AD_____

Award Number: DAMD17-03-1-0255

TITLE: Mapping Interactive Cancer Susceptibility Genes in Prostate Cancer

PRINCIPAL INVESTIGATOR: Theodore G. Krontiris, M.D., Ph.D.
Garry P. Larson, Ph.D.
Yan Ding, Ph.D.

CONTRACTING ORGANIZATION: Beckman Research Institute of the City of Hope
Duarte, CA 91010

REPORT DATE: April 2007

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-04-2007	Final	1 Apr 2003– 31 Mar 2007

4. TITLE AND SUBTITLE

Mapping Interactive Cancer Susceptibility Genes in Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-03-1-0255

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Theodore G. Krontiris, M.D., Ph.D., Garry P. Larson, Ph.D., Yan Ding, Ph.D.

Email: tkrontiris@coh.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Beckman Research Institute of the City of Hope
Duarte, CA 91010

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We have employed a large prostate cancer affected sibling pair cohort for candidate gene based linkage analyses. In this work we sought to enlarge a pre-existing cohort of CaP (Prostate Cancer) ASP with continued institutional recruitment of brothers affected with disease. We performed candidate gene based fine structure linkage analysis on approximately 2 dozen genes previously implicated in CaP risk. We also tested gene x gene interactions with a new paradigm based upon allele sharing enrichment. Our major finding was the localization of a susceptibility locus to intron 5 of the FHIT gene. By utilizing a combination of extensive mutation/single nucleotide polymorphism (SNP) discovery efforts in select disease cases in conjunction with linkage disequilibrium (LD) mapping and association testing we identified a SNP, rs760317, showing strong association with disease in affected brothers sharing 2 alleles identify by descent (IBD). The findings were published in 2005 and have recently been replicated by independent researchers in both a family-based Caucasian patient cohort and an African American patient cohort. Our efforts represent a significant accomplishment in the identification of a new gene associated with CaP risk as quite often promising initial linkage or association results fail to be replicated in independent studies. We continue our efforts today with the hope of finding the causative allele(s) in FHIT and it/their possible function using population genetic tools. This represents extreme challenges as the mechanistic basis for how disease allelesresiding deep within the intronscontribute to disease risk.

15. SUBJECT TERMS
Genetics, prostate cancer, affected sib pair, linkage, candidate gene, linkage disequilibrium, allelic association, cancer risk, disease susceptibility, SNP, haplotype, microsatellite, Identity by Descent, Identity by State, single nucleotide polymorphism.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	33	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

The quest to identify inherited risk alleles in genes that increase a man's chances of prostate cancer (CaP) have been difficult though there is a strong inherited component to this disease. A powerful approach to identifying these disease alleles is to use affected sibling pairs (ASP) where both brothers are affected with disease. The analysis is based on a very simple proposition that ASP that inherit disease-causing alleles at a given locus will share these alleles more often than chance alone. This project deals with collecting ASP with CaP through a collaboration with the Department of Urologic Oncology at the City of Hope National Medical Center (Dr. Mark Kawachi) to add to a pre-existing cohort of CaP ASP patients (Aim 1). Additionally, we are attempting to test for linkages in approximately two-dozen candidate genes previously implicated in CaP pathogenesis from published reports (Aim 2). We also sought to develop strategies that enrich for the likelihood of finding disease alleles by hypothesized gene-gene interactions (Aim 3). Our test utilized the joint sharing distribution of an important cell cycle gene (*CDKN1A*) and a transcription factor (*TP53*) that activates this gene. Finally, we continued the characterization of one promising linkage signal proposed in the original application by more narrowly defining the linkage interval in the FHIT gene. We describe a combination of linkage disequilibrium (LD) and association studies in an effort to identify disease alleles in this gene. This has resulted in the publication of one manuscript describing our findings at the FHIT locus (*Ca Res* 65:805-814, 2005). We continue to narrow the disease interval through a combination of single nucleotide polymorphism (SNP) discovery efforts (mutation detection), LD mapping and association studies. This provides many challenges as the target region resides deep within a large intron of the *FHIT* gene. Our efforts focused on a 28.5 kb interval within intron 5 of *FHIT*. Since non-exonic causative mutations are difficult to identify, we employed an approach looking for signatures of natural selection in this region within human populations to better understand the potential nature of any disease mutation(s). Since non-exonic causative mutations are difficult to identify, we are employing an approach looking for signatures of natural selection in this region within human populations to better understand the potential nature of any disease mutation(s). Thus, a detailed resequencing survey in Europeans, Africans, Japanese, and several non-human primates was conducted (Aims 4 & 5a). We have refined the region associated with prostate cancer risk to a 9-kb LD block and discovered a strong signature of selection in multiple human populations and other primate species. This suggests the existence of functionally important elements within the intronic sequences analyzed. Recently our findings of an association of CaP with SNP rs760317 was replicated in a large independent case-control setting (*Ca Epi Biomrkrs Prev* 16(6):1-4, 2007) thereby supporting our findings from this project. Our approach illustrates the continued usefulness of linkage studies in identifying disease susceptibility genes and the difficulties involved in elucidating disease alleles in non-coding regions of the genome.

Task 1(A & B): Recruit unaffected siblings from our prexisting families and new CaP ASP to add to our pre-existing cohort.

Task 1a - Our collaborator, Dr. Kawachi, Department of Urologic Oncology, and Clinical Research Associates (CRA) actively recruited CaP probands into the study. We attempted to inform prospective patients with a poster in the clinic and informational pamphlets (IRB approved) describing the study. We also provided informational articles about our study to patient support organizations (Prostate Cancer Research Institute, Los Angeles, CA) in an attempt to 'reach-out' to potential patients that may be distant from the City of Hope. The recruitment of siblings proved to be ineffective primarily based upon protocol modifications by the US Army Medical Research and Materiel Command Human Subjects Research Review Board (HSRRB) and our Institutional Review Board (IRB #02175). In prior patient recruitment projects of similar nature we were given permission to directly contact the sibling to explain the purpose of our study and attempt to recruit. However, in the current study, recruitment of both brothers with CaP to form an affected sibling pair component has been compromised by our inability to directly communicate with the CaP sibling-instead relying on the proband to convey the information. This has drastically compromised our ability to effectively recruit new families. Relying on the proband/index case (identified in the Department of Urology) to communicate information regarding this study to his affected brother and subsequently have that brother contact us was largely ineffective. Upon study completion we had collected 8 complete affected sibling pairs and 10 index cases where we still await the brother's sample. We have collected a third sibling in one case. In total, we have 42 individuals in the study, including unaffecteds. This falls short of our initial recruitment goals of 100 CaP ASP and is, in itself, a large disappointment. The completed ASP have been integrated into our genotyping flow after being subjected to whole genome amplification (WGA) to boost DNA amounts (Holbrook, Stabley et al. 2005),

Our protocol amendments were designed to boost recruitment numbers and proposed contacting the index case with a strategy to collect buccal cells from his saliva sample and saliva from his affected brother. The index case then forwards a similar kit containing the saliva collection sampler, consent forms, and family history questionnaire to his sibling in a prepaid mailer. DNAs were later prepared from the saliva samples. It was the hope that this recruitment strategy would increase the number of participants since neither the proband or sibling need visit the hospital for sampling. In addition a family history questionnaire could be filled out in the privacy of their home. We have abandoned the recruitment of unaffected siblings from sib pairs previously recruited from our Eastern Cooperative Oncology Group (ECOG) study since it has been determined it is too difficult to communicate with these potential participants while abiding by the IRB and HSRRB approved patient recruitment protocols..

Task 2: Fine-structure linkage analysis with multiple physically close markers in approximately two dozen candidate genes relevant in CaP.

Tasks 2a-e - Most of the preliminary linkage analyses were reported in Table 1 of the 2005 Annual Progress Report. Of note male siblings had previously been screened with the Y-chromosome marker *DYS413* (het 0.71) since brothers must share a common Y chromosome. Two additional Y-chromosome markers (*DYS385* and *DYS389*, hets 0.79 and 0.70), both duplicated on the Y-chromosome, identified 3 additional sibling pairs not sharing a common paternity (Thomas, Bradman et al. 1999) (Butler, Schoske et al. 2002). These pairs were removed from further analyses. The identified non-shared paternity rate is approximately 2-3% in our patient population.

Task 2f – Table 1 gives linkage results for our candidate genes. In our 2005 Annual Report we realize we listed Identity by State (IBS) sharing statistics (Annual Report Table 1). We have calculated the Identify by Descent (IBD) mean sharing statistics, a much more powerful statistic to detect linkage with the SIBPAL component of the statistical genetics package S.A.G.E. 5.3 (Elston 2006). These data for all ASP are presented in Table 1. We elected the means statistic (signified by "π") for sharing as this is the most sensitive to detect linkage in the absence of a genetic model (Blackwelder and Elston 1985). As with *FHIT*, we stratified our ASP by clinical co-variates such as: family history of disease (≥3 affected siblings), and combined Gleason Score. Those markers showing significant evidence of single point linkage ($p<0.05$) (**, asterisks in Table 1). Four markers (and thus their associated genes) show excess sharing ($H_o=0.5$, $H_A>0.5$): D17S1353 (*TP53*), D17S947 (*ELAC2*), D17S1147 (*HSD17β1*), and D17S1322 (*BRCA1*). To rule out artifacts from multiple testing, we performed multi-point analyses with additional markers (listed in Table 1). The tumor suppressor gene *TP53* survived a 3-point analysis (D17S1353 and P53_VNTR) (mean sharing (π) =0.538, p=0.046). Germline p53 mutations have been identified in cancer predisposition syndromes such as Li-Fraumeni (Evans, Mims et al. 1998). It is reasonable that germline mutations reside in *TP53* that influence CaP risk and this represents a promising lead for future research.

Task 2g– All CaP ASP were genotyped; however, we were unable to genotype unaffected brothers due to issues surrounding patient recruitment (see Task 1 above). In addition, we discovered that only ~5% of CaP ASP families had a 3rd sib (brother) available for sampling. The purpose of genotyping unaffected brothers is to compare allele sharing between concordant sibs (both sibs affected) versus discordant sibs (1 affected and 1 unaffected). The comparison of allele sharing between concordant versus discordant sibs allows one to identify areas of excess sharing due to transmission distortion (ie-evidence of linkage due to causes other than the phenotype for which the patient was ascertained) (Zollner, Wen et al. 2004). When concordant and discordant sibs demonstrate the same sharing across an interval, these areas are much less likely to harbor susceptibility genes (Wiesner, Daley et al. 2003). In the absence of unaffecteds sibs we routinely interrogated our candidate gene intervals by examining publicly available genotype data for the CEPH families (http://www.cephb.fr/cephdb/php/eng/index.php). Each CEPH family has a large pedigree of minimally 10 children. Though this represents a small number of <10 sibships (families) it identifies areas of concern for our linkage analysis where excess allele sharing is observed. We did not observe excess sharing across intervals showing significance in single-point linkages.

Task 3: Employ a marker-guided strategy for the discovery of risk alleles and potential gene-gene interactions of candidates noted in Task 2 above.

We have already detailed in our 2005 Annual Report (Fig. 2, Table 2) a preliminary gene-gene interaction test which we call *DABLS* (*D*isease *A*ssociation *b*y *L*ocus *S*tratification). DABLS relies on partitioning a select group of ASP by allele sharing enrichment with microsatellite markers to generate 9 compartments much like a tic-tac-toe pattern. We hypothesize that probands will be enriched for low-frequency, disease-causing haplotype variants, possibly in both genes, compared to the entire sample population.

Task 3a/b– Our goal was to screen for interactions between and transcription factor and its downstream target. With this test we explored transcriptional interactions between *CDKN1A* (6p21) and a transcriptional activator *TP53* (17p13). Two binding sites for the TP53p tumor suppressor transcriptional activator reside *CDKN1A* upstream region (Chin, Momand et al. 1997) in conjunction with additional *cis*-acting elements that are responsive to RAS, TGFβ Vitamin D Receptor, various STAT proteins, and C/EBPα (Roninson 2002).

Task 3c– We had previously determined the haplotype spectrum in *CDKN1A* in breast cancer patients and defined 10 haplotypes with these 9 SNPs that span *CDKN1A*. As described in the 2005 Annual Report (Task 3), we utilized multiplex SNP genotyping on all CaP Index cases from our ASP cohort. Unfortunately we did not observe any significant differences in the distribution of haplotypes when we compared the 2 x 2 Target Group to the All Index Cases (χ^2 test 9 df, not significant). Though unsuccessful in our initial attempt we continue to examine this approach with other gene-gene interactions.

