

Bao-Li Chang · Ethan M. Lange · Latchezar Dimitrov
Christopher J. Valis · Elizabeth M. Gillanders
Leslie A. Lange · Kathleen E. Wiley · Sarah D. Isaacs
Fredrik Wiklund · Agnes Baffoe-Bonnie
Carl D Langefeld · S. Lilly Zheng · Mika P. Matikainen
Tarja Ikonen · Henna Fredriksson · Teuvo Tammela
Patrick C. Walsh · Joan E. Bailey-Wilson
Johanna Schleutker · Henrik Gronberg
Kathleen A. Cooney · William B. Isaacs · Edward Suh
Jeffrey M. Trent · Jianfeng Xu

Two-locus genome-wide linkage scan for prostate cancer susceptibility genes with an interaction effect

Received: 24 June 2005 / Accepted: 26 September 2005 / Published online: 23 November 2005
© Springer-Verlag 2005

Abstract Prostate cancer represents a significant world-wide public health burden. Epidemiological and genetic epidemiological studies have consistently provided data supporting the existence of inherited prostate cancer susceptibility genes. Segregation analyses of prostate cancer suggest that a multigene model may best explain familial clustering of this disease. Therefore, modeling gene–gene interactions in linkage analysis may improve

the power to detect chromosomal regions harboring these disease susceptibility genes. In this study, we systematically screened for prostate cancer linkage by modeling two-locus gene–gene interactions for all possible pairs of loci across the genome in 426 prostate cancer families from Johns Hopkins Hospital, University of Michigan, University of Umeå, and University of Tampere. We found suggestive evidence for an epistatic

Bao-Li Chang and Ethan M. Lange have contributed equally to the work

B.-L. Chang · L. Dimitrov · S. L. Zheng · J. Xu
Center for Human Genomics,
Wake Forest University School of Medicine,
Winston-Salem, NC, USA

C. J. Valis · C. D. Langefeld · J. Xu
Department of Public Health Sciences,
Wake Forest University School of Medicine,
Winston-Salem, NC, USA

E. M. Lange · L. A. Lange
Department of Genetics,
University of North Carolina,
Chapel Hill, NC, USA

E. M. Lange
Department of Biostatistics,
University of North Carolina,
Chapel Hill, NC, USA

E. M. Gillanders · J. E. Bailey-Wilson
National Human Genome Research Institute,
National Institute of Health,
Bethesda, MD, USA

K. E. Wiley · S. D. Isaacs · P. C. Walsh · W. B. Isaacs
Department of Urology, Johns Hopkins Medical Institutions,
Baltimore, MD, USA

K. A. Cooney
Department of Internal Medicine and Urology,
University of Michigan, Ann Arbor, MI, USA

F. Wiklund · H. Gronberg
Department of Radiation Sciences, Oncology,
University of Umeå, Umeå, Sweden

A. Baffoe-Bonnie
Division of Population Science, Fox Chase Cancer Center,
Philadelphia, PA, USA

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

T. Ikonen · H. Fredriksson · J. Schleutker · M. P. Matikainen
Institute of Medical Technology,
University of Tampere and Tampere University Hospital,
Tampere, Finland

E. Suh · J. M. Trent · J. Xu
Translational Genomics Research Institute (TGen),
Phoenix, AZ, USA

J. Xu (✉)
Medical Center Blvd, Winston-Salem, NC, 27157 USA
E-mail: jxu@wfubmc.edu
Tel.: +1-336-7137500
Fax: +1-336-7137566

interaction for six sets of loci (target chromosome-wide/reference marker-specific $P \leq 0.0001$). Evidence for these interactions was found in two independent subsets from within the 426 families. While the validity of these results requires confirmation from independent studies and the identification of the specific genes underlying this linkage evidence, our approach of systematically assessing gene–gene interactions across the entire genome represents a promising alternative approach for gene identification for prostate cancer.

