highlighted in two recent reviews (3, 11). Collectively, no single gene identified to date has been implicated by itself as being responsible for a large portion of familial prostate cancer.

Association studies using biologically plausible candidate genes have showed variable success. A number of polymorphisms associated with some candidates are fairly common in the population and are believed to function as low-penetrance disease alleles influencing risk, prognosis, or response to therapy. Two types of polymorphisms have been described in the androgen receptor (AR) gene and are associated with risk. Polymorphisms alleles encoded by polymorphic CAG repeats in the transcriptional activation domain show an inverse relationship between CAG length and risk (12). Other exonic AR mutations seem to be associated with the metastatic or growth potential of CaP tumors (13).
Polymorphisms in the CYP gene family influence the age of onset and the metabolism of chemotherapeutic drugs. A promoter polymorphism in CYP3A4 is a prognostic indicator for the likelihood of patients with benign prostatic hyperplasia developing CaP (14). Studies also found CaP risk associated with mutations in genes involved in breast cancer risk, BRCAl2 and CHEK2, both involved in DNA repair (15–17). Thus, there is growing evidence of low-penetrance disease alleles playing a role in multiple cancer types. We have conducted linkage analyses of candidate genes in a cohort of CaP-affected sibling pairs (ASP). Among our targets was an extensive list of genes involved in DNA metabolism, cell cycle control, and steroid and xenobiotic metabolism. Genes/loci implicated in cancer risk from previously published studies were also included. We genotyped preexisting or newly developed microsatellite markers for these candidate genes. Here we report linkage results for our candidate genes located on chromosome 3 and subsequent support of linkage using single nucleotide polymorphism (SNP) haplotype association tests.

Materials and Methods

Subjects

All siblings affected with CaP were recruited through a consortium of institutions involved with the Eastern Cooperative Oncology Group, the City of Hope National Medical Center, and the Department of Radiation Medicine at Loma Linda University Medical Center. Our ascertainment criteria were a proband (index case) with documented prostatic adenocarcinoma verified by medical records and self-reported additional affected brother(s) (full sibling) who was alive and willing to participate in the studies. We obtained and verified pathology reports for all but three index cases. Combined Gleason scores of needle biopsies and/or surgical specimens were available for 88% of the index cases. The accuracy of sibling- and self-reporting of prostate cancer was supported by 28 pathology reports we have collected for siblings. Other researchers have also concluded that overreporting of cancer incidence is rare among first-degree relatives (18). Each institution’s Institutional Review Board approved this study. Informed consent was obtained from all participants.

Our initial ASP cohort consisted of 433 patients in 207 families. Data of cancer incidence among first-degree relatives of probands were collected in 93% (193/207) of the families for parents and in 57% (118/207) of the families for siblings. Among these families, 38 reported a CaP-affected father. Thirty-nine families reported three or more affected brothers, of which 14 each contributed samples for three affected brothers. One family had seven affected brothers sampled. We were able to obtain samples from only the proband and one sibling in the remaining 24 families. Additional affected brothers were not recruited due to death or refusal to participate. Parents were not collected in this study because we observed that fewer than 5% of siblings had both parents available for sampling. Six sibling pairs from six families were removed from linkage analysis because they were either identified as monozygotic twins or unrelated through paternal descent. For an initial screen of candidate genes, we assembled a “primary pair group” (including the family with seven affected brothers), which consisted of the index case and the first affected sibling recruited into the study. In the “all pair group,” we omitted the seven-sibling family. Unless otherwise stated, the seven-sibling family was conservatively omitted from all analyses because this family alone contributed 21 possible pairing combinations, whereas other families presented three pairs at most. Its inclusion could greatly inflate the type I error rate in those analyses that assume all pairs are independent. We also did subgroup analyses based on family history and age at diagnosis. The first subgroup consisted of families that reported three or more affected brothers (“multiple-affected group,” 66 pairs from 38 families). The second subgroup consisted of families in which the age at diagnosis for all brothers was <65 years (“age at diagnosis <65 group,” 66 pairs from 60 families). Sixteen pairs from 10 families were shared between the two subgroups. The mean age at diagnosis for index cases from the multiple-affected group was not statistically different from that of all ASPs (63.6 versus 65.8). The mean age at diagnosis for index cases from the age at diagnosis <65 group was 58.7 years. The overall characteristics of our cohort are summarized in Table 1.

We collected self-reported ethnicity data for both maternal and paternal grandparents from -75% of our patients. Our patient population was predominantly of European origin. Among families that provided information, ~96% reported Caucasian ancestry, ~2% African American, ~1% Native American, and ~1% other. For association analyses, we assembled 1 sibling from each family into a case population, totaling 207. The control population consisted of 146 individuals of Caucasian ancestry. It consisted of three subgroups: cancer-free individuals with a mean age of 42 years (range, 17 to 81, n = 73), prostate cancer–free parents of breast cancer sister pairs (mean age, 73, range 57 to 85, n = 34, obtained in the same Eastern Cooperative Oncology Group study), and prostate cancer–free males at least 65 years of age (n = 39). All cases and controls were subjected to population structure analyses as discussed below.