Task 4: Conduct linkage disequilibrium analysis to identify genes and haplotypes that are responsible for PCa in *CDC25a/FHIT* and *CDC2* and any genes demonstrating positive results from Aim 2.

Task 4b/d –New short tandem repeat (STR) and SNP markers for the *FHIT* interval were reported in Annual Reports for 2004 and 2006 respectively. These included the fine structure STR linkage markers in and around *FHIT* (reported in Appendix 1, Table 2), along with known and newly-defined SNPs from this work (2006 Annual Report Supporting Data)

Task 4c – We were unable to recruit unaffected individuals from these families due to IRB/HSRB protocol restrictions. Reference DNAs from the HapMap Reference Panel (http://www.hapmap.org/) along with various primate DNAs from either the Coriell Repository (http://ccr.coriell.org/nigms/) or the Center for the Reproduction of Endangered Species at the San Diego Zoo were described in the 2006 Annual Report Task 5.

We pursued the refinement of the linkage signal at *FHIT*, and conducted linkage disequilibrium (LD) analysis and association tests within intron 5 of *FHIT* based on resequencing data from effort detailed in Task 5. These works resulted in one publication in Cancer Research and a manuscript in preparation. To briefly summarize our findings, linkage analysis identified an interval showing excess sharing highlighting intron 5 of FHIT gene on chromosome 3 (Fig.1 in manuscript Larson, et al. Ca Res. 65:805-14). Initial association tests were performed with 16 single nucleotide polymorphisms (SNPs) in this interval and revealed maximum signal at SNP rs760317 within a 28.5 kb region bracketed by two SNPs, hCV8351378 and rs722070 (Table 3 in manuscript Larson, et al. Ca Res. 65:805-14). LD measurements (Table 3 in manuscript Larson, et al. *Ca Res*. 65:805-14) suggested the need to examine the area at a higher resolution with additional SNPs to define the risk interval. We therefore extensively sequenced the 28.5 kb interval (Task 5) and characterized local LD structure (Fig. 1 & 2 in 2006 Annual Report). Additional association tests were performed with SNPs capturing most of the LD information. Significant association (cutoff $p = 0.05$) was detected for multiple SNPs within a 24 kb interval and maximized at SNP rs760317 (Pearson's $\chi^2 = 9.12$, df 1, $p = 0.003$) (Fig. 3 in 2006 Annual Report).

Recently, the association of rs760317 to CaP risk has been confirmed in two independent sample sets, one family-based Caucasian samples (434 with and 383 without prostate cancer) and another unrelated cases and controls of African Americans (133 with and 342 without prostate cancer), by another group of researchers (Levin and Cooney 2007) utilizing our initial findings (Appendix 2). We have included a copy of their soon to be published manuscript in June 2007 since it represents a validation of our efforts. During their study, we collaborated with Michigan based group by exchanging anonymous DNA samples to control genotyping errors and discussed and shared data from our ongoing investigations of *FHIT*.

To search for potential risk alleles across the 1.5 Mb region of *FHIT* gene, we genotyped three additional SNPs exhibiting low p-values in a large scale genome wide association study on CaP (Cancer Genetic Markers of Susceptibility Study, CGEMS Prostate Ca WGAS Phase 1A) (http://cgems.cancer.gov/index.asp). These genome-wide datasets examining 550,000 SNPs for 1172 cases and

1157 controls of European origin have identified SNPs associated with disease risk (Yeager, Orr et al. 2007). In addition, we elected to screen 19 of the top 200 scoring SNPs from the CGEMS project and another 4 SNPs covering 3 candidates genes (originally proposed in a prior NIH grant) that also scored high in the genome-wide association test (Table 2, candidate genes denoted with asterisk,*). One SNP within *FHIT*, rs6779755, showed evidence of association to disease risk (p=0.014 comparing allele counts and p=0.055 comparing genotype counts). Similarly, one of the 4 SNPs, rs2295348 for *CDC25B*, covering one of our candidate genes also generated significant p-values (0.035 comparing alleles and 0.093 comparing genotypes). In contrast, none of the additional 17 SNPs selected from the most significant SNPs in the CGEMS project exhibited a p-value lower than 0.05 in our sample set. These results suggest additional risk alleles in *FHIT* and possibly other candidate genes.

Task 5: Conduct mutation detection in appropriate candidate genes among individuals identified in Aims 3 & 4

Our 2006 Annual Report detailed the resequencing effort that provided data to investigate local LD structure and natural selection within the 28.5 kb interval in human populations. We analyzed the resequencing data and detected strong signatures of natural selection in the European American (Fig. 4 in 2006 Annual Report) and Japanese populations, providing strong evidence for a functional role for this intronic region.

To investigate if natural selection was restricted to human populations, we also sequenced the 1 kb region of maximum selection signature in 13 unrelated common western chimpanzees and 6 bonobos. These data revealed potential natural selection in common western chimpanzee and bonobos. Although the common chimpanzee possessed a completely different collection of SNPs compared to the human, their haplotype distribution exhibited a pattern similar to that of the Japanese: predominantly one haplotype with extremely high frequencies of the derived allele for multiple SNPs (Tajima's D = -1.81, FuLi D = -3.02, Pi = 0.0015). A significantly high Fay & Wu's H (8.62 for 12 SNPs, p = 0.0001 assuming standard neutral model) suggested a hitchhiking effect under a recent positive selection pressure. Briefly, Tajima's D, FuLi and Fay and Wu's H statistics are population genetic parameters which measure selective pressures on nucleotide sequences. The Bonobo individuals were all homozygous for the major haplotype observed in chimpanzees with two new rare SNPs. Both of them were observed only once in the 6 individuals (Tajima's D = -1.45, FuLi D = -1.72, Pi = 0.00034). We have not observed fixed nucleotide changes within the 1 kb window between the Chimpanzee and the bonobo. This pattern is consistent with background selection.

Task 5a – All SNP discovery efforts has utilized conventional ABI based bidirectional fluorescent DNA sequencing. Details of the SNP discovery efforts within the *FHIT* gene were provided in the 2005 Annual Report.

Task 5b – We initially utilized the ABI SNaPShot assay for Single Nucleotide Polymorphism (SNP) genotyping of both our case and control patient populations (Makridakis and Reichardt 2001). This assay has limited throughput potential (up to 13 SNPs in our hands). In the final year of the program institutional acquisition of a Sequenom mass spectrometer genotyping system has facilitated higher genotyping throughput (up to 28-plex) at a reduced cost. We have therefore migrated all SNP assays to the mass spec platform. Much of the new genotype data generated over the last year of the project in Task 4 from the CaP Genetic Markers of Susceptibility (CGEMS) program was generated on this platform.

Task 5c – Our efforts to replicate any findings in our DABLS analyses (Aim 3) were hampered by our inability to robustly recruit new ASP families into the study (see Aim 1 above). This prevented us from defining replication sample sets large enough to have sufficient power for the analyses. Nonetheless, new SNPs

identified in the linkage interval of *FHIT* were tested independently by the Michigan group (discussed in Aim 4 above).

KEY RESEARCH ACCOMPLISHMENTS

- Recruitment of 8 CaP affected sibling pair families through collaboration with the Department of Urologic Oncology, City of Hope National Medical Center.

- Utilization online databases to identify newly defined microsatellite markers for CaP associated candidate genes for linkage analysis.

- Integration of high throughput SNP genotyping via mass spectroscopy (Sequenom) for patient samples.

- Identification of 203 SNPs in a 28kb interval for association testing and LD mapping. Seventy-eight of these represents newly defined SNPs not previously identified in public databases (HapMap or dbSNP).

- Publication of manuscript in Cancer Research "Genetic Linkage of Prostate Cancer Risk to the Chromosome 3 Region Bearing *FHIT*" (Ca Res 65:805, 2005). Replicated in independent study.

- Significant association detected for multiple SNPs located within a 9 Kb LD block within a refined block of intron 5 within *FHIT*.

- Initiation of gene x gene interaction testing (DABLS) for the *TP53* and *CDKN1A* genes

- Manuscript in preparation defining the population genetic analysis of the interval associated with CaP risk. (Ding, Y *et al*. Strong Signature of Natural Selection within an *FHIT* Intron Implicated in Prostate Cancer Risk)

REPORTABLE OUTCOMES

Published Scientific Articles

Larson, G., Y. Ding, et al.Genetic linkage of prostate cancer risk to the chromosome 3 region bearing FHIT. Cancer Res 65(3): 805-14, 2005.

Manuscripts in Preparation

Ding, Y *et al*. Strong Signature of Natural Selection within an *FHIT* Intron Implicated in Prostate Cancer Risk

Invited Scientific Sessions

Invited Poster Presentation, Annual Meeting of the American Society for Human Genetics, Toronto, Canada October, 2004. Sibpair linkage analyses using SNP genotypes as covariant suggests that two candidate genes 11 cM apart on chromosome 3 may independently contribute to prostate cancer risk Y. Ding, G. Larson, T.G. Krontiris, The ECOG E1Y97 Study Group Beckman Res Institute, City of Hope, Duarte CA.

Abstract We conducted single point linkage analysis of over 80 candidate genes in 402 brothers affected with prostate cancer from 201 families. Markers representing two adjacent candidate genes on chromosome 3p, *CDC25A* and *FHIT*, demonstrated suggestive evidence for linkage with identity by descent (IBD) allele-sharing statistics. Fine-structure multipoint linkage analyses were performed using LODPAL (S.A.G.E.) and MERLIN. The strongest evidence of linkage was detected for D3S1234 (located in intron 5 of *FHIT*) at 81.23 cM (maximum LOD score = 3.15, p = 0.00007) using LODPAL, and for both CDC25a2 (15 kb downstream of *CDC25A*) at 70.55 cM (NPL_{all} = 1.90, p = 0.03) and D3S1234 (NPL_{all} = 1.84, P = 0.03) using MERLIN. For a subset of 38 families in which three or more affected brothers were reported, LODPAL generated a maximum LOD of 3.83 (p = 0.00001) at D3S1234 and a secondary peak of 2.19 at CDC25a2), while MERLIN produced a maximum NPL_{all} of 2.94 (p = 0.002) at CDC25a2 and a smaller peak of 2.38 (p = 0.009) at D3S1234. We then genotyped 16 SNPs covering a 381 kb region surrounding D3S1234 and 5 SNPs spanning 148 kb region surrounding *CDC25A* on one case from each family. Using LODPAL with one-parameter model incorporating individual SNPs as covariate, we evaluated each SNP for their genotype correlation with excessive IBD sharing in all families. We found one SNP from each region with significantly increased maximum LOD scores of 5.02 and 4.72 at D3S1234 (alpha = 100) and CDC25a2 (alpha = 7), respectively. Permutation tests of random SNP genotype designation to each family assuming the same genotype frequency, missing data, and value of alpha demonstrated a p value of ~ 0.01 for the associated SNP at D3S1234 and $p < 0.001$ for the SNP at CDC25a2 to generate maximum LOD exceeding observed ones. These results suggest that both candidate genes *CDC25A* and *FHIT* may independently be involved in prostate cancer risk. They also demonstrate potential advantages using SNP genotypes as covariate to reduce heterogeneity and to pinpoint disease locus in the absence of unaffected controls.

Invited Poster Presentation, Annual Meeting of the American Society for Human Genetics, Salt Lake City, UT, October, 2005. Evidence for Balancing Selection within an FHIT Intronic Region Implicated in Prostate Cancer Author: Y. Ding, G. P. Larson, G. Rivas, L. Geller, C. Lundberg, C. Ouyang, T. G. Krontiris.