Introduction

Genetic predisposition to prostate cancer has been well established through genetic epidemiological studies (Isaacs et al. 2001). Evidence for major prostate cancer susceptibility genes segregating in families has been obtained from several complex segregation analyses, with the majority supporting a dominant mode of inheritance (Carter et al. 1992; Schaid et al. 1998; Gronberg et al. 2000; Verhage et al. 2001), while others support a recessive or X-linked mode of inheritance (Cui et al. 2001; Gong et al. 2002). The hypothesis that multiple major genes interact to increase prostate cancer susceptibility is supported by our understanding of tumorigenesis of cells, by the parallel evidence that multiple major genes contribute to cancer in other organ sites such as the breast (Miki et al. 1994; Wooster et al. 1995) and colon (Groden et al. 1991; Kinzler et al. 1991; Fishel et al. 1993; Leach et al. 1993; Bronner et al. 1994; Nicolaidis et al. 1994; Miyaki et al. 1997), and by observational data from molecular and genetic epidemiological studies. A segregation study in 263 prostate cancer families found that the disease is more likely due to the contributions of two to four prostate cancer susceptibility genes than one gene (Conlon et al. 2003). The interaction of genes that leads to a disease may be

described as consistent with either a heterogeneity model in which alterations in any of several genes is sufficient, or an epistatic model in which several simultaneous genetic alterations are required.

Linkage analysis methods that model interactions may increase the statistical power to detect linkage when interactions among genes exist. For example, under a two-locus epistatic model, the power to detect linkage at one susceptibility locus would increase if we assess the linkage among families that are linked at another susceptibility locus and vice versa. Several published studies have empirically demonstrated the increased power to detect linkage by modeling gene–gene interactions. For example, although there is a lack of linkage evidence at either the *PTEN* or *CDKN1B* chromosomal regions when each region was studied individually among 188 prostate cancer families ascertained from Johns Hopkins Hospital, a modeled interaction of these two regions in a linkage analysis provided significant evidence for linkage (Xu et al. 2004).

In spite of our understanding that multiple major genes likely contribute to prostate cancer susceptibility and that there are advantages in modeling interactions in linkage analyses, current studies to identify these major genes continue to primarily rely on single gene approaches. This is particularly true when exploring for novel regions of linkage using genome-wide scans. Among a dozen genome-wide scans for prostate cancer susceptibility genes published to date, none of these have modeled gene–gene interactions (Smith et al. 1996; Witte et al. 2000; Hsieh et al. 2001; Cunningham et al. 2003; Edwards et al. 2003; Janer et al. 2003; Lange et al. 2003; Schleutker et al. 2003; Wiklund et al. 2003; Xu et al. 2003; Maier et al. 2005). This gap may be primarily due to a combination of factors, such as a lack of standard analytical methods to model interactions and sample sizes that are not large enough to investigate interactions.

As an attempt to close this gap, we have designed and implemented a study to identify prostate cancer loci by modeling interactions in a linkage analysis of 426 prostate cancer families. We limited the analyses of interactions to two loci at a time and systematically evaluated evidence for prostate cancer linkage among all possible combinations of two loci across the entire genome using an ordered subset analysis (OSA) to model epistatic interactions (Hauser et al. 2004).

Table 1 Characteristics of families in the combined analysis

Family characteristics	Number of families (genotyped individuals)				
	JHH	Michigan	Umea	Tampere	Total
All	188 (1,033)	175 (640)	50 (190)	13 (87)	426 (1,950)
Mean age at diagnosis (years)					
< 65	96	91	10	4	201
≥ 65	92	84	40	9	225
Number of affected members					
≤ 3	28	96	13	4	141
= 4	47	45	17	5	114
≥ 5	113	34	20	4	171
Race					
White	171	157	50	13	391
Black	15	16	0	0	31
Others	2	2	0	0	4

JHH Johns Hopkins Hospital

Methods

Study populations

The 426 prostate cancer families were ascertained from four independent studies, including 188 families from Johns Hopkins Hospital (Xu et al. 2003), 175 families from University of Michigan (Lange et al. 2003), 50 families from University of Umeå, Sweden (Wiklund

et al. 2003), and 13 families from University of Tampere, Finland (Schleutker et al. 2003). A previous genome-wide linkage scan for a prostate cancer susceptibility locus combining these 426 families gave a nonparametric multipoint LOD score of 3.16 at chromosome 17q22 and LOD scores greater than 2.0 at chromosomes 2q32, 15q11, and Xq27 (Gillanders et al. 2004). A detailed description of the family ascertainment has been described elsewhere (Gillanders et al. 2004). The clinical characteristics of these families are summarized in Table 1. Informed consent was obtained from all participants, and study protocols were reviewed and approved by the Institutional Review Boards at each institution.