Genotyping

DNA was extracted from peripheral blood samples using a modified salting-out procedure (19). Genotyping for microsatellite markers was done on all ASP samples using routine multiplex methodologies on an ABI 377 sequencer. On average one to two microsatellite markers were genotyped per candidate locus in the first round of screening. Six of our candidate genes resided on chromosome 3 (VHL, PCAF, MLH1, CDC25A, FHIT, and MCM2). For multipoint analysis on chromosome 3, samples were typed for a total of 28 microsatellite markers (Table 2). Two of these markers were newly developed intronic markers from BAC genomic sequence (CDC25a2, BAC AC09207, primers GGGGTGCGAGGTTTGG and TCCCCAGGCT-CAGGTGAT; and pCAFa, BAC AC104190, primers AATAAAACCAACC-CAATGA and GAGGAAAACGGAAGAAAAT). SNP genotyping was done on cases and controls using a modified, multiplex protocol based on ABI SNaPshot Multiplex Kit on an ABI 377 sequencer (20). The length of

![Table 1. Characteristics of prostate cancer ASP families](http://cancerres.aacrjournals.org)
extension primers was modified by the addition of a poly(dA) tail at the 5' end to achieve variable sizes from 18 to 50 nucleotides for electrophoresis multiplexing. Size standards for SNP genotyping consisted of X-rhodamine–labeled 16, 32, and 52 mers of poly(dGACT)n. Alleles were identified using GeneScan 3.0. Nonspecific extension of one allele was observed for one SNP and a high failure rate was found for another. Both were discarded from subsequent analysis. Extreme deviation from Hardy-Weinberg equilibrium was observed for one SNP and a high failure rate was found for another. Both were discarded from subsequent analysis.

Table 2. Markers used for multipoint analysis

<table>
<thead>
<tr>
<th>Markers</th>
<th>Heterozygosity rate</th>
<th>Position (cM)†</th>
<th>UCSC position, July 2003</th>
<th>Comments ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1317</td>
<td>0.706</td>
<td>27.68</td>
<td>10298658</td>
<td>VHL</td>
</tr>
<tr>
<td>D3S1335</td>
<td>0.767</td>
<td>27.94†</td>
<td>10254548</td>
<td>VHL</td>
</tr>
<tr>
<td>pCAFa</td>
<td>0.825</td>
<td>40.68†</td>
<td>20138241</td>
<td>PCAF</td>
</tr>
<tr>
<td>D3S1561</td>
<td>0.698</td>
<td>61.92</td>
<td>36444920</td>
<td>MLH1</td>
</tr>
<tr>
<td>D3S1298</td>
<td>0.885</td>
<td>62.93†</td>
<td>38009388</td>
<td>MLH1</td>
</tr>
<tr>
<td>D3S2304</td>
<td>0.588</td>
<td>67.22</td>
<td>42775941</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S3647</td>
<td>0.746</td>
<td>67.73</td>
<td>43539737</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S2420</td>
<td>0.788</td>
<td>70.55</td>
<td>48028036</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S3560</td>
<td>0.669</td>
<td>70.58†</td>
<td>48155020</td>
<td>CDC25a</td>
</tr>
<tr>
<td>CDC25a2</td>
<td>0.857</td>
<td>70.59†</td>
<td>48170150</td>
<td>CDC25a</td>
</tr>
<tr>
<td>D3S1581</td>
<td>0.884</td>
<td>70.66†</td>
<td>48557869</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S1588</td>
<td>0.807</td>
<td>72.68</td>
<td>54055293</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S2408</td>
<td>0.697</td>
<td>76.58</td>
<td>55667768</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S3048</td>
<td>0.592</td>
<td>77.38</td>
<td>56095168</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S2402</td>
<td>0.792</td>
<td>78.91</td>
<td>58174295</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S3553</td>
<td>0.912</td>
<td>78.96†</td>
<td>58401230</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S1540</td>
<td>0.918</td>
<td>79.99†</td>
<td>59484073</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S3577</td>
<td>0.725</td>
<td>80.10</td>
<td>59576704</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S1523</td>
<td>0.692</td>
<td>81.23</td>
<td>60064809</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S14103</td>
<td>0.831</td>
<td>82.01†</td>
<td>60389874</td>
<td>FHT</td>
</tr>
<tr>
<td>D3S1300</td>
<td>0.83</td>
<td>82.23</td>
<td>60467319</td>
<td>FHT</td>
</tr>
<tr>
<td>D3S1481</td>
<td>0.839</td>
<td>82.58†</td>
<td>60615893</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S1312</td>
<td>0.767</td>
<td>85.07</td>
<td>62363525</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S1600</td>
<td>0.768</td>
<td>86.78†</td>
<td>63277480</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S1287</td>
<td>0.646</td>
<td>88.25</td>
<td>64164382</td>
<td>Multipoint</td>
</tr>
<tr>
<td>D3S3584</td>
<td>0.666</td>
<td>134.26</td>
<td>128497626</td>
<td>MCM2</td>
</tr>
<tr>
<td>D3S3606</td>
<td>0.834</td>
<td>134.60</td>
<td>128521221</td>
<td>MCM2</td>
</tr>
<tr>
<td>D3S3607</td>
<td>0.734</td>
<td>135.10</td>
<td>128593996</td>
<td>MCM2</td>
</tr>
</tbody>
</table>