Abstract Previously, we identified a locus for prostate cancer susceptibility at D3S1234 within *FHIT* (maximum LOD = 3.17, LODPAL) using a candidate gene-based linkage approach on 228 brother pairs (200 families) affected with prostate cancer. Subsequent association tests in Americans of European descent on 16 SNPs spanning approximately 400 kb surrounding D3S1234 revealed significant evidence of association for a single SNP (Pearson's $\chi^2 = 8.54$, df = 1, p = 0.0035) within intron 5 of FHIT. Genotyping 40 tagging SNPs within a 30 kb region surrounding this SNP further delineated association of prostate cancer risk to a 10 Kb region. Population studies (13 Americans of European descent and 16 Yorubans) revealed strong signatures of balancing selection within the European population, but not within the African population. A sliding window analysis of resequencing data from individuals of European descent revealed a 13 Kb region of peaks and plateaus of Pi > 0.004 and Tajima's D > 2.0 (max. Pi = 0.0074, max. Tajima's D = 3.06, p < 0.001 under a standard neutral model). The elevated Pi and Tajima's D extends across three LD blocks, suggesting the possibility of multiple sites under selection. Decay of these D statistic elevations elsewhere suggests that population structure and past demographic events do not account for our result. Within the LD block associated with prostate cancer, the haplotype enriched in the control group is the most common haplotype in European descent (40%) compared to only 10% in the Yoruban population. In contrast, the putative risk haplotype is 28% in Americans of European descent and occurs as the most common haplotype (33%) within the Yoruban population. Our study, which suggests an important selectable function within intron 5, also represents an additional corroborative approach for gene-disease associations.

SUPPORTED PROJECT PERSONNEL

Personnel receiving pay from DAMD-03-1-0255 during the project period included:

Dr. Garry Larson, Ph.D., Division of Molecular Medicine
Dr. Yan Ding, Ph.D., Division of Molecular Medicine
Mr. Guillermo Rivas, B.S., Division of Molecular Medicine
Dr. Li Cheng, Ph.D. (left the institution after Y1), Division of Information Sciences
Mr. Virgil Gagalang, B.S. (left institution after Y1), Division of Molecular Medicine

CONCLUSIONS

Prior family-based linkage studies in CaP have utilized genome-wide scanning approaches to identify regions of interest (Schaid 2004). In contrast, our approach targeted candidate genes and/or intervals previously implicated in CaP risk. Our methodology relied on the careful selection of candidate genes via curation of extant literature followed by fine-structure linkage analysis. In an era where genome-wide association (GWA) testing is the norm with especially large affected and unaffected cohorts we feel family-based linkage analyses in rather small cohorts (~200 ASP) still provides a valuable tool to identify important genomic regions that should be explored with association testing in larger, independent patient cohorts. The identification of nearly 100 novels SNPs and insertion/deletion polymorphisms in the *FHIT* intron 5 region indicates the need for deep sequencing of previously less-explored regions of the genome. Our major accomplishment has been the identification of a putative disease locus associated with increased CaP cancer risk in families of brothers sharing 2 alleles IBD in the *FHIT* interval. Our efforts represent a significant accomplishment in the identification of a new CaP susceptibility gene. Publication of our results in *Cancer Research* in 2005 lead to sharing our data in the *FHIT* gene with an independent group at the University of Michigan (Dr, Kathleen Cooney, Department of Urology, member International Consortium for Prostate Cancer Genetics, ICPCG). Based upon our linkage guided analyses and subsequent association testing, Dr. Cooney's research team was able to validate our findings with SNP rs760317 in a family-based set of Caucasian samples and an independent African American cohort (Levin A, et al. *Ca Epid Biomrk Prev* June 2007). We feel our efforts facilitated their subsequent confirmation via association analyses and may also hold promise for African American men who are acknowledged to be at a higher risk for disease than their Caucasian counterparts. We continue the effort to identify the disease susceptibility allele(s) within *FHIT* and their possible function using population genetic tools. This represents extreme challenges as it is not intuitively obvious how these disease alleles function since they reside deep within *FHIT* intron 5. We feel that funding provided by the DOD PCRP enabled this discovery and in the future it may have applicability to multiple ethnic groups.

REFERENCES

Blackwelder, W. C. and R. C. Elston (1985). "A comparison of sib-pair linkage tests for disease susceptibility loci." Genet Epidemiol **2**(1): 85-97.

Butler, J. M., R. Schoske, et al. (2002). "A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers." Forensic Sci Int **129**(1): 10-24.

Chin, P. L., J. Momand, et al. (1997). "In vivo evidence for binding of p53 to consensus binding sites in the p21 and GADD45 genes in response to ionizing radiation." Oncogene **15**(1): 87-99.

Elston, R. C. (2006). "S.A.G.E. [2006] Statistical Analysis for Genetic Epidemiology,."

Evans, S. C., B. Mims, et al. (1998). "Exclusion of a p53 germline mutation in a classic Li-Fraumeni syndrome family." Hum Genet **102**(6): 681-6.

Holbrook, J. F., D. Stabley, et al. (2005). "Exploring whole genome amplification as a DNA recovery tool for molecular genetic studies." J Biomol Tech **16**(2): 125-33.

Levin, A., Anna M. Ray, Kimberly A. Zuhlke, Julie A. Douglas, and a. K. A. Cooney (2007). "Association between Germ line Variation in the FHIT Gene and Prostate Cancer in Caucasians and African Americans." Cancer Epidemiol Biomarkers Prev **16**(6): 1.

Makridakis, N. M. and J. K. Reichardt (2001). "Multiplex automated primer extension analysis: simultaneous genotyping of several polymorphisms." Biotechniques **31**(6): 1374-80.

Roninson, I. B. (2002). "Oncogenic functions of tumour suppressor p21(Waf1/Cip1/Sdi1): association with cell senescence and tumour-promoting activities of stromal fibroblasts." Cancer Lett **179**(1): 1-14.

Schaid, D. J. (2004). "The Complex Genetic Epidemiology of Prostate Cancer." Hum Mol Genet.

Thomas, M. G., N. Bradman, et al. (1999). "High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome." Hum Genet **105**(6): 577-81.

Wiesner, G. L., D. Daley, et al. (2003). "A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2." Proc Natl Acad Sci U S A **100**(22): 12961-5.

Yeager, M., N. Orr, et al. (2007). "Genome-wide association study of prostate cancer identifies a second risk locus at 8q24." Nat Genet **39**(5): 645-649.

Zollner, S., X. Wen, et al. (2004). "Evidence for extensive transmission distortion in the human genome." Am J Hum Genet **74**(1): 62-72.

SUPPORTING DATA

Table 1 –
Identify by Descent (IBD) linkage analyses of candidate genes using SIBPAL in S.A.G.E. 5.3. Mean sharing calculation (pi, π) and p-values listed.

Table 2 –
Association Testing of CaP ASP Probands and controls with top scoring SNPs from CGEMS study.

Table 1- Single Point IBD Linkage Analysis of Candidate Genes (S.A.G.E. 5.3, SIBPAL)

Candidate Gene	Chrom	Marker	GDB Acc. No.[a]	Het[b]	UCSC Pos.(MB) May 1, 2004	deCode Pos. (cM)	No. ASP Analyzed	SIBPAL_mean Sharing (π)	p-value
RNASEL					*179.274-179.287*				
	1	*RNaseL*	*de novo*[c]	*0.74*	*179.231*		*180*	*0.459*	*0.984*
		D1S413	199102	*0.75*	*195.352*		*168*	*0.479*	*0.853*
		D1S466	199681	*0.77*	*179.035*	*183.53*	*169*	*0.483*	*0.787*
HSD3β2	1				*119.669-119.677*				
		HSD3β2	134044	*0.67*	*119.675*		*186*	*0.487*	*0.765*
		D1S534	686478	*nd*	*119.39*		*198*	*0.500*	*0.508*
SRD5A2	2				*31.661-31.717*				
		D2S2203	607887	*0.72*	*31.518*	*55.37*	*173*	*0.490*	*0.692*
NFKB1	4				*103.779-103.895*				
		NFKB1	*nd*	*0.8*	*103.909*		*203*	*0.525*	*0.109*
		D4S3043	614211	*0.67*	*103.931*	*107.52*	*186*	*0.515*	*0.216*
hTERT	6				*1.306-1.348*				
		D5S678	200148	*0.61*	*1.418*		*158*	*0.478*	*0.834*
		D5S417	188326	*0.73*	*3.174*	*8.66*	*169*	*0.497*	*0.563*
CDKN1A	6				*36.754-36.760*				
		p21B	*de novo*	*0.81*	*36.755*		*215*	*0.523*	*0.126*
CYP3A4	7				*98.999-99.026*				
		D7S647	199496	*0.79*	*98.913*		*195*	*0.510*	*0.300*
EZH2	7				*147.961-147.982*				
		D7S688	199984	*0.84*	*147.981*		*49*	*0.478*	*0.687*
PTEN	10				*89.613-89.716*				
		D10S1765	613080	*0.85*	*89.591*	*107.92*	*189*	*0.513*	*0.257*
CYP17	10				*104.580-104.587*				
		D10S1692	608877	*0.87*	*104.579*		*162*	*0.492*	*0.640*
CDKN1B	12				*12.761-12.766*				
		D12S358	199945	*0.76*	*12.53*		*192*	*0.527*	*0.081*
		D12S1580	598965	*0.77*	*13.239*	*30.91*	*196*	*0.512*	*0.262*
VDR	12				*46.521-46.585*				
		VDRga27	*de novo*	*0.86*	*46.49*		*188*	*0.517*	*0.220*
BRCA2	13				*31.787-31.871*				
		BRCA2b	*de novo*	*0.83*	*31.651*		*180*	*0.522*	*0.149*
		BRCA2c	*de novo*	*0.85*	*31.12*		*174*	*0.524*	*0.140*
CYP19	15				*49.288-49.418*				
		CYP19	119830	*0.73*	*49.307*		*188*	*0.503*	*0.437*
		D15S220	214954	*0.57*	*49.861*	*49.94*	*138*	*0.487*	*0.767*
		D15S992	608919	*0.81*	*46.627*	*47.52*	*137*	*0.500*	*0.500*
TP53	17				*7.512-7.531*				
		D17S1353	435120	*0.89*	*7.558*		*218*	*0.546*	*0.016***
		p53_VNTR	61990	*0.6*	*7.588*		*213*	*0.505*	*0.363*
ELAC2					*12.836-12.861*				
	17	*D17S947*	199816	*0.9*	*12.747*		*196*	*0.538*	*0.048***
		D17S1803	607137	*0.81*	*12.504*	*35.32*	*151*	*0.522*	*0.161*
		D17S799	188235	*0.69*	*13.111*	*37*	*188*	*0.531*	*0.055*
HSD17β1	17				*37.957-37.960*				
		D17S1147	287521	*0.7*	*38.033*		*192*	*0.530*	*0.049***
BRCA1	17				*38.450-38.530*				
		D17S1322	375323	*0.63*	*38.465*		*64*	*0.543*	*0.052*
		D17S855	192761	*0.84*	*38.458*		*150*	*0.494*	*0.604*
TYMS	18				*0.647-0.663*				
		D18S59	188185	*0.85*	*0.636*	*1.39*	*182*	*0.497*	*0.552*
KLK3	19				*56.050-56.055*				
		D19S553	314825	*0.94*	*56.241*		*104*	*0.465*	*0.858*
AR	X				*66.546-66.727*				
		A1/A2	176283	*0.9*	*66.548*				

** *p-value<0.05*

a GDB-Genome Database Accession Number (http://www.gdb.org/)

b Het, heterozygosity

c *de novo* - newly developed candidate gene markers in this study from human genome resources