Marker genotyping

A detailed description of our methods for marker genotyping has been presented elsewhere (Gillanders et al. 2004). Briefly, genomic DNA was prepared from blood samples using standard techniques. All DNA samples were genotyped in a single laboratory using 406 short tandem repeat markers with an average inter-marker spacing of ~ 10 cM and an average heterozygosity of 80%. PCR products were separated using the ABI 377 or 3100 DNA sequencers, allowing multiple fluorescently labeled markers to be run in a single lane. Allele sizing was calculated using a local southern algorithm available in the GENESCAN software program (Applied Biosystems, Foster City, CA). Allele calling and binning was done using GENOTYPER software (Applied Biosystems, Foster City, CA). All genotyping included a CEPH control individual (1347-02) for quality control purposes. Additionally, 1% of samples were included as blind duplicates in order to evaluate the genotyping error rate. Of the 12,992 duplicated genotypes, eleven genotyping errors were detected (0.085% error rate).

Statistical analyses

Nonparametric multipoint linkage analysis was first performed using the computer program Genehunter-Plus (Kong and Cox 1997). Maximum likelihood marker allele frequency estimates were calculated from pedigree founders. Nonparametric-based LOD scores were calculated using the 'Z-all' allele-sharing statistic (Whittemore and Halpern 1994), with an equal weight assigned to each family. LOD scores were calculated at each marker location and at four evenly spaced locations between each pair of consecutive markers.

OSA conditional linkage analyses were performed to model epistatic interactions (Hauser et al. 2003). Briefly, given a pair of unlinked loci, a series of linkage analyses at the "target" region (locus 1) were evaluated conditional on the linkage results at the "reference" region

(locus 2). Family-specific LOD scores at locus 2 were ranked from largest to smallest. The family with the largest LOD score at locus 2 was entered into the analysis and the corresponding LOD score was computed and recorded at locus 1, for that family. Next, a second linkage analysis at locus 1 was computed and recorded by combining the two families with the two largest LOD scores at locus 2. The i th OSA analysis proceeded by computing linkage at locus 1 using the subset of families with the i largest LOD scores at locus 2. This process was repeated until all families were sequentially added to the linkage analysis at the target region. For each target/reference marker pairing, we defined $LOD_{\text{delta}} = LOD_{\text{conditional}} - LOD_{\text{unconditional}}$, where $LOD_{\text{conditional}}$ was determined by the OSA analysis that provided the maximum LOD at locus 1.

To minimize the impact of multiple correlated tests, the reference markers were restricted to the 406 markers and did not include locations between the markers. The target loci consisted of the 406 markers and four evenly spaced locations between each pair of adjacent markers. Reference and target loci were restricted from being on the same chromosome. The statistical significance of the change in the LOD score (LOD_{delta}) was evaluated by a permutation test under the null hypothesis that linkage at the target region was independent of linkage at the reference region. To further minimize the impact of multiple tests, 22 P -values representing chromosome-wide significance were estimated for each reference marker. Specifically, for each reference marker and for each of the 22 chromosomes not containing the reference marker in question, the LOD_{delta} statistic was calculated at each target locus across the chromosome. The largest value of the LOD_{delta} statistic across all target loci on the chromosome was used as the test statistic in the permutation testing procedure. Family-specific LOD scores at all target loci across the chromosome were randomly and jointly permuted with respect to the ordering defined by the reference locus, thus preserving the correlation structure of LOD scores between linked markers. The maximum LOD_{delta} over all target loci on the chromosome was determined for each randomly permuted data set and this value was compared to the observed maximum LOD_{delta} for the chromosome. The empirical P -value for each target chromosome-wide/reference marker-specific pairing was set equal to the number of replicated chromosome-wide maximum LOD_{delta} s that were greater than or equal to our observed chromosome-wide maximum LOD_{delta} . Using this target-chromosome-by-reference-marker design resulted in $22 \times 406 = 8,932$ hypotheses being tested. To account for the multiple tests, we considered target chromosome-wide/reference marker-specific P -values $< \sim 5.6 \times 10^{-6}$ ($0.05/8,932$) to be globally statistically significant at the 0.05 level. This threshold is conservative as it does not account for the correlation of test results between linked reference markers. The number of permutations performed for the OSA varied as a function of the magnitude of the target chromosome-wide/

reference marker-specific P -value for the sake of computational efficiency, and was set at a minimum of 100 and a maximum of 1,000,000.