*deCode map position (Kong et al., ref. 24).
†Candidate gene or multipoint marker.
‡Interpolated genetic position using flanking markers of known deCode genetic location.

based on the UCSC July 2003 assembly. Beyond identifying of Mendelian inconsistencies, microsatellite genotyping errors were identified using the error function in MERLIN and supported by inspection of identity-by-descent (IBD) output files from both MERLIN and GENIBD (S.A.G.E.). These genotypes were treated as missing values in multipoint analyses. Empirical P values were calculated using MERLIN to simulate replicates of random genotypes of markers with the allele frequencies, assuming no linkage.

Analysis of Population Structure in Cases and Controls. Analyses of population structure were done on 550 cancer cases and 146 controls using STRUCTURE (25) with 116 unlinked microsatellites across the genome. The cases comprised one individual from each of the 207 CaP families in this study and an additional 343 breast cancer cases to increase the number of non-European individuals in the data set, which provided a more reliable characterization of population structure. Without using prior information on ethnic background, each of 10 runs was done with 10^6 iterations after 10^6 iterations of burn-in period under the option of correlated allele frequencies. All seven known African American cases, two of which are prostate cancer cases, and one Puerto Rican case were found to cluster tightly together. None of the controls was clustered with African Americans but three were clustered close to African Americans. We observed consistent results in all 10 runs assuming the presence of two to five populations. Excluding African American and the Puerto Rican samples from the data set, STRUCTURE was unable to detect any population...
structure. Rosenberg et al. reported similar difficulty detecting population structure in European populations, allowing the possibility of subtle population stratifications among individuals of European descent (26). Aside from three individuals that clustered close to African Americans, we were able to cluster the remaining cases of unknown ethnicity with other cases of known European descent and included them when testing association. After the removal of 5 CaP cases and 3 controls that were clustered with or close to African Americans, our cases and controls of matching genetic background used in subsequent association tests were 202 and 143 individuals respectively.

Association Tests. For SNP data, we did \( \chi^2 \) tests of Hardy-Weinberg equilibrium for each marker. Haplotypes of SNP markers were reconstructed combining data from cases and controls using PHASE 2.0 (27). Genotype and haplotype frequencies were compared between case and control groups using Pearson's \( \chi^2 \) test. Empirical \( P \) values were calculated using a permutation test of the null hypothesis that cases and controls were random draws from a common set of haplotype frequencies using PHASE 2.0 (PHASE 2.0 Instruction Manual, M. Stephens, 2003).

Homogeneity Tests. Because our controls consisted of three subgroups, we tested the associated SNPs for homogeneity across the three sets using \( \chi^2 \) tests with 6 degrees of freedom (df) in a 4 \( \times \) 3 contingency table for neighboring pairwise haplotypes (i.e., haplotypes formed by the alleles at two neighboring SNPs), and with 2 df in a 2 \( \times \) 3 contingency table for single SNP genotypes.

Results

Candidate Gene Screening. We systematically conducted single point IBD sharing calculations (SIBPAL, S.A.G.E. 4.3) for 118 markers tightly linked to 80 candidate genes, covering \( \sim 80 \) cM, in the primary pair group (Supplemental Fig. S1). The candidates were previously implicated in pathways involving DNA repair, cell cycle control, and steroid hormone metabolism. Among markers that exceeded an initial criterion of one-sided \( P < 0.05 \) were those for three candidate genes D3S1561 (MLH1), D3S3560 (CDC25A), and D3S4103 (FHIT), which showed IBD mean sharing of 0.536 (SE \( \pm 0.021 \), \( P = 0.097 \)), 0.532 (SE \( \pm 0.015 \), \( P = 0.034 \)), and 0.539 (SE \( \pm 0.021 \), \( P = 0.065 \)). These three markers resided within an interval of \( \sim 18.7 \) and 20.1 cM, respectively, on the Marshfield and deCode (24) genetic maps, and so may be within a single linkage region.