Table 2-Association Testing of SNPs from CGEMS study

SNP	UCSC Chromosome	Location (Mb)	Associated Gene	P-value in CGEMS	P-value Rank Order in CGEMS	Alelle & Genotype	Cases Count	Cases Freq	Controls Count	Controls Freq		χ^2	P-value
						C	36	0.091	25	0.089	Alelle	0.005	0.944
						T	360	0.909	255	0.911			
rs7541350	1	37860631	*LOC440580*	0.000009	6	CC	0	0.000	2	0.014	Genotype	0.175**	0.676
						CT	36	0.182	21	0.150			
						TT	162	0.818	117	0.836			
						A	288	0.783	220	0.775	Alelle	0.059	0.808
						G	80	0.217	64	0.225			
rs11118988	1	204684478	*PLXNA2*	0.000075	42	AA	115	0.625	83	0.585	Genotype	2.190	0.335
						AG	58	0.315	54	0.380			
						GG	11	0.060	5	0.035			
						A	41	0.103	36	0.131	Alelle	1.288	0.256
						C	357	0.897	238	0.869			
rs2033404	4	163179911	*FSTL5*	0.000161	31	AA	0	0.000	1	0.007	Genotype	1.133**	0.287
						AC	41	0.206	34	0.248			
						CC	158	0.794	102	0.745			
						C	359	0.902	244	0.871	Alelle	1.562	0.211
						T	39	0.098	36	0.129			
rs1440606	4	163184382	*FSTL5*	0.000161	30	CC	160	0.804	105	0.750	Genotype	1.405**	0.236
						CT	39	0.196	34	0.243			
						TT	0	0.000	1	0.007			
						C	304	0.772	199	0.711	Alelle	3.202	0.074
						T	90	0.228	81	0.289			
rs604490	6	65378410	*LOC389405*	0.000034	17	CC	119	0.604	68	0.486	Genotype	4.907	0.086
						CT	66	0.335	63	0.450			
						TT	12	0.061	9	0.064			
						C	340	0.859	254	0.888	Alelle	1.288	0.256
						T	56	0.141	32	0.112			
rs7384464	7	12261775	*LOC389465 FLJ14712*	0.000004	2	CC	149	0.753	111	0.776	Genotype	5.169	0.075
						CT	42	0.212	32	0.224			
						TT	7	0.035	0	0.000			
						A	95	0.238	72	0.252	Alelle	0.184	0.668
						G	305	0.763	214	0.748			
rs9649913	8	98455684		0.000044	21	AA	15	0.075	12	0.084	Genotype	0.167	0.920
						AG	65	0.325	48	0.336			
						GG	120	0.600	83	0.580			
						A	47	0.118	25	0.089	Alelle	1.458	0.227
						C	353	0.883	257	0.911			
rs1447295	8	128554220		0.000408	164	AA	2	0.010	3	0.021	Genotype	4.136	0.126
						AC	43	0.215	19	0.135			
						CC	155	0.775	119	0.844			
						A	46	0.115	28	0.098	Alelle	0.507	0.476
						G	354	0.885	258	0.902			
rs4242382	8	128586755		0.000112	44	AA	2	0.010	3	0.021	Genotype	2.312	0.315
						AG	42	0.210	22	0.154			
						GG	156	0.780	118	0.825			
						A	342	0.859	254	0.894	Alelle	1.850	0.174
						C	56	0.141	30	0.106			
rs7017300	8	128594450		0.000199	74	AA	147	0.739	116	0.817	Genotype	3.892	0.143
						AC	48	0.241	22	0.155			
						CC	4	0.020	4	0.028			
						A	170	0.425	138	0.483	Alelle	2.230	0.135
						G	230	0.575	148	0.517			
rs2038946	13	74019203		0.000007	9	AA	34	0.170	32	0.224	Genotype	2.325	0.313
						AG	102	0.510	74	0.517			
						GG	64	0.320	37	0.259			
						A	290	0.729	210	0.739	Alelle	0.099	0.753
						G	108	0.271	74	0.261			
rs1570555	13	75269877	*LMO7*	0.000042	13	AA	103	0.518	77	0.542	Genotype	0.264	0.876
						AG	84	0.422	56	0.394			
						GG	12	0.060	9	0.063			
						C	47	0.118	27	0.094	Alelle	0.924	0.336
						T	353	0.883	259	0.906			
rs8030745	15	71920144		0.000061	117	CC	4	0.020	0	0.000	Genotype	0.352**	0.553
						CT	39	0.195	27	0.189			
						TT	157	0.785	116	0.811			
						A	186	0.467	128	0.448	Alelle	0.262	0.609
						G	212	0.533	158	0.552			
rs1872694	16	47435132		0.000012	15	AA	41	0.206	31	0.217	Genotype	1.391	0.499
						AG	104	0.523	66	0.462			
						GG	54	0.271	46	0.322			

Table 2- Continued

						C	287	0.736	207	0.724	Alelle	0.123	0.726
						T	103	0.264	79	0.276			
rs2058005	17	66757330		0.000014	12	CC	109	0.559	73	0.510	Genotype	2.125	0.346
						CT	69	0.354	61	0.427			
						TT	17	0.087	9	0.063			
						G	291	0.739	208	0.727	Alelle	0.108	0.742
						T	103	0.261	78	0.273			
rs11077554	17	66798276		0.000009	8	GG	110	0.558	74	0.517	Genotype	1.386	0.500
						GT	71	0.360	60	0.420			
						TT	16	0.081	9	0.063			
						C	106	0.268	78	0.275	Alelle	0.041	0.840
						T	290	0.732	206	0.725			
rs4468671	17	66802264		0.000022	19	CC	17	0.086	9	0.063	Genotype	1.490	0.475
						CT	72	0.364	60	0.423			
						TT	109	0.551	73	0.514			
						A	301	0.760	202	0.706	Alelle	2.484	0.115
						G	95	0.240	84	0.294			
rs465543	19	6892867	*EMR1*	0.000076	34	AA	115	0.581	75	0.524	Genotype	3.138	0.208
						AG	71	0.359	52	0.364			
						GG	12	0.061	16	0.112			
						A	260	0.667	189	0.665	Alelle	0.001	0.975
						G	130	0.333	95	0.335			
rs6076157	20	23810844	*CST5*	0.00009	131	AA	86	0.441	63	0.444	Genotype	0.031	0.985
						AG	88	0.451	63	0.444			
						GG	21	0.108	16	0.113			
						A	343	0.871	262	0.929	Alelle	5.988	0.014
						G	51	0.129	20	0.071			
rs6779755	3	60006999	*FHIT**	0.018674	11894	AA	150	0.761	122	0.865	Genotype	5.810	0.055
						AG	43	0.218	18	0.128			
						GG	4	0.020	1	0.007			
						A	325	0.813	240	0.839	Alelle	0.816	0.366
						G	75	0.188	46	0.161			
rs2594264	3	60489776	*FHIT**	0.003449	1140	AA	132	0.660	103	0.720	Genotype	1.910	0.385
						AG	61	0.305	34	0.238			
						GG	7	0.035	6	0.042			
						A	127	0.324	90	0.324	Alelle	0.000	1.000
						G	265	0.676	188	0.676			
rs9879276	3	60928629	*FHIT**	0.000597	575	AA	28	0.143	14	0.101	Genotype	2.886	0.236
						AG	71	0.362	62	0.446			
						GG	97	0.495	63	0.453			
						C	352	0.880	259	0.906	Alelle	1.122	0.289
						T	48	0.120	27	0.094			
rs10137185	14	63845529	*ESR2**	0.003468	699	CC	154	0.770	120	0.839	Genotype	5.485	0.064
						CT	44	0.220	19	0.133			
						TT	2	0.010	4	0.028			
						C	94	0.241	65	0.227	Alelle	0.173	0.677
						T	296	0.759	221	0.773			
rs2281479	20	3710095	*20ORF28|CDC25I*	0.007316	1203	CC	15	0.077	11	0.077	Genotype	0.300	0.861
						CT	64	0.328	43	0.301			
						TT	116	0.595	89	0.622			
						A	93	0.233	87	0.304	Alelle	4.429	0.035
						G	307	0.768	199	0.696			
rs2295348	20	3733034	*CDC25B**	0.009022	2558	AA	10	0.050	11	0.077	Genotype	4.757	0.093
						AG	73	0.365	65	0.455			
						GG	117	0.585	67	0.469			
						A	26	0.065	20	0.070	Alelle	0.068	0.794
						G	372	0.935	264	0.930			
rs8116803	20	39167195	*TOP1**	0.009713	6678	AA	2	0.010	2	0.014	Genotype	0.122	0.941
						AG	22	0.111	16	0.113			
						GG	175	0.879	124	0.873			

* indicates candidate genes

** Degree of freedom 1 instead of 2 due to combined genotype counts

APPENDICES

Appendix 1 –
Larson, G. P., Y. Ding, et al. (2005). "Genetic linkage of prostate cancer risk to the chromosome 3 region bearing *FHIT*." Cancer Res 65(3): 805-14.

Appendix 2 –
Levin, A., Anna M. Ray, Kimberly A. Zuhlke, Julie A. Douglas, and a. K. A. Cooney (2007). "Association between Germ line Variation in the *FHIT* Gene and Prostate Cancer in Caucasians and African Americans." Cancer Epidemiol Biomarkers Prev 16(6): 1.

Appendix 1 -
Larson, G. P., Y. Ding, et al. (2005). "Genetic linkage of prostate cancer risk to the chromosome 3 region bearing *FHIT*." Cancer Res 65(3): 805-14.

Research Article

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

Garry P. Larson,[1] Yan Ding,[1] Li S-C. Cheng,[2] Cathryn Lundberg,[1] Virgil Gagalang,[1] Guillermo Rivas,[1] Louis Geller,[1] Jeffrey Weitzel,[1,3] Deborah MacDonald,[3] John Archambeau,[4] Jerry Slater,[4] Donna Neuberg,[5] Mary B. Daly,[6] Irene Angel,[6] Al B. Benson III,[7] Kimberly Smith,[8] John M. Kirkwood,[9] Peter J. O'Dwyer,[10] Barbara Raskay,[10] Rebecca Sutphen,[11] Rosalind Drew,[11] James A. Stewart,[12] Jae Werndli,[12] David Johnson,[13] John C. Ruckdeschel,[14] Robert C. Elston,[15] and Theodore G. Krontiris[1]

[1]Divisions of Molecular Medicine and [2]Information Sciences, Beckman Research Institute and [3]Department of Cancer Genetics, City of Hope National Medical Center, Duarte, California; [4]Department of Radiation Medicine, Loma Linda University Medical Center, Loma Linda, California; [5]Division of Biostatistics, Dana-Farber Cancer Institute, Harvard Medical School and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; [6]Department of Population Science, Fox Chase Cancer Center, Philadelphia, Pennsylvania; [7]Department of Medicine, Division of Hematology/Oncology and [8]Clinical Research Office, Robert J. Lurie Comprehensive Cancer Center, Northwestern University School of Medicine, Chicago, Illinois; [9]Department of Medicine, Division of Hematology/Oncology, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania; [10]Department of Medicine, Hematology-Oncology Division, University of Pennsylvania Cancer Center, Philadelphia, Pennsylvania; [11]Interdisciplinary Oncology Program, H. Lee Moffitt Cancer Center and Research Institute, University of South Florida, Tampa, Florida; [12]University of Wisconsin Comprehensive Cancer Center and University of Wisconsin School of Medicine, Madison, Wisconsin; [13]Division of Hematology & Oncology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee; [14]Barbara Ann Karmanos Cancer Institute, Detroit, Michigan; and [15]Department 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 A allele of single nucleotide polymorphism rs760317, located within intron 5 of *FHIT* (Pearson's χ^2 = 8.54, df = 1, P = 0.0035). Our results strongly suggest involvement of germline 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, *HPC1*), *ELAC2* (17p11, *HPC2*), 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. Polyglutamine 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).

Note: G.P. Larson and Y. Ding contributed equally to this work.

Requests for reprints: Theodore G. Krontiris, Division of Molecular Medicine, Beckman Research Institute, City of Hope National Medical Center, 1500 East Duarte Road, Duarte, CA 91010-3000. Phone: 626-359-8111 ext. 64297; Fax: 626-930-5394; E-mail: tkrontir@coh.org.