In addition to the genome-wide interaction analyses, we focused particular attention on interactions with the region of prior, “main effects”, linkage on chromosome 17q22. Given the evidence for linkage at 17q22 (Gillanders et al. 2004), we performed analyses that focused specifically on 17q22 to reduce the impact of multiple testing and to subsequently reduce the required threshold for statistical significance. Specifically, we tested the hypotheses that if we conditioned on the reference marker D17S787, the marker directly under the linkage peak at 17q22, we would find increased evidence for linkage due to epistasis at other chromosomal locations. Given this *a priori* hypothesis, we used a threshold of $0.05/22=0.0023$ to assess whether interactions with the reference marker S17S787 were statistically significant at the 0.05 level. In addition, we used the 1-LOD support interval about our linkage peak at 17q22 (from Gillanders et al. 2004) as the target region and looked for interactions conditioning on all 391 reference markers not on chromosome 17. We used a threshold for statistical significance in these analyses of $0.05/391=0.00013$.

Results

Selected demographic and clinical characteristics (mean age of diagnosis, number of affected individuals, and ethnicity) are presented in Table 1. Collectively, this set of families represents one of the largest collections of prostate cancer families ascertained and analyzed to date. In particular, this set of families includes a large number of pedigrees that most likely segregate highly penetrant prostate cancer genes based on family characteristics. For example, there are 285 families with at least four men affected with prostate cancer, and 201 families with a mean age of diagnosis under 65 years.

We performed two-locus conditional linkage analyses for all possible pairs of loci across the genome in the 426 prostate cancer families using the OSA method. Multiple loci had significantly increased evidence for linkage compared to the unconditional single locus linkage analyses (Fig. 1). None of the target chromosome/reference marker combinations reached global statistical significance assuming our strict genome-wide threshold for statistical significance of $P=5.6\times 10^{-6}$. As detailed in Table 2, the differences in the LOD scores between conditional and unconditional analyses (LOD_{delta}) in six target chromosome/reference marker combinations reached target chromosome-wide/reference marker-specific significance levels of $P\leq 0.0001$. For example, when linkage analysis at a target region of chromosome 12 was evaluated conditional on the linkage results at reference regions across the genome, the evidence for

linkage was maximized at 12q24 ($LOD=5.69$) when the 78 families with the highest LOD scores at 16p13 were included. This LOD score was significantly higher than the LOD of 0.29 in the unconditional analysis, with a LOD_{delta} of 5.41. Among 1,000,000 randomly permuted datasets with respect to the relationship between target and reference regions, only 22 LOD_{delta} s reached 5.41 ($P=0.000022$). This result revealed that increased allele sharing at 12q24 among affected men was observed most strongly in families that also had increased allele sharing at 16p13, a phenomenon that is consistent with an epistatic interaction of two prostate cancer susceptibility genes in these two regions. Evidence for interaction effects using a target chromosome-wide/reference marker-specific significance level of $P<0.0001$ were also observed for five other pairs of loci at 11q13, 22q13, 8q24, 20p13, and 5p13, when conditioning on reference markers at 13q12, 21q22, 7q21, 16q21, and 16p12, respectively.

The evidence for epistatic interactions for these six sets of loci were consistently supported by two prior defined subsets of families; 188 families from Johns Hopkins and 238 families from the three other groups (Michigan, Umeå, and Tampere), as shown in Table 3. For example, when linkage analysis at chromosome 5 was evaluated conditional on the linkage results at reference regions across the genome among the 188 families from Johns Hopkins, the evidence for linkage was maximized at 5p13 ($LOD=3.15$) among families with the highest LOD score at 16p12. This LOD_{delta} between conditional and unconditional analyses was 2.88 with a chromosome-wide significance level of $P=0.009$. Similarly, when linkage analysis at chromosome 5 was evaluated conditional on the linkage results at reference regions across the genome among the 238 families from the three other groups, the evidence for linkage was also maximized at 5p13 ($LOD=3.57$) among families with the highest LOD score at 16p12. This LOD_{delta} between conditional and unconditional analyses was also 2.88 with a target chromosome-wide/reference marker-specific significance level of $P=0.02$. Consistent support of epistatic interactions from these two sets of families was also observed at the other five sets of loci (Table 3).