Multipoint Linkage Analysis. Using a two-stage approach as suggested by Elston et al. (28), we expanded the preliminary analysis of linkage results for these three candidate genes (MLH1, CDC25A, and FHIT) by genotyping 26 additional markers spanning 107 cM across chromosome 3 (Table 2). Eight of these markers were tightly linked to three additional candidate genes (VHL, pCAF, and MCM2) from our initial screen, whereas the remaining 18 markers were located in a 21-cM interval surrounding D3S3560 and D3S4103. Markers at two of the candidates (pCAF and CDC25a2) were newly described. We did linkage analysis on the entire cohort (200 families) using the S.A.G.E. program LODPAL (29) and MERLIN (23). For the 14 sibships with three affected brothers available for analysis, we assumed that all pairs were independent (30). The results are shown in Fig. 1A. The strongest evidence of linkage was detected for D3S1234 (located in intron 5 of FHIT) at 81.23 cM (LOD score = 3.15, \( P = 0.00007 \)) using LODPAL; there were peaks for both CDC25a2 (15 kb downstream of CDC25a) at 70.55 cM (NPLall = 1.90, \( P = 0.03 \)) and D3S1234 at 81.23 cM (NPLall = 1.84, \( P = 0.03 \)) using MERLIN (Fig. 1A). This broad linkage region encompassed peaks at both candidate genes.

To reduce potential heterogeneity in our sample, we tested the linkage signal on chromosome 3 in the two stratified data sets (multiple affecteds and age at diagnosis <65) and found significantly stronger linkage in the subgroup consisting of those families with more than two affected siblings (Fig. 1B). Again, we detected two linkage peaks at the two candidate genes in the multiple-affected group. LODPAL generated the maximum LOD of 3.83 (N = 0.00001) at 81.23 cM (D3S1234) and a secondary peak of 2.19 at 70.59 cM (CDC25a2). Adding the 21 pairs from the family with seven affected brothers, the maximum LOD increased to 4.46. On the other hand, MERLIN produced a maximum NPLall of 2.94 (N = 0.002) at 70.59 cM and a smaller peak of 2.38 (N = 0.009) at 81.23 cM. For the multiple-affected group, the empirical \( P \) value was <0.002 for the peak at 70.55 cM and <0.015 for the peak at 81.23 cM.

Further Characterization of the Linkage Region. Because the maximum peaks produced by the two programs were 11 cM apart, we compared IBD allele-sharing distributions calculated by the two programs. In the multiple-affected subgroup, both programs produced a maximum 2 allele IBD sharing of 0.49 and a minimum 1 allele IBD sharing of 0.21 at D3S1234 (Fig. 2A), corresponding to the major LOD score peak from LODPAL and the secondary NPL peak from MERLIN. Assuming a dominant mode of inheritance (achieved by setting the \( \alpha \) parameter equal to 1 in LODPAL; ref. 31), the maximum LOD score was 2.1 at CDC25a2. Assuming a recessive locus (\( \alpha = 100 \)), the maximum LOD score was 8.2 at D3S1234.
LOD score was 3.7 at D3S1234 (Fig. 2B). In a detailed model-based analysis of the data set using GENEHUNTER, we tested a series of models with a fixed 0.95 penetrance for the susceptible genotype(s) and a 0.05 phenocopy penetrance for the other genotype(s); the disease allele frequencies tested were 0.001 to 0.1 for dominant models and 0.001 to 0.2 for recessive models. The best fit was a recessive model with a disease allele frequency of 0.07, producing a maximum LOD score of 3.64 at D3S1234 ($P = 0.00004$; Fig. 2C). Given these results, we focused further analysis around this FHIT marker.

Under the assumption of a recessive model, we attempted to narrow the disease interval by examining key meiotic recombinants in which 2 allele IBD decayed on either side of D3S1234. We examined IBD output files from GENIBD (S.A.G.E.) and, from 10 families in the entire cohort, identified 10 sibling pairs that may define a minimum region of 2 alleles shared IBD surrounding D3S1234 (Fig. 3B and C). Therefore, we concentrated our subsequent SNP based studies on a 2.23-cM (1.1 Mb) interval encompassing D3S1234.