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, *BRCA2* 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 1 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 AC069207, primers GGGGTGCAGGTGGTTTG and TCCCCAGGCT-CAGGTGAT; and pCAFa, BAC AC104190, primers AATAAACCAACCC-CAAATGA and GAGGAAAGCGGAAGAAAGTT). 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

Group	No. of families analyzed	Total individuals genotyped	Age at diagnosis, mean ± SD (range)	Mean Gleason score (range)
All subjects	207	433	65.8 ± 7.5 (36-90)	6.3 (3-9)
Primary pair group	201	402	65.8 ± 7.5 (36-90)	6.3 (3-9)
All pair group	200	414	65.8 ± 7.5 (36-90)	6.3 (3-9)
Multiple-affected group	38	90	64.5 ± 6.6 (48-75)	6.3 (4-9)
Age at diagnosis <65 group	60	123	58.7 ± 4.1 (48-65)	6.4 (4-9)

Table 2. Markers used for multipoint analysis

Markers	Heterozygosity rate	Position (cM)*	UCSC position, July 2003	Comments†
D3S1317	0.706	27.68	10208658	*VHL*
D3S1335	0.767	27.94‡	10254548	*VHL*
pCAFa	0.825	40.68‡	20138241	*PCAF*
D3S1561	0.698	61.92	36444920	*MLH1*
D3S1298	0.885	62.93‡	38009388	*MLH1*
D3S2304	0.588	67.22	42775941	Multipoint
D3S3647	0.746	67.73	43539737	Multipoint
D3S2420	0.788	70.55	48028036	Multipoint
D3S3560	0.669	70.58‡	48155020	*CDC25a*
CDC25a2	0.857	70.59‡	48170150	*CDC25a*
D3S1581	0.884	70.66‡	48557869	Multipoint
D3S1588	0.807	72.68	54055293	Multipoint
D3S2408	0.697	76.58	55667768	Multipoint
D3S3048	0.592	77.38	56095168	Multipoint
D3S2402	0.792	78.91	58174295	Multipoint
D3S3553	0.912	78.96‡	58401230	Multipoint
D3S1540	0.918	79.99‡	59484073	Multipoint
D3S3577	0.725	80.10	59576704	Multipoint
D3S1234	0.692	81.23	60064809	Multipoint
D3S4103	0.831	82.01‡	60389874	*FHIT*
D3S1300	0.83	82.22	60467319	*FHIT*
D3S1481	0.839	82.58‡	60615893	Multipoint
D3S1312	0.767	85.07	62363825	Multipoint
D3S1600	0.768	86.78‡	63277480	Multipoint
D3S1287	0.646	88.25	64164382	Multipoint
D3S3584	0.666	134.26	128497626	*MCM2*
D3S3606	0.834	134.60	128521221	*MCM2*
D3S3607	0.734	135.10	128593996	*MCM2*

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

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 Genotyper 2.1 and individually verified in GeneScan 3.0. We selected SNPs with minor allele frequencies >10% in the European Caucasian population from the Applied Biosystems SNP Genotyping database and verified their positions on the July 2003 University of California at Santa Cruz (UCSC) genome build. We genotyped a total of 24 SNPs with an overall success rate greater than 95% using ABI SNaPshot. 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 in case or control populations was not observed for the remaining 22 SNPs (data not shown). We also checked the reproducibility of allele calling and found only 0.87% (7/805) of the genotypes differed between independent experiments.

Statistical Analysis

Linkage Analysis. For ASP allele-sharing data, we used three packages of programs to conduct linkage analysis: S.A.G.E. (version 4.3; ref. 21), GENEHUNTER (22), and MERLIN (23). We used the deCode genetic map (24) and integrated any marker not present on that map by interpolating its position using the physical location of the closest flanking markers of known genetic location, as well as the local recombination rate of the region 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 same 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 χ^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 χ^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 χ^2 tests with 6 degrees of freedom (*df*) in a 4×3 contingency table for neighboring pairwise haplotypes (i.e., haplotypes formed by the alleles at two neighboring SNPs), and with 2 *df* in a 2×3 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 ~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 ± 0.021, $P = 0.097$), 0.532 (SE ± 0.015, $P = 0.034$), and 0.539 (SE ± 0.021, $P = 0.065$). These three markers resided within an interval of ~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 107cM 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 (pCAFa 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. 1*A*. 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 ($NPL_{all} = 1.90$, $P = 0.03$) and D3S1234 at 81.23 cM ($NPL_{all} = 1.84$, $P = 0.03$) using MERLIN (Fig. 1*A*). 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. 1*B*). Again, we detected two linkage peaks at the two candidate genes in the multiple-affected group. LODPAL generated the maximum LOD of 3.83 ($P = 0.00001$) 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 NPL_{all} of 2.94 ($P = 0.002$) at 70.59 cM and a smaller peak of 2.38 ($P = 0.009$) 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. 2*A*), 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 α parameter equal to 1 in LODPAL; ref. 31), the maximum LOD score was 2.1 at CDC25a2. Assuming a recessive locus (α = 100), the maximum

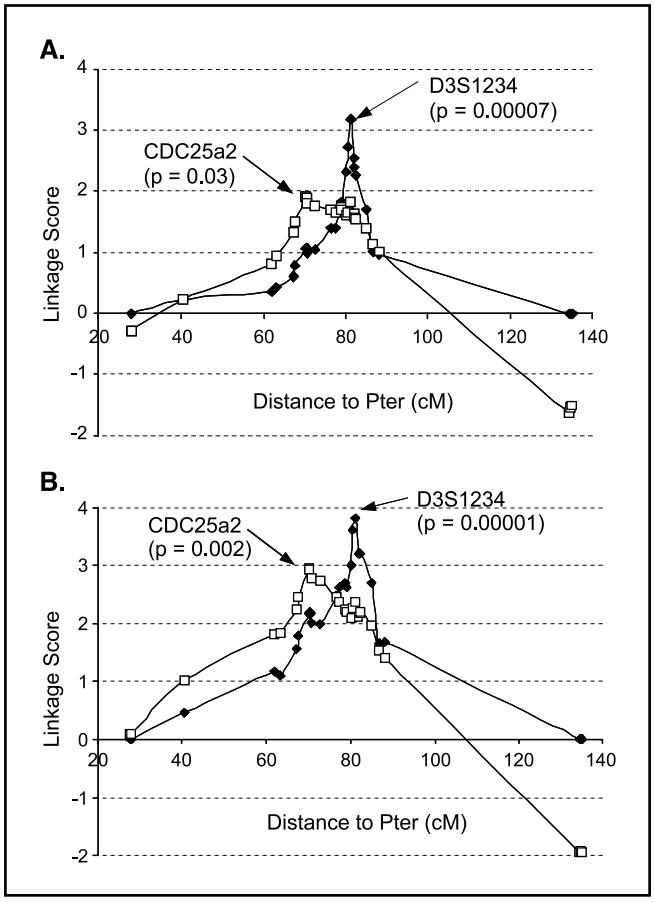


Figure 1. Multipoint model-fee linkage analyses of CaP susceptibility loci using 28 microsatellite markers (Table 2) on chromosome 3. ♦, results from LODPAL (S.A.G.E. 4.3); □, results from MERLIN. *A,* all pairs group. *B,* multiple-affected group.

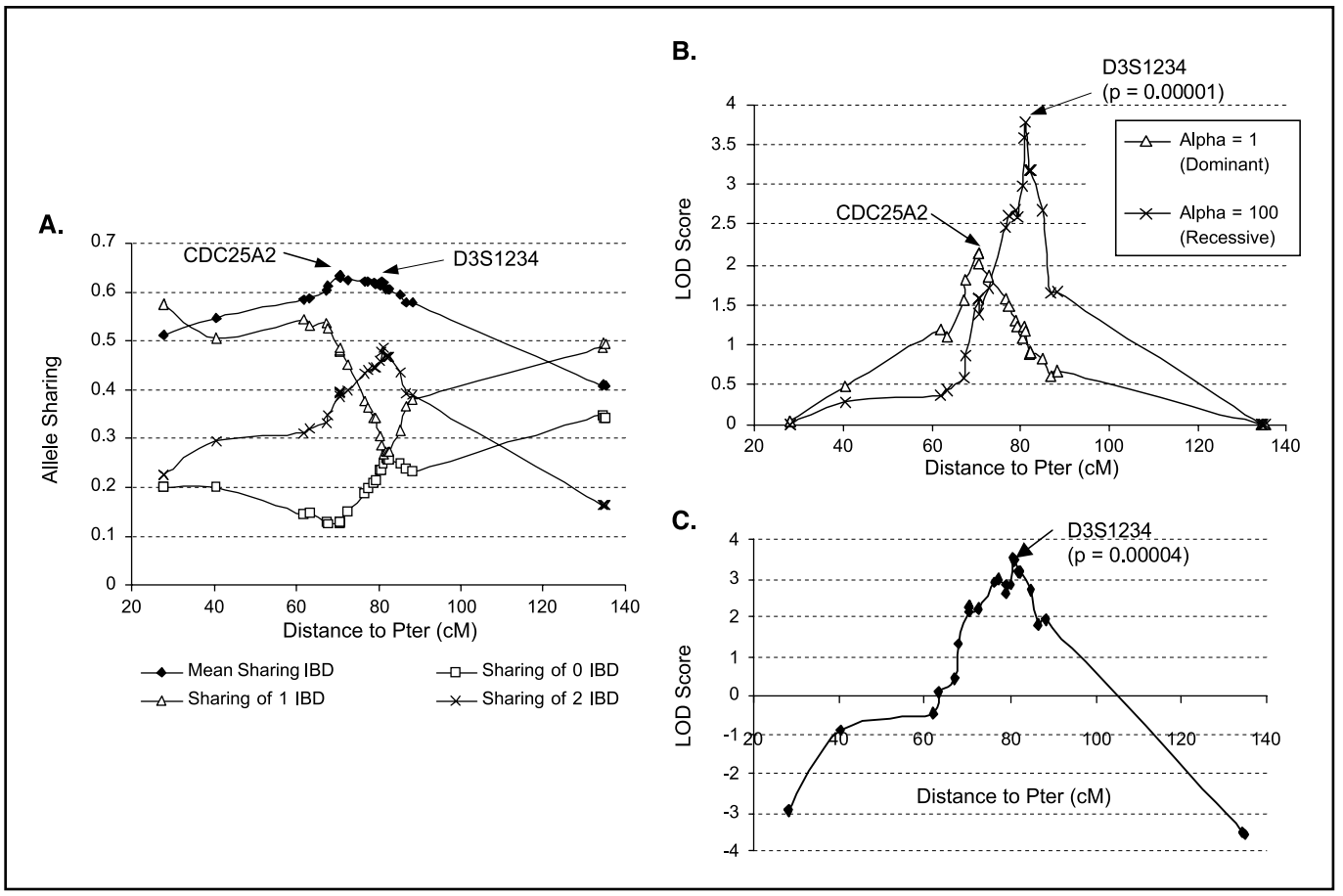


Figure 2. Testing inheritance mode in multiple-affected group. *A,* IBD distribution within the linkage interval using GENIBD (S.A.G.E.). *B,* parametric LOD score calculation using LODPAL (S.A.G.E.) with a one-parameter model. *C,* model-based LOD score calculation using GENEHUNTER under a recessive model, assuming a penetrance of 0.95 for homozygotes, phenocopy rate of 0.05, and disease allele frequency of 0.07.

LOD score was 3.7 at D3S1234 (Fig. 2*B*). 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. 2*C*). 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. 3*B* 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 3*B*). Because linkage disequilibrium was not observed in the 7-SNP set, we next selected a denser 16-SNP set encompassing D3S1234 (Fig. 3*A*). 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 Δ^2 (low) values, involving SNPs of relatively lower minor allele frequencies. Zabetian et al. (32) suggested Δ^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 χ^2 tests on frequency distributions of genotypes and haplotypes between these case-control groups. The maximum association was detected for the