Considering that the 17q22 region was implicated by a single gene approach where a LOD of 3.16 was observed using nonparametric linkage analysis (Gillanders et al. 2004), we examined the results of the various interaction models involving this region. A marginally globally statistically significant interaction was observed in the epistatic interaction models using our reduced thresholds for statistical significance for this *a priori* plausible region. Specifically, when conditioning on the reference marker D17S787, the target region on chromosome 4q35 (target chromosome-wide/reference marker-specific $P=0.0018$) was marginally statistically significant at the 0.0023 level (deemed equivalent to a global significance level of 0.05) and the target region on chromosome 11q14 (target chromosome-wide/reference

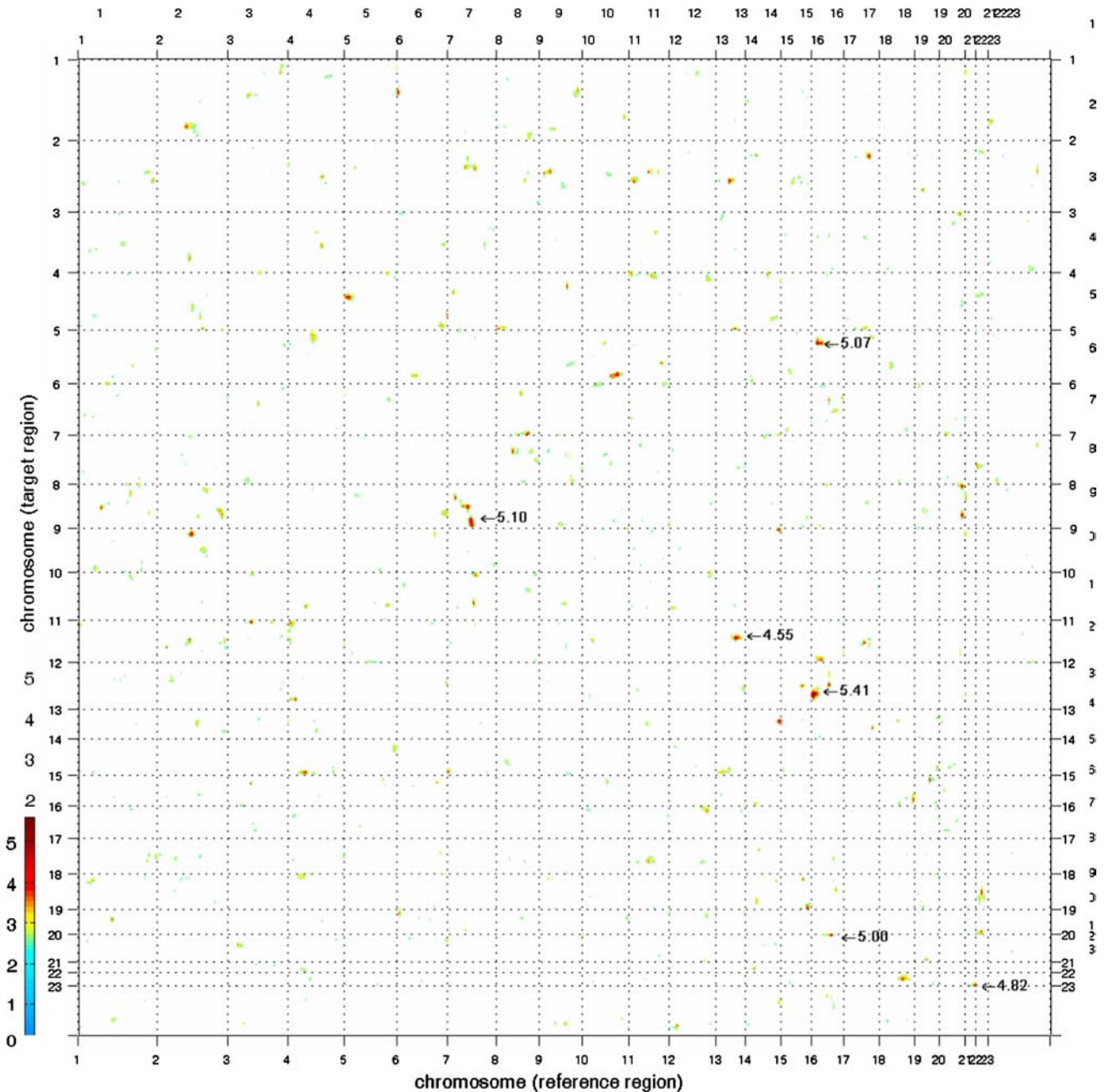


Fig. 1 Results from two-locus conditional linkage analyses for all possible pairs of loci across the genome among 426 prostate cancer families using the OSA method are presented using a contour plot.