**Association Tests.** We initially explored linkage disequilibrium within this interval using a coarse set of seven SNPs (Fig 3B). Because linkage disequilibrium was not observed in the 7-SNP set, we next selected a denser 16-SNP set encompassing D3S1234 (Fig. 3A). These SNPs, including rs212004 from the initial set, spanned a 381-kb region between rs639244 and rs732380 with an average spacing between adjacent SNPs of 25 kb (range, 7-69 kb). Table 3 lists the minor allele nucleotides, their frequencies, location within FHIT, and adjacent pairwise linkage disequilibrium measurements. As shown in Table 3 (last two columns), we found evidence of high linkage disequilibrium for only three neighboring SNPs (rs802774-rs810615, rs760317-rs722070, and rs213294-rs213408). Two additional pairs of SNPs (rs212046-rs212004 and rs1882904-rs213294) displayed inconsistent $D'$ (high) and $\Delta^2$ (low) values, involving SNPs of relatively lower minor allele frequencies. Zabetian et al. (32) suggested $\Delta^2$ as the better predictor of phenotype correlation to the degree of linkage disequilibrium between a marker and a disease mutation. Association tests were then done between cases and controls on both individual SNP genotypes and haplotypes formed from pairs of adjacent loci.

Assuming a recessive inheritance model, we analyzed genotype and haplotype data in two comparisons. First, we compared frequencies for all index cases against controls ("All cases" in Table 3). Second, we compared the subgroup of cases that shared 2 alleles in the region with their brother(s) against the controls ("2 IBD cases" in Table 3). Table 3 lists the $\chi^2$ tests on frequency distributions of genotypes and haplotypes between these case-control groups. The maximum association was detected for the...
SNP pair hCV8351378-rs760317 (Pearson’s $\chi^2 = 15.84$, df 3, $P = 0.0012$) between the 2 IBD subset and all controls (Table 3, columns 12 and 13). Significant association was also detected for a single SNP rs760317 (Pearson’s $\chi^2 = 8.54$, df 1, $P = 0.0035$; Table 3, columns 8 and 9). There was no evidence of heterogeneity among the three control subgroups for these SNPs (Pearson’s $\chi^2 = 2.03$, df 6, $P = 0.917$ for SNP pair hCV8351378-rs760317 and Pearson’s $\chi^2 = 0.091$, df 2, $P = 0.956$ for rs760317). Testing the null hypothesis (PHASE 2.0) for the SNP pair hCV8351378-rs760317 under 10,000 permutations yielded an empirical $P$ value of 0.003. The enrichment of the A allele of rs760317 in the 2 IBD subset and in all cases was consistently observed when compared separately to each of the three subgroups of controls (data not shown). $\chi^2$ tests based on haplotypes delineated by three adjacent SNPs revealed that the association is defined by hCV8351378, rs760317, and rs722070, which collectively spanned D3S1234 (data not shown).

**Discussion**

Several previous investigations have suggested the involvement of recessive or X-linked loci with high lifetime risks for prostate cancer (33–37). All reported a higher risk for men with an affected brother than for men with an affected father; that is, the families analyzed tended to exhibit horizontal transmission, a major characteristic of recessive or X-linked traits (38). In the current study, families were ascertained with at least one CaP brother pair. Only 19.7% reported an affected father in the 207 families we collected. In the multiple-affected group, in which 38 families reported three or more affected brothers, a slightly smaller proportion (15.8%) reported an affected father. Had these been solely dominant inheritance, at least one parent would carry the dominant allele and we would have expected at least 50% of the fathers to be affected. Using this cohort, we localized a recessive candidate for prostate cancer susceptibility to a chromosome 3 region bearing the $FHIT$ gene. Although the search was initiated on 80 candidate genes, the final evidence of linkage ($P = 0.00001$) for the $FHIT$ gene exceeded the stringent threshold of genome-wide significance ($P = 0.000022$) proposed by Lander and Kruglyak (39). A subsequent association study using 16 SNPs extending over 381 kb around the LOD maximum identified a single SNP and haplotype that were associated with disease status. The minimum
### Table 3. SNP association tests in the FHIT region