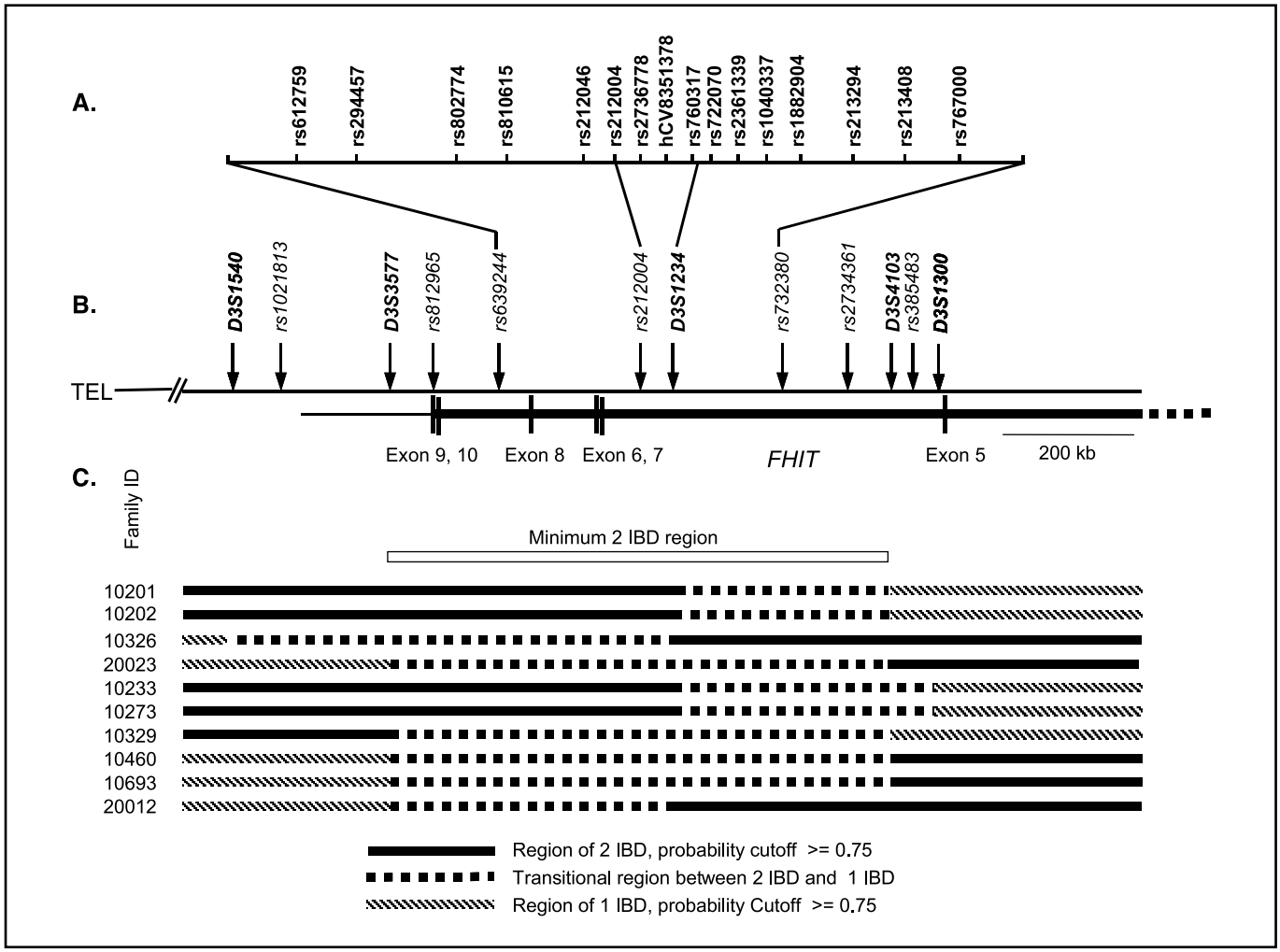


Figure 3. High-resolution marker map and inference of common 2 allele IBD region by examining key meiotic recombinants. *A* and *B,* physical map illustrating marker and *FHIT* exon locations. *Solid bar, FHIT* gene boundary; *vertical bars,* exons 5 to 10. *Bold italic font,* microsatellite markers; *bold font,* 16 SNPs used for association testing. *C,* IBD sharing distribution of selected ASPs. Patient pairs are listed to the left; *lines of various patterns,* region of IBD transition (based on sharing probability computed by GENIBD, S.A.G.E.). *Open box,* region subjected to SNP genotyping and association analyses.

SNP pair hCV8351378-rs760317 (Pearson's χ^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 χ^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 χ^2 = 2.03, *df* 6, P = 0.917 for SNP pair hCV8351378-rs760317 and Pearson's χ^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). χ^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

Marker name	Distance (kb) to next SNP	Marker location	Minor allele frequency/ *N* in cases	Minor allele frequency/ *N* in controls	χ^2 Test for single SNPs				χ^2 Test for pairwise haplotypes				LD measurement	
					All cases (202)		2 IBD cases (75)		All cases (202)		2 IBD cases (75)		Pairwise	
					χ^2	*P*	χ^2	*P*	χ^2 (*df**)	*P*	χ^2 (*df*)	*P*	Δ^2	*D′*
rs612759		*FHIT* intron 8	0.482/*G*	0.486/*G*	0.01	0.93	0.08	0.78						
	45								3.47 (3)	0.33	2.21 (3)	0.53	0.030	0.432
rs294457		*FHIT* intron 8	0.143/*T*	0.121/*T*	0.62	0.43	0.75	0.39						
	69								1.57 (3)	0.67	0.85 (3)	0.84	0.005	0.106
rs802774		*FHIT* intron 7	0.273/A	0.268/*A*	0.02	0.88	0.00	1.00						
	24								5.26 (2)	0.072	3.47 (2)	0.18	0.358	0.873
rs810615		*FHIT* intron 7	0.419/*C*	0.479/*C*	2.39	0.12	1.78	0.18						
	45								4.75 (3)	0.19	3.89 (3)	0.27	0.001	0.084
rs212046		*FHIT* intron 5	0.179/G	0.163/G	0.27	0.60	0.83	0.36						
	13								3.40 (2)	0.18	1.59 (3)	0.45	0.049	1.000
rs212004		*FHIT* intron 5	0.163/*A*	0.218/*A*	3.22	0.07	1.15	0.28						
	17								5.54 (3)	0.14	3.43 (3)	0.33	0.162	0.572
rs2736778		*FHIT* intron 5	0.288/*A*	0.355/*A*	3.33	0.07	2.45	0.12						
	16								7.97 (3)	0.047	7.69 (3)	0.053	0.011	0.104
hCV8351378		*FHIT* intron 5	0.300/C	0.350/C	0.81	0.37	0.34	0.56						
	16								13.10 (3)	0.0044	15.84 (3)	0.0012	0.142	0.543
rs760317		*FHIT* intron 5	0.490/*G*	0.427/*A*	4.64	0.03	8.54	0.0035						
	13								5.19 (2)	0.075	8.44 (2)	0.015	0.745	1.000
D3S1234														
rs722070		*FHIT* intron 5	0.433/*A*	0.482/*A*	1.54	0.21	3.53	0.060						
	7								2.05 (2)	0.36	3.79 (2)	0.15	0.017	0.556
rs2361339		*FHIT* intron 5	0.0718/T	0.0522/T	1.02	0.31	1.27	0.26						
	23								2.03 (2)	0.36	1.75 (2)	0.42	0.048	0.627
rs1040337		*FHIT* intron 5	0.350/C	0.366/C	0.19	0.67	0.01	0.91						
	9								0.39 (3)	0.94	0.84 (3)	0.84	0.021	0.321
rs1882904		*FHIT* intron 5	0.274/*A*	0.252/A	0.43	0.51	0.04	0.85						
	34								1.43 (2)	0.49	0.91 (2)	0.64	0.088	0.932
rs213294		*FHIT* intron 5	0.239/*T*	0.209/*T*	0.81	0.37	1.59	0.21						
	23								6.17 (2)	0.1	6.12 (2)	0.11	0.330	0.790
rs213408		*FHIT* intron 5	0.322/*A*	0.369/*A*	1.64	0.20	0.41	0.52						
	27								3.40 (3)	0.33	2.80 (3)	0.43	0.017	0.144
rs767000		*FHIT* intron 5	0.322/*G*	0.369/*G*	0.11	0.74	0.35	0.56						

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 brothers. 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 LODPAL and model-based analysis using GENEHUNTER yielded a maximum peak at D3S1234 (Fig. 2*B* 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 LODPAL. 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 inherence, 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. LODPAL 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; χ^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 FRA3B 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

boundary defined by SNPs rs2736778 and rs213294 using GeneScan and Grail, none of these corresponds 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 explication 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 9q22.2-31.2. 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/ng917_S1.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/OMIM; (CaP MIM 176807; *FHIT*, MIM 601153, *CDCD25a*, MIM 116947)

SNP DB, http://www.ncbi.nlm.nih.gov/SNP/

Human Genome Browser Gateway, http://genome.ucsc.edu/cgi-bin/hgGateway

Applied Biosystems SNP Genotyping database, http://myscience.appliedbiosystems.com/genotype/search.jsp?assayType=genotyping

Acknowledgments

Received 6/2/2004; revised 11/5/2004; accepted 11/17/2004.

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 RR03655 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.

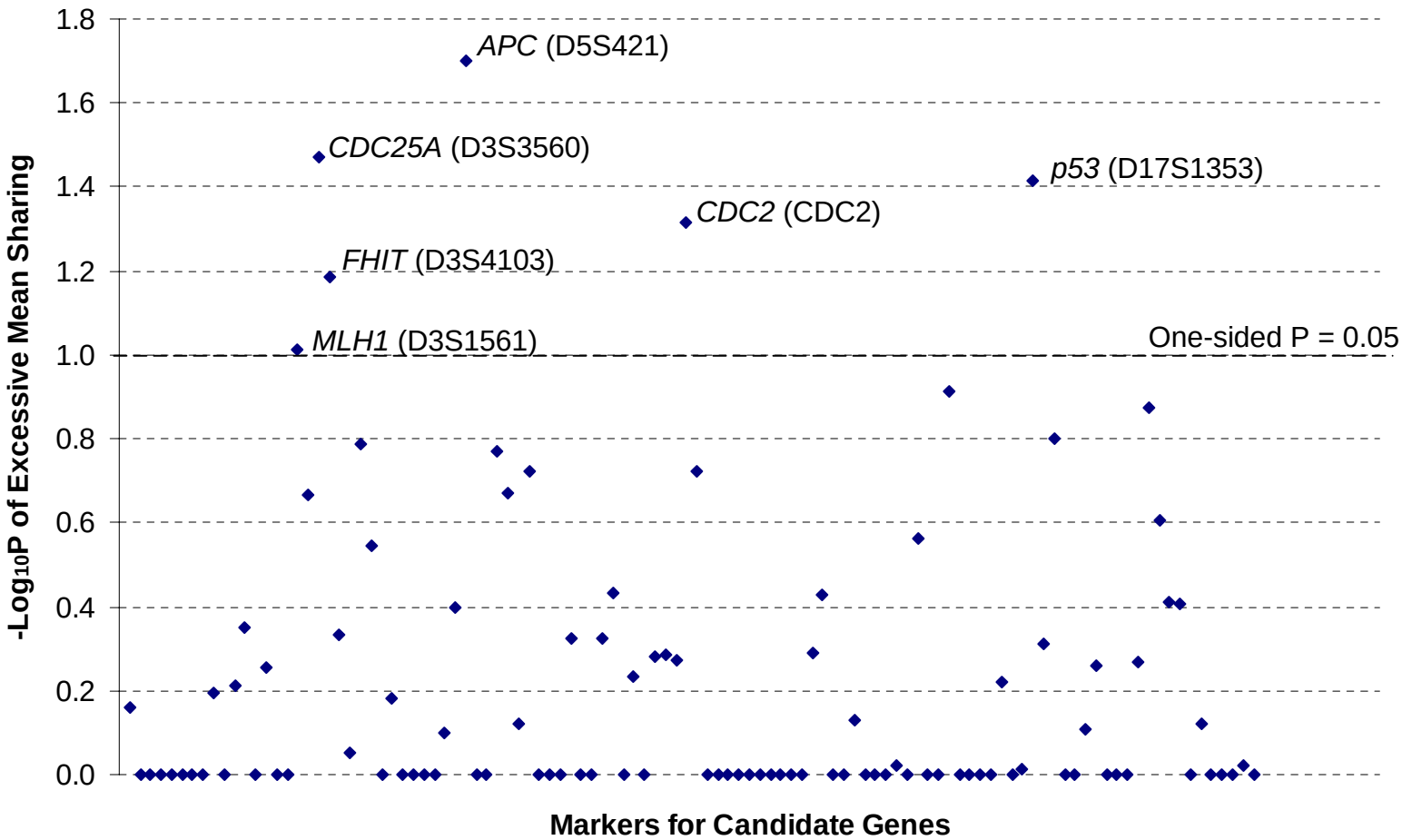
We thank all subjects for their participation, Mary Booth for help in initiating this study, and Dr. Robert Comis, Eastern Cooperative Oncology Group chair, for support in establishing this study.