The LOD_{Δ} ($LOD_{\Delta} = LOD_{\text{conditional}} - LOD_{\text{unconditional}}$) is plotted using a color scheme as indicated by the legend on the left of the figure. The six sets of strongest interaction ($P < 0.0001$) are labeled

marker-specific $P=0.0045$) was suggestive. When linkage at chromosome 17q22 [using a 1-LOD support interval of our linkage peak in Gillanders et al. (2004) to define the target region] was evaluated conditional on the linkage results at the 391 nonchromosome 17 reference markers across the genome, the conditional LOD scores were maximized when conditioning on markers located at 1q44, 2p21, 4p15, 8q21, and 11q14, respectively (Table 4). None of the target 1-LOD-support-interval/reference marker-specific P -values reached the

threshold of 0.00013 for global statistical significance. Interestingly, 11q14 showed up as suggestive in both sets of analyses.

Discussion

Because multiple genetic alterations are likely required for tumor development (Land et al. 1983), the inheritance of germline mutations in several genes is expected

Table 2 Summary results from two-locus epistatic interaction linkage analysis in the genome

Target region			Reference region			LOD scores			P-value	Number of families included
Cyto-band	Position (cM)	Nearest marker	Marker	Cyto-band	Position (cM)	Conditional	Unconditional	Difference		
11q13	65.4	D11S987	D13S1241	13q32	70.2	6.12	1.57	4.55	0.000015	201
22q13	47.5	D22S274	D21S266	21q22	36.2	5.15	0.33	4.82	0.000019	55
12q24	119.9	D12S79	D16S404	16p13	7.7	5.69	0.29	5.41	0.000022	78
8q24	126.9	D8S514	D7S669	7q21	48.2	5.66	0.55	5.10	0.000083	76
20p13	0	D20S117	D16S503	16q21	73.2	5.13	0.13	5.00	0.000055	81
5p13	50.3	D5S426	D16S3103	16p12	21.7	5.94	0.87	5.07	0.0001	97

to confer stronger susceptibility to cancer. Therefore, modeling gene–gene interactions in linkage analysis may improve the power to detect chromosomal regions that harbor these disease susceptibility genes, as demonstrated in our previous linkage study of *PTEN* and *CDKN1B* (Xu et al. 2004). The ability to systematically model interactions in linkage analyses has largely been hindered by a lack of appropriate analytical methods, inadequate computing power, and small sample sizes. In this study, we applied the OSA method to identify prostate cancer susceptibility genes by systematically modeling gene–gene interactions in linkage analyses over all possible pairs of loci across the genome among 426 prostate cancer families. We found evidence (defined by a target chromosome-wide/reference marker-specific $P \leq 0.0001$) of an epistatic interaction for six sets of loci. These results did not meet our strict conservative criterion, $P = 5.6 \times 10^{-6}$, to be deemed globally statistically significant. However, it would require very strong interaction effects to overcome the burden of reaching global-statistical significance after accounting for the large number of tests. Our results should be useful for future prostate cancer interaction studies because we identified specific *a priori* locus combinations that should be followed up. We did, however, find evidence for significant/suggestive interactions between chromosome 17q22 loci and loci on chromosomes 4q35/11q14, respectively, using the reduced threshold level of statistical significance assumed for this *a priori* plausible region. Clearly our results will require confirmation

from independent studies and from the identification of specific genes underlying this linkage evidence. Our approach to systematically assess gene–gene interactions across the genome represents a potentially powerful alternative approach for gene identification for complex diseases using linkage studies.

Increased ability to identify disease susceptibility genes by modeling gene–gene interaction using the OSA method was demonstrated in our linkage analysis of the *PTEN* and *CDKN1B* regions in the subset of 188 families from Johns Hopkins Hospital (Xu et al. 2004). It is worth noting that this gene–gene interaction between the *PTEN* and *CDKN1B* regions was also observed in our search for novel interactions across the entire genome among the full set of 426 prostate cancer families. For example, when conditioning on the linkage result at *CDKN1B*, the LOD score at the *PTEN* region was 3.50, which is significantly increased compared to the unconditional LOD score of 0.15 ($P = 0.007$). Although this result did not meet our stringent criteria for genome-wide significance, this specific result can be considered significant in the case of candidate genes for which there is strong biological evidence.