<table>
<thead>
<tr>
<th>Marker name</th>
<th>Distance (kb) to next SNP</th>
<th>Marker location</th>
<th>Minor allele frequency/ N in cases</th>
<th>Minor allele frequency/ N in controls</th>
<th>$\chi^2$ Test for single SNPs</th>
<th>$\chi^2$ Test for pairwise haplotypes</th>
<th>LD measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All cases (202)</td>
<td>2 IBD cases (75)</td>
<td>All cases (202)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\chi^2$ $P$ $\chi^2$ $P$ $\chi^2$ (df) $P$</td>
<td>$\chi^2$ (df) $P$ $\Delta^2$ $D^\prime$</td>
<td></td>
</tr>
<tr>
<td>rs612759</td>
<td>45</td>
<td>FHIT intron 8</td>
<td>0.482/G</td>
<td>0.486/G</td>
<td>0.01 0.93 0.08 0.78</td>
<td>3.47 (3) 0.33 2.21 (3) 0.53</td>
<td>0.030 0.432</td>
</tr>
<tr>
<td>rs294457</td>
<td>69</td>
<td>FHIT intron 8</td>
<td>0.143/T</td>
<td>0.121/T</td>
<td>0.62 0.43 0.75 0.39</td>
<td>1.57 (3) 0.67 0.85 (3) 0.84</td>
<td>0.005 0.106</td>
</tr>
<tr>
<td>rs802774</td>
<td>24</td>
<td>FHIT intron 7</td>
<td>0.273/A</td>
<td>0.268/A</td>
<td>0.02 0.88 0.00 1.00</td>
<td>5.26 (2) 0.072 3.47 (2) 0.18</td>
<td>0.358 0.873</td>
</tr>
<tr>
<td>rs810615</td>
<td>45</td>
<td>FHIT intron 7</td>
<td>0.419/C</td>
<td>0.479/C</td>
<td>2.39 0.12 1.78 0.18</td>
<td>4.75 (3) 0.19 3.89 (3) 0.27</td>
<td>0.001 0.084</td>
</tr>
<tr>
<td>rs212046</td>
<td>13</td>
<td>FHIT intron 5</td>
<td>0.179/G</td>
<td>0.163/G</td>
<td>0.27 0.60 0.83 0.36</td>
<td>3.40 (2) 0.18 1.59 (3) 0.45</td>
<td>0.049 1.000</td>
</tr>
<tr>
<td>rs212004</td>
<td>17</td>
<td>FHIT intron 5</td>
<td>0.163/A</td>
<td>0.218/A</td>
<td>3.22 0.07 1.15 0.28</td>
<td>5.54 (3) 0.14 3.43 (3) 0.33</td>
<td>0.162 0.572</td>
</tr>
<tr>
<td>rs2736778</td>
<td>16</td>
<td>FHIT intron 5</td>
<td>0.288/A</td>
<td>0.355/A</td>
<td>3.33 0.07 2.45 0.12</td>
<td>7.97 (3) 0.047 7.69 (3) 0.053</td>
<td>0.011 0.104</td>
</tr>
<tr>
<td>hCV8351378</td>
<td>16</td>
<td>FHIT intron 5</td>
<td>0.300/C</td>
<td>0.350/C</td>
<td>0.81 0.37 0.34 0.56</td>
<td>13.10 (3) 0.0044 15.84 (3) 0.0012</td>
<td>0.142 0.543</td>
</tr>
<tr>
<td>rs760317</td>
<td>13</td>
<td>FHIT intron 5</td>
<td>0.490/G</td>
<td>0.427/A</td>
<td>4.64 0.03 8.54 0.0035</td>
<td>5.19 (2) 0.075 8.44 (2) 0.015</td>
<td>0.745 1.000</td>
</tr>
<tr>
<td>D3S1234</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs722070</td>
<td>7</td>
<td>FHIT intron 5</td>
<td>0.433/A</td>
<td>0.482/A</td>
<td>1.54 0.21 3.53 0.060</td>
<td>2.05 (2) 0.36 3.79 (2) 0.15</td>
<td>0.017 0.556</td>
</tr>
<tr>
<td>rs2361339</td>
<td>23</td>
<td>FHIT intron 5</td>
<td>0.0718/T</td>
<td>0.0522/T</td>
<td>1.02 0.31 1.27 0.26</td>
<td>2.03 (2) 0.36 1.75 (2) 0.42</td>
<td>0.048 0.627</td>
</tr>
<tr>
<td>rs1040337</td>
<td>9</td>
<td>FHIT intron 5</td>
<td>0.350/C</td>
<td>0.366/C</td>
<td>0.19 0.67 0.01 0.91</td>
<td>0.39 (3) 0.94 0.84 (3) 0.84</td>
<td>0.021 0.321</td>
</tr>
<tr>
<td>rs1882904</td>
<td>34</td>
<td>FHIT intron 5</td>
<td>0.274/A</td>
<td>0.252/A</td>
<td>0.43 0.51 0.04 0.85</td>
<td>1.43 (2) 0.49 0.91 (2) 0.64</td>
<td>0.088 0.932</td>
</tr>
<tr>
<td>rs213294</td>
<td>23</td>
<td>FHIT intron 5</td>
<td>0.239/T</td>
<td>0.209/T</td>
<td>0.81 0.37 1.59 0.21</td>
<td>6.17 (2) 0.1 6.12 (2) 0.11</td>
<td>0.330 0.790</td>
</tr>
<tr>
<td>rs213408</td>
<td>27</td>
<td>FHIT intron 5</td>
<td>0.322/A</td>
<td>0.369/A</td>
<td>1.64 0.20 0.41 0.52</td>
<td>3.40 (3) 0.33 2.80 (3) 0.43</td>
<td>0.017 0.144</td>
</tr>
</tbody>
</table>

Abbreviation: LD, linkage disequilibrium.