References

1. Risch N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. Cancer Epidemiol Biomarkers Prev 2001;10:733–41.
2. Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. Eur J Cancer 2001;37 Suppl 8:S4–66.
3. Simard J, Dumont M, Labuda D, et al. International Congress on Hormonal Steroids and Hormones and Cancer: prostate cancer susceptibility genes: lessons learned and challenges posed. Endocr Relat Cancer 2003;10:225–59.
4. Nwosu V, Carpten J, Trent JM, Sheridan R. Heterogeneity of genetic alterations in prostate cancer: evidence of the complex nature of the disease. Hum Mol Genet 2001;10:2313–8.
5. Wiklund F, Jonsson BA, Goransson I, Bergh A, Gronberg H. Linkage analysis of prostate cancer susceptibility: confirmation of linkage at 8p22-23. Hum Genet 2003;112:414–8.
6. Xu J, Zheng SL, Chang B, et al. Linkage of prostate cancer susceptibility loci to chromosome 1. Hum Genet 2001;108:335–45.
7. Carpten J, Nupponen N, Isaacs S, et al. Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat Genet 2002;30:181–4.
8. Tavtigian SV, Simard J, Teng DH, et al. A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 2001;27:172–80.
9. Xu J, Zheng SL, Komiya A, 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.
10. Camp NJ, Tavtigian SV. Meta-analysis of associations of the Ser217Leu and Ala541Thr variants in ELAC2 (HPC2) and prostate cancer. Am J Hum Genet 2002;71:1475–8.
11. Schaid DJ. The complex genetic epidemiology of prostate cancer. Hum Mol Genet 2004;13:103–21.
12. Beilin J, Ball, EM, Favaloro JM, Zajac JD. Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. J Mol Endocrinol 2000;25:85–96.
13. Marcelli M, Ittmann M, Mariani S, et al. Androgen receptor mutations in prostate cancer. Cancer Res 2000;60:944–9.
14. Tayeb MT, Clark C, Sharp L, et al. CYP3A4 promoter variant is associated with prostate cancer risk in men with benign prostate hyperplasia. Oncol Rep 2002;9:653–5.
15. Thorlacius S, Olafsdottir G, Tryggvadottir L, et al. A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. Nat Genet 1996;13:117–9.
16. Edwards SM, Kote-Jarai Z, Meitz J, 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.
17. Dong X, Wang L, Taniguchi K, et al. Mutations in CHEK2 associated with prostate cancer risk. Am J Hum Genet 2003;72:270–80.
18. Ziogas A, Anton-Culver H. Validation of family history data in cancer family registries. Am J Prev Med 2003;24:190–8.
19. Larson GP, Zhang G, Ding S, et al. An allelic variant at the ATM locus is implicated in breast cancer susceptibility. Genet Test 1997;1:165–70.
20. Makridakis NM, Reichardt JK. Multiplex automated primer extension analysis: simultaneous genotyping of several polymorphisms. Biotechniques 2001;31:1374–80.
21. Statistical analysis for genetic epidemiology. Release 4.2 S.A.G.E. [computer program package]. Cork (Ireland): Statistical Solutions; 2002.
22. 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.
23. 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.
24. Kong A, Gudbjartsson DF, Sainz J, et al. A high-resolution recombination map of the human genome. Nat Genet 2002;31:241–7.
25. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–59.
26. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. Science 2002;298:2381–5.
27. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–89.
28. Elston RC, Guo X, Williams LV. Two-stage global search designs for linkage analysis using pairs of affected relatives. Genet Epidemiol 1996;13:535–58.
29. Olson JM. A general conditional-logistic model for affected-relative-pair linkage studies. Am J Hum Genet 1999;65:1760–9.
30. Greenwood CM, Bull SB. Down-weighting of multiple affected sib pairs leads to biased likelihood-ratio tests, under the assumption of no linkage. Am J Hum Genet 1999;64:1248–52.
31. 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.
32. Zabetian CP, Buxbaum SG, Elston RC, et al. The structure of linkage disequilibrium at the DBH locus strongly influences the magnitude of association between diallelic markers and plasma dopamine β-hydroxylase activity. Am J Hum Genet 2003;72:1389–400.
33. Monroe KR, Yu MC, Kolonel LN, et al. Evidence of an X-linked or recessive genetic component to prostate cancer risk. Nat Med 1995;1:827–9.
34. Cui J, Staples MP, Hopper JL, et al. Segregation analyses of 1,476 population-based Australian families affected by prostate cancer. Am J Hum Genet 2001;68: 1207–18.
35. Paiss T, Herkommer K, Chab A, et al. Familial prostate carcinoma in Germany. Urologe A 2002;41:38–43.
36. Valeri A, Briollais L, Azzouzi R, 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.
37. Zeegers MP, Jellema A, Ostrer H. Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma: a meta-analysis. Cancer 2003;97: 1894–903.
38. Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 1990;46:229–41.
39. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–7.
40. Smith JR, Freije D, Carpten JD, et al. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 1996; 274:1371–4.
41. Gibbs M, Stanford JL, Jarvik GP, et al. A genomic scan of families with prostate cancer identifies multiple regions of interest. Am J Hum Genet 2000; 67:100–9.
42. Xu J. Combined analysis of hereditary prostate cancer linkage to 1q24-25: results from 772 hereditary prostate cancer families from the International Consortium for Prostate Cancer Genetics. Am J Hum Genet 2000;66:945–57.
43. Hsieh CL, Oakley-Girvan I, Balise RR, et al. A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. Am J Hum Genet 2001;69:148–58.
44. Witte JS, Goddard KA, Conti DV, et al. Genomewide scan for prostate cancer-aggressiveness loci. Am J Hum Genet 2000;67:92–9.
45. Zhang W, Tapper W, Collins A, et al. A tournament of linkage tests in complex inheritance. Hum Hered 2001;52:140–8.
46. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
47. Fouts RL, Sandusky GE, Zhang S, et al. Down-regulation of fragile histidine triad expression in prostate carcinoma. Cancer 2003;97:1447–52.
48. Becker NA, Thorland EC, Denison SR, Phillips LA, Smith DI. Evidence that instability within the FRA3B region extends four megabases. Oncogene 2002; 21:8713–22.
49. Maruyama R, Toyooka S, Toyooka KO, et al. Aberrant promoter methylation profile of prostate cancers and its relationship to clinicopathological features. Clin Cancer Res 2002;8:514–9.
50. Harland M, Mistry S, Bishop DT, Bishop JA. A deep intronic mutation in CDKN2A is associated with disease in a subset of melanoma pedigrees. Hum Mol Genet 2001;10:2679–86.
51. Wiesner GL, Daley D, Lewis S, et al. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci U S A 2003;100:12961–5.

Supplemental Figure S1



Appendix 2 -
Levin, A., Anna M. Ray, Kimberly A. Zuhlke, Julie A. Douglas, and a. K. A. Cooney (2007). "Association between Germ line Variation in the *FHIT* Gene and Prostate Cancer in Caucasians and African Americans." Cancer Epidemiol Biomarkers Prev 16(6): 1.

Short Communication

Association between Germ line Variation in the *FHIT* Gene and Prostate Cancer in Caucasians and African Americans

Albert M. Levin,[1] Anna M. Ray,[2] Kimberly A. Zuhlke,[2] Julie A. Douglas,[1] and Kathleen A. Cooney[2,3]

Departments of [1]Human Genetics, [2]Internal Medicine, and [3]Urology, University of Michigan Medical School, Ann Arbor, Michigan

Abstract

Many studies have established that loss of heterozygosity and/or altered expression of the fragile histidine triad (*FHIT*) gene is a common event in a number of tumor types including prostate carcinoma. Encompassing the most active fragile site in the human genome, *FRA3B*, *FHIT* has become the model fragile site–associated tumor suppressor gene. In a recent study, linkage and association between germ-line genetic variation in *FHIT* (specifically single nucleotide polymorphism rs760317) and prostate cancer were reported. We sought to confirm this finding in two independent samples: (*a*) a family-based sample of 817 men with (n = 434) and without (n = 383) prostate cancer from 323 Caucasian families, and (*b*) a community-based case-control sample of African American men with (n = 133) and without (n = 342) prostate cancer. Using a family-based association test, rs760317 was associated with prostate cancer in Caucasians (P = 0.031), with a reduction in the risk of prostate cancer among carriers of the minor allele (odds ratio, 0.66; 95% confidence interval, 0.42-1.04; P = 0.074). African American carriers experienced a similar risk reduction (odds ratio, 0.63; 95% confidence interval, 0.42-0.96; P = 0.032). These results are remarkably consistent across ethnic samples but are in opposition to results from the original study, which showed an association between the minor allele of rs760317 and an increased risk of prostate cancer. Taken together, the consistently significant but flipped association between single nucleotide polymorphism rs760317 and prostate cancer in three independent samples suggests that rs760317 may be in linkage disequilibrium with one or more prostate cancer susceptibility variants in or near *FHIT*. (Cancer Epidemiol Biomarkers Prev 2007;16(6):1–4)

Introduction

Since its discovery in 1996, the fragile histidine triad (*FHIT*) gene has been established as the model fragile site–associated tumor suppressor gene. This large gene (~1.5 Mb) resides at chromosome 3p14.2 and encompasses the common fragile site *FRA3B*, overlapping exons 4 and 5. Whereas there is evidence of loss of heterozygosity and/or protein in many tumor types, the function of this gene and the mechanism by which its loss leads to tumor initiation and/or progression are still unclear.

Studies of *FHIT* in prostate cancer have been sparse relative to cancers of the gastrointestinal tract, colon, cervix, lung, and breast (reviewed in ref. 1). However, among the few published reports, there is some consensus that *FHIT* protein expression is down-regulated in primary prostate carcinomas (2, 3) and that this decrease is not the result of loss of heterozygosity within the gene (3). In a recent study, Larson et al. (4) reported suggestive evidence for linkage between prostate cancer and a microsatellite marker within *FHIT*. Following up their linkage signal with a denser set of single-nucleotide polymorphisms (SNP), these authors found a significant association between prostate cancer and SNP rs760317 (in intron 5 of *FHIT*) and a two-SNP haplotype (containing rs760317 and rs6791450). The present report examines these two *FHIT* SNPs in independent samples of Caucasians and African Americans.

Materials and Methods

Study Subjects. The first sample consisted of Caucasian families with at least one sibling pair discordant for prostate cancer. Men from these families (5) were recruited as part of the Prostate Cancer Genetics Program at the University of Michigan. Prostate Cancer Genetics Program families were primarily recruited from the University of Michigan Comprehensive Cancer Center. Other sources included direct patient or physician referrals. Prostate Cancer Genetics Program enrollment was restricted to (*a*) families with two or more living members with prostate cancer in a first- or second-degree relationship or (*b*) men diagnosed with prostate cancer at ≤55 years of age without a family history of the disease. All participants were asked to provide a blood sample for DNA extraction, extended family history information, and access to medical records. For this sample, the oldest available unaffected brother from each family was preferentially enrolled to maximize the probability that unaffected men were truly unaffected and not simply unaffected by virtue of being younger than their affected brother(s). Additional male siblings and multiple sibships from the same family were included if DNA was available. For this analysis, 323 Caucasian families were genotyped.

The second sample consisted of African American men with and without prostate cancer, who were recruited as part of the Flint Men's Health Study (6). Starting in 1996, 943 potentially eligible men were selected from a probability sample of African American men ages 40 to 79 years in Flint, Michigan,

Received 12/20/06; revised 4/4/07; accepted 4/10/07.

Grant support: NIH Specialized Program of Research Excellence in Prostate Cancer grant P50 CA69568, NIH grant R01 CA79596 (K.A. Cooney), the Department of Urology at the University of Michigan Medical School, and the University of Michigan Comprehensive Cancer Center.

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.

Requests for reprints: Kathleen A. Cooney, 7310 CCGC, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0946. Phone: 734-764-2248; Fax: 734-736-7376. E-mail: kcooney@umich.edu

doi:10.1158/1055-9965.EPI-06-1054

and neighboring Beecher Township (Genesee County, Michigan). Unaffected men were excluded if they were previously diagnosed with prostate cancer and/or had a previous operation involving the prostate gland. A total of 379 eligible unaffected men completed urologic and physical examinations in conjunction with prostate-specific antigen screening, a blood draw, and questionnaire, and 342 unaffected men had DNA available for this study. African American men with prostate cancer diagnosed between 1995 and 2002 were identified through the Genesee County Community-Wide Hospital Oncology program registry, which covers the three local hospitals servicing the community. Between 1999 and 2002, 138 men with prostate cancer agreed to participate in the study, and 133 had DNA available for this analysis.

Below, we refer to the Prostate Cancer Genetics Program sample as the ''Caucasian sample'' and the Flint Men's Health Study sample as the ''African American sample.'' The Institutional Review Board at the University of Michigan Medical School approved all aspects of both study protocols, and all participants gave written informed consent.

Genotyping Assays. Two SNPs in intron 5 of *FHIT* (rs760317 and hCV8351378/rs6791450) were genotyped using TaqMan SNP assays (Applied Biosystems). Genotyping call rates for rs760317 and rs6791450 were 98.9% and 97.9%, respectively, and the undetermined samples were sequenced to achieve a final call rate of 100% for both SNPs. A subset of genotypes was duplicated by TaqMan (5.5%) or direct sequencing (3.0%) for each SNP, and no discrepancies were observed.