The increased evidence for linkage between conditional and unconditional analyses in the six sets of loci suggested herein is consistent with the hypothesis of an epistatic interaction of two prostate cancer susceptibility genes; i.e., mutations in two genes are needed to increase prostate cancer risk. The significant LOD_{delta} , however, may also represent type I errors (false positives). The

Table 3 Comparisons of top six strongest two-locus epistatic interaction regions between two independent sets of families

Target regions	Reference regions	JHU families (N = 188)				P-value	Michigan, Tampere, Umeå families (N = 238)			
		LOD scores			P-value		LOD scores			P-value
		Conditional	Unconditional	Difference			Conditional	Unconditional	Difference	
11q13	13q32	2.27	0.29	1.98	0.06	5.69	1.28	4.41	0.0002	
22q13	21q22	2.39	0.77	1.62	0.03	3.43	0.00	3.43	0.0001	
12q24	16p13	4.21	0.78	3.43	0.0006	3.60	0.33	3.27	0.0015	
8q24	7q21	4.60	0.41	4.19	0.002	2.15	0.25	1.90	0.07	
20p13	16q21	3.06	0.78	2.28	0.01	3.16	0.00	3.16	0.0008	
5p13	16p12	3.15	0.27	2.88	0.009	3.57	0.69	2.88	0.02	

Table 4 Results of two-locus interaction involving the 17q22 region

Target region			Reference region			LOD scores			P-value	Number of families included
Cyto-band	Position (cM)	Nearest marker	Marker	Cyto-band	Position (cM)	Conditional	Unconditional	Difference		
<i>Linkage analysis at chromosome 17 conditional on the results at other region in the genome</i>										
17q22	70	D17S787	D1S2836	1q44	281.5	6.18	3.11	3.07	0.0033	236
17q22	72.2	D17S787	D2S391	2p21	62.7	6.18	3.14	3.04	0.0065	310
17q22	72.22	D17S787	D2S151	2q22	144.4	5.44	3.14	2.31	0.0026	288
17q22	75.9	D17S787	D4S403	4p15	21.2	5.82	3.01	2.81	0.0053	294
17q22	74.4	D17S787	D8S1771	8p21	49.3	5.75	3.08	2.67	0.0079	228
17q22	72.2	D17S787	D11S901	11q14	83.4	6.06	3.14	2.92	0.0029	301
<i>Linkage analysis in the genome conditional on the linkage result at 17q22</i>										
4q35	202.2	D4S426	D17S787	17q22	74.4	4.13	0.62	3.52	0.0018	55
11q14	82.3	D11S901	D17S787	17q22	74.4	4.51	0.50	4.01	0.0045	54

correlation between results at neighboring loci (due to linkage between adjacent reference markers) and the large number of tests makes the probability of a type I error in this type of two-locus genome-wide scan difficult to assess. It is important to note that the OSA *P*-values we report are adjusted for all possible subsets of families and multiple target points along each chromosome. The reported *P*-values do not account for the multiple tests corresponding to the numerous different target chromosome/reference marker combinations. Thus, our type I error rate for the hypothesis of an interaction is protected and does not suffer from multiplicity due to evaluating a variable number of subsets of families and a large number of target loci per chromosome, but does suffer from the multiple tests corresponding to the different target chromosome/reference marker combinations. Support for interaction effects for all six sets of suggestive loci was observed in two independent subsets of families; however, it is important to note that the consistent support from these two subsets of families was not independent from the overall results. Finally, this study was based on the largest number of prostate cancer families studied to date and therefore, the results are less likely to be influenced by variation due to small sample size. Furthermore, this set of families is enriched for characteristics consistent with inherited prostate cancer susceptibility; more than 86% of these families have either ≥ 4 affected members in a family or have a mean age at diagnosis of < 65 years.

In each of the six sets of interactions identified in this study, evidence for interactions was supported from both directions of the two involved regions, although one direction was stronger than the reverse direction. For example, the LOD_{delta} was 5.41 ($P < 0.000022$) at 12q24 when conditioning on the reference marker at 16p13. Conversely, the LOD_{delta} was 3.18 ($P = 0.002$) at 16p13 when conditioning on the reference marker at 12q24. There are at least two possible interpretations for this observation. The observation remains consistent with a two-locus epistatic interaction model