*Four haplotypes detected, 3 df; three haplotypes detected, 2 df.
P value of a single SNP association at 0.0035 was significant after a conservative Bonferroni correction (0.0035 × 16 = 0.056) for multiple testing. Considering several SNPs tested displayed certain degrees of linkage disequilibrium, the total number of independent SNP would decrease to ~16.

The chromosome 3 region bearing the FHIT gene has not been reported in previous genome-wide linkage scans, probably for a variety of reasons. Most previous studies used hereditary prostate cancer families that ascertained families with three or more cases among first- or second-degree relatives (40–43), resulting in a tendency toward vertical transmission, with a higher probability of fathers being affected—a major characteristic of dominant traits (38). Interestingly, the location of a linkage signal at ~80 cM on chromosome 3 reported in the current study corresponds to smaller peaks in the same region in genome-wide scans that were based on families ascertained in a similar way to ours (31, 44). Minor peaks in the same region are also evident in one genome-wide scan based on hereditary prostate cancer families (43). Our stronger linkage signal was likely the result of location of markers quite close to the candidate region, a consequence of the candidate gene approach we used, together with the probable reduction of locus heterogeneity achieved by testing linkage in the subset of multiple-affected siblings.

Although the linkage signal was elevated significantly for a subset of families that reported three or more affected brothers, it was not restricted to this subset (data not shown). Subsequent association tests also suggested the occurrence of homozygotes of the putative risk haplotype for a number of individuals outside the multiple-affected subset. In our cohort, nearly half the families did not report information on additional siblings, and 14% reported no more than two siblings. These families were not included in the subset. A higher rate of unawareness of cancer incidence among male first-degree relatives of probands may also be a factor (18).

Both model-free analysis using LOPAL and model-based analysis using GENEHUNTER yielded a maximum peak at D3S1234 (Fig. 2B and C) on the assumption of recessive inheritance. Similarly, analysis with these programs assuming a dominant model yielded smaller peak maxima at CDC25a2. The location of maximum sharing of 2 alleles IBD correlated with that of minimum sharing of 1 allele IBD and with the LOD score maximum of LOPAL. Thus, our IBD sharing distribution data point to a recessive locus centered on D3S1234, but the possibility remains that an additional dominant locus resides near CDC25A.

Due to the complex nature of human diseases, different programs available for linkage analyses may deal with certain problems, such as missing data, conflicting data, large and extended family data, better than others. Each program may have different assumptions on the mode of inheritance, use distinct algorithms to calculate IBD sharing status, and assess significance with different statistics (45). As a result, these programs can produce different linkage locations or these magnitude of LOD scores. Inasmuch as MERLIN and GENEHUNTER calculate the same NPL score, we only reported the result from MERLIN. LOPAL and MERLIN use different methods of analysis that have their best power against different alternatives, and it is not surprising for the two programs to yield distinct linkage peaks that were 11 cM apart. We chose first to focus our analysis on the D3S1234 signal, but we are currently beginning to construct SNP-based linkage disequilibrium blocks extending from the CDC25A peak marker, CDC25a2, to determine if one or more risk haplotypes may be identified there and if inheritance of the risk alleles there is independent of FHIT.

The controls we used in the current study were not age-matched men without prostate cancer. We attempted to estimate allele (haplotype) frequencies in individuals without prostate cancer from the same ethnic population to compare them with our CaP cases. The fact that women and underaged men were included in two of the control subgroups implies that risk alleles (haplotypes) may be present in our controls at a higher frequency than in age-matched men without CaP, because women cannot develop the disease and younger men may not be old enough to develop the disease despite being homozygous for risk allele(s). This would have biased our finding toward the null hypothesis. Although the consistency of genotype and haplotype frequencies we observed among the three control subgroups suggested their homogeneity, additional tests in an independent set of age, ethnicity, and gender-matched cases and healthy controls will be required to replicate our observations.

With the SNPs described in Table 3, we detected association closely localized to, and surrounding, the D3S1234 marker. Significant association was detected for the single SNP, rs760317. Association was also observed to a lesser degree for an adjacent SNP, rs722070, showing significant linkage disequilibrium with rs760317. A stronger correlation was revealed through haplotype analyses, identifying haplotype A-A of SNPs hCV8351378-rs760317 that was significantly enriched in cases versus controls (Table 3; \( \chi^2 = 15.84, df = 3, P = 0.0012 \)). The haplotype association with disease status decreased significantly for the adjacent SNP pair rs760317-rs722070, although these two SNPs display significant linkage disequilibrium. These observations suggest the existence of additional SNPs in the vicinity that may be more strongly associated with the disease than rs760317. Other pairs of SNPs displaying linkage disequilibrium (e.g., rs802774-rs810615) showed no significant disease association. Our association seems to extend over a broader region with haplotypes than with single SNPs, consistent with a previous conclusion that haplotypes may be used to screen for associations initially (46). Completing our linkage disequilibrium mapping of the region around D3S1234 will require a much higher density of SNPs than is available in current public databases because of a much higher local recombination rate in this region (2.6 cM/Mb) than the genome-wide average (~1 cM/Mb). We are currently conducting extensive resequencing in the region to acquire additional markers and investigate detailed linkage disequilibrium structure.