To test for potential population substructure in the African American sample, 42 unlinked microsatellite markers were genotyped by deCODE Genetics in a separate collaborative project (7). These markers are located on the Marshfield genetic map and were selected to distinguish between European, African, and Asian ancestry.

Statistical Methods. Within each sample, observed genotype distributions were tested for departure from Hardy-Weinberg equilibrium in a subset of unrelated, unaffected men. For the Caucasian sample, this subset consisted of the oldest unaffected man from each family. SNP genotypes did not depart from Hardy-Weinberg equilibrium in either sample at a significance level of 0.05. Haplotype frequencies were estimated using the expectation-maximization algorithm and were used to calculate the linkage disequilibrium measure r^2.

For the Caucasian sample, we used the family-based association method (ref. 8; implemented in the FBAT software, version 1.5.5) to test for association between single SNPs and prostate cancer. To maximize power, we analyzed the combined set of affected and unaffected men using the offset option to test the null hypothesis of no association and no linkage. To account for the possible misclassification of unaffected men, we analyzed only affected men using the empirical variance estimate to test the null hypothesis of no association in the presence of linkage. Conditional logistic regression, coupled with a robust variance estimate that incorporates familial correlations (9), was used to generate odds ratios (OR) and 95% confidence intervals (95% CI). Two-SNP haplotypes were analyzed using the haplotype FBAT (HBAT) method (10).

For the African American sample, we used logistic regression to test for association between each SNP and prostate cancer and to estimate ORs and 95% CIs. Tests of association between two-SNP haplotypes and prostate cancer were conducted using the haplotype generalized linear model method proposed by Lake et al. (11). Individual haplotypes were evaluated using a model-specific Wald test. In all African American analyses, age and family history of prostate cancer in a first-degree relative were included as potential confounders.

To test for population substructure in the African American sample, we implemented the method of Pritchard and Rosenberg (12) using 42 unlinked microsatellite markers. The observed summary χ^2 measure was 133.13 with 142 degrees of freedom ($P = 0.96$), suggesting that hidden population substructure is unlikely to generate false-positive evidence for association.

For both samples, we calculated single SNP and haplotype association tests under additive, dominant, and recessive models. For single SNPs, an additional genotype model (2 degree of freedom test) was used. All statistical tests were two sided, with the significance level set at 0.05. Conditional logistic regression was conducted using version 8.2 of the SAS programming language. All remaining analyses were carried out using the R-language.[4]

Table 1. Minor allele frequencies in affected and unaffected men

Sample (no. affected/ no. unaffected)	dbSNP ID	Minor allele frequency		
		Affected	Unaffected	*P**
Caucasian (434/383)	rs760317[†]	0.45	0.50	0.047
	rs6791450[‡]	0.32	0.33	0.524
African American (133/342)	rs760317	0.23	0.29	0.105
	rs6791450	0.47	0.47	0.995

*P value from the Z test of proportions assuming independence of all individuals.
†rs760317 (G > A) is located at base pair 60,074,196 on chromosome 3.
‡rs6791450 (T > C) is located at base pair 60,057,979 on chromosome 3 and is recorded as hCV8351378 by Larson et al.

Table 2. Family-based association test results from the Caucasian sample

dbSNP ID	Model*	Affecteds and unaffecteds			Affecteds only		
		n[†]	*Z* score	*P*	n[†]	*Z* score	*P*
rs760317	Additive	162	−2.22	0.026	152	−2.31	0.021
	Dominant	96	−2.15	0.031	92	−2.04	0.041
rs6791450	Additive	152	−0.85	0.396	141	−0.91	0.363
	Dominant	121	−1.11	0.266	123	−1.09	0.276

*Both models are with respect to the minor allele, which is ''A'' for rs760317 and ''C'' for rs6791450.
†Number of informative families.

Results

The Caucasian sample included 434 men with prostate cancer and 383 unaffected men from 323 families with at least one pair of brothers discordant for prostate cancer. Of these families, 221 included only a single discordant sibling pair (DSP). The remaining families included additional DSPs from the same sibship (e.g., two brothers with and one without prostate cancer or two DSPs) or from the same family but different sibships (e.g., a pair of DSPs related as first cousins), resulting in a total sample of 516 DSPs. The median age at diagnosis for Caucasian men with prostate cancer was 55 years (interquartile range, 50-63 years), and the median age at consent for unaffected men was 56 years (interquartile range, 50-63 years).

The minor allele frequency of rs760317 was 5% greater in unaffected men compared with affected men ($P = 0.047$; Table 1). Consistent with this difference, we also detected

[4] http://www.R-project.org

Table 3. Estimated ORs from logistic regression

dbSNP ID	Model*	Sample			
		Caucasian		African American †	
		OR (95% CI)	*P*	OR (95% CI)	*P*
rs760317	Additive	0.77 (0.57-1.03)	0.073	0.71 (0.51-1.00)	0.050
	Dominant	0.66 (0.42-1.04)	0.074	0.63 (0.42-0.96)	0.032
rs6791450	Additive	0.91 (0.69-1.19)	0.483	1.00 (0.75-1.33)	0.997
	Dominant	0.81 (0.56-1.17)	0.267	0.96 (0.61-1.51)	0.862

*Both models are with respect to the minor allele, which is ''A'' for rs760317 and ''C'' for rs6791450.
†All logistic regression models for the African American sample were adjusted for age and family history of prostate cancer in a first-degree relative.

significant overtransmission of the minor allele of rs760317 to unaffected men compared with affected men in our family-based analysis. In the combined sample of affected and unaffected men, both additive and dominant models for rs760317 showed significant evidence for prostate cancer association (Table 2). Before estimating ORs, we excluded 18 men who were not brothers of the index case from seven multisibship families, resulting in a reduced sample size of 799 men and 506 DSPs. Conditional logistic regression results are presented in Table 3. The OR associated with each minor allele at rs760317 was 0.77 (95% CI, 0.57-1.03; $P = 0.073$).

The African American sample included 133 affected and 342 unaffected men. The median age at diagnosis for African American men with prostate cancer was 63 years (interquartile range, 56-69 years) and the median age at consent for unaffected men was 55 years (interquartile range, 49-63 years). Similar to the Caucasian sample, the rs760317 minor allele frequency was 6% greater in unaffected men compared with affected men (Table 1). Using logistic regression (Table 3), the OR associated with each minor allele at rs760317 was 0.71 after adjustment for age and family history of prostate cancer (95% CI, 0.51-1.00; $P = 0.050$). Under a dominant model, the effect of the minor allele was also significant ($P = 0.032$).

SNP rs6791450 was not associated with prostate cancer in either sample (Tables 2 and 3). Notably, rs6791450 is located ~16 kb from rs760317 and was not in strong linkage disequilibrium with rs760317 in either the Caucasian ($r^2 = 0.18$) or African American ($r^2 < 0.01$) sample. In the Caucasian sample, the haplotype defined by the major alleles of both SNPs was overtransmitted to affected men under additive ($P = 0.041$) and recessive models ($P = 0.045$), consistent with the single SNP result for rs760317. In the African American sample, there was a reduction in risk associated with the haplotype defined by the minor allele of rs760317 and the major allele of rs6791450 under additive ($P = 0.003$) and dominant ($P = 0.005$) models.

Discussion

In summary, our results show association between genetic variation in *FHIT* (specifically rs760317) and prostate cancer in two independent samples. The association between rs760317 and prostate cancer was remarkably similar in direction and magnitude in Caucasian and African American samples. Whereas our data indicated a protective effect associated with the minor allele of rs760317, Larson et al. (4) found the opposite effect. In their study, men homozygous for the minor allele showed an ~2-fold increased risk of prostate cancer in comparison with carriers of at least one copy of the major allele.[5] We were able exclude the possibility that genotyping error was the source of this allelic reversal through a mutual exchange of 12 anonymous DNA samples with Larson et al. group (i.e., there were no discrepancies; data not shown).

This pattern of allelic reversal has been noted in replication studies of other candidate SNPs (13, 14), and several such discrepancies have been shown to differ beyond what would occur by chance alone (14). Further, in a recently published study investigating the potential causes of this ''flip-flop'' phenomenon, Lin et al. (15) suggested that a genotyped SNP interacting with a nongenotyped causal SNP may show a flipped association when the minor allele frequency of the genotyped SNP is high (~0.5), the pair is in relatively low linkage disequilibrium ($r^2 < 0.3$), and the interaction of the two is not accounted for in the model. Given the relatively high minor allele frequency of rs760317, this explanation of the observed result is plausible. Of note, rs760317 was not genotyped in the International HapMap project[6] or the recent prostate cancer genome-wide association study conducted by the Cancer Genetic Markers of Susceptibility initiative.[7]

Whereas a functional relationship between *FHIT* and tumorigenesis and/or progression is still unknown, data from the mouse suggest that *FHIT* haploinsufficiency predisposes to a wide range of tumors (16). In addition, alternatively spliced *FHIT* transcripts have been shown to occur in nonneoplastic tissue (17), some of which lead to loss of a functional protein product. Whereas rs760317 does not directly alter a known splice site (18), it could be in linkage disequilibrium with another SNP that influences alternative splicing of the gene, potentially reducing the amount of the functional protein product. Further, rs760317 resides in a region of intron 5 that is commonly deleted in tumor cell lines (19), suggesting an important role for sequence variation in this region. Additional resequencing and functional work will be required to evaluate the direct or indirect influence of rs760317 on the integrity of normal *FHIT* expression. In view of the data presented here, this additional work seems justified.

Acknowledgments

We thank Joe Washburn and the Microarray Core Facility at the University of Michigan Comprehensive Cancer Center for assistance with genotyping assays, and Drs. Theodore G. Krontiris and Yan Ding (City of Hope National Medical Center) for openly discussing and sharing data from their ongoing investigation of *FHIT*.

References

1. Huebner K, Croce CM. Cancer and the FRA3B/FHIT fragile locus: it's a HIT. Br J Cancer 2003;88:1501–6.
2. Fouts RL, Sandusky GE, Zhang S, et al. Down-regulation of fragile histidine triad expression in prostate carcinoma. Cancer 2003;97:1447–52.
3. Guo Z, Johansson SL, Rhim JS, Vishwanatha JK. Fragile histidine triad gene expression in primary prostate cancer and in an *in vitro* model. Prostate 2000;43:101–10.
4. Larson GP, Ding Y, Cheng LSC, et al. Genetic linkage of prostate cancer risk to the chromosome 3 region bearing FHIT. Cancer Res 2005;65:805–14.

Q3 [5] Personal communication.
[6] http://www.hapmap.org
[7] http://cgems.cancer.gov/

5. Douglas JA, Zuhlke KA, Beebe-Dimmer J, et al. Identifying susceptibility genes for prostate cancer—a family-based association study of polymorphisms in CYP17, CYP19, CYP11A1, and LH-β. Cancer Epidemiol Biomarkers Prev 2005;14:2035–9.
6. Cooney KA, Strawderman MS, Wojno KJ, et al. Age-specific distribution of serum prostate-specific antigen in a community-based study of African-American men. Urology 2001;57:91–6.
7. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet 2006;38:652–8.
8. Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 2000;50:211–23.
9. Siegmund KD, Langholz B, Kraft, P, Thomas DC. Testing linkage disequilibrium in sibships. Am J Hum Genet 2000;67:244–8.
10. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 2004;26:61–9.
11. Lake SL, Lyon H, Tantisira K, et al. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. Hum Hered 2003;55:56–65.
12. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65:220–8.
13. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003;33:177–82.
14. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001;29:306–9.
15. Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. Am J Hum Genet 2007;80:531–8.
16. Zanesi N, Fidanza V, Fong LY, et al. The tumor spectrum in FHIT-deficient mice. Proc Natl Acad Sci U S A 2001;98:10250–5.
17. Panagopoulos I, Thelin S, Mertens F, Mitelman F, Aman P. Variable FHIT transcripts in non-neoplastic tissues. Genes Chromosomes Cancer 1997;19:215–9.
18. Ohta M, Inoue H, Cotticelli MG, et al. The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. Cell 1996;84:587–97.
19. Iliopoulos D, Guler G, Han SY, et al. Roles of FHIT and WWOX fragile genes in cancer. Cancer Lett 2006;232:27–36.