FHIT is composed of 10 short exons spanning a ~1.5-Mb genomic interval and encoding a small 16.8-kDa peptide involved in nucleoside binding (47). Because our linkage and preliminary association studies have located the presumed disease locus to intron 5, a mechanistic basis for our result is not evident. For example, FHIT resides at the FRAXB fragile site of 3p14.2 and is one of the most frequently deleted regions in multiple cancers (48). Yet none of the previously identified landmarks characteristic of the fragile region, such as aphidicolin-induced hybrid breaks, HPV16 integration sites, pSV2neo integration sites, and deletion end points in cancer cell lines, overlaps with the region defined in this study. In this regard, however, it is worth noting that although FHIT expression is absent or significantly reduced in many types of cancer (including prostate cancer; ref. 47, 49), usually, as noted above, allelic losses of large regions bearing this gene have rarely been observed in prostate cancer. Whereas several exons apparently unrelated to FHIT have been predicted within the

Cancer Res 2005; 65: (3). February 1, 2005 812 www.aacrjournals.org

Downloaded from cancerres.aacrjournals.org on January 6, 2018. © 2005 American Association for Cancer Research.
boundary defined by SNPs rs2736778 and rs213294 using GeneScan and Grail, none of these correspond to conserved segments that have been identified among humans, mice, or rats. Thus, there is no clear evidence for new genes within our candidate interval. It is possible that although the intronic position we described may not lie within canonical splice recognition signals, disease alleles may nonetheless alter the splicing pattern, leading to an aberrantly spliced gene product, such as the phenomenon observed for a mutation residing deep within intron 2 of CDKN2A (50). In recent years, there has also been accumulating evidence indicating conserved intronic sequences playing a regulatory role in gene expression. In any event, it is clear that further elucidation of a disease mechanism must await sequence characterization of disease alleles.

Finally, another notable outcome of our study was the finding that although a FHIT linkage signal was present in the analysis of all primary pairs, the signal was considerably enhanced in the 66 ASPs in 38 families chosen for multiple-affected brothers. Although the signal strength was partly attributable to the likely recessive mode of inheritance, there was also a significant contribution from reduction of locus heterogeneity by stratifying on that phenotype. We are currently evaluating two independent linkage signals, each obtained in a phenotypic subset of prostate cancer siblings: with higher Gleason scores or younger age at diagnosis. Our findings echo those of Wiesner et al. (51) in which siblings characterized by disease diagnosis at ≤65 with colon cancer or advanced colon adenomas >1 cm in size, or those who showed high-grade dysplasia, showed linkage to 9q22.2-2.1. Thus, when phenotypic characterization is successfully applied, smaller numbers of affected siblings may provide robust identification of loci important to the development of common adult cancers in a substantial proportion of cases.

Electronic Database Information

URLs for data presented herein as follows:

- Center for Medical Genetics, http://research.marshfieldclinic.org/genetics/
- DeCode Genetic Map, http://www.nature.com/ng/journal/v31/n3/suppinfo/ngv17_s1.html
- Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/OMIM; (CaP MIM 176807; FHIT, MIM 601153, CDDC25a, MIM 116947)
- Human Genome Browser Gateway, http://genome.ucsc.edu/cgi-bin/hgGateway

Acknowledgments


Grant support: National Institute on Aging grant AG15720, Department of Defense grant PC020680, the National Institute of General Medical Sciences grant GM28356, USPHS resource grant RB03565 from the National Center for Research Resources, and funds from the Beckman Research Institute of the City of Hope.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank all subjects for their participation, Mary Booth for help in initiating this study, and Dr. Robert Comis, Eastern Cooperative Oncology Group chair, for support in establishing this study.

References

31. Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM. Model-free linkage analysis with covariates


Genetic Linkage of Prostate Cancer Risk to the Chromosome 3 Region Bearing \textit{FHIT}

Garry P. Larson, Yan Ding, Li S-C. Cheng, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/65/3/805

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2005/02/18/65.3.805.DC1

Cited articles
This article cites 47 articles, 10 of which you can access for free at:
http://cancerres.aacrjournals.org/content/65/3/805.full#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/65/3/805.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/65/3/805.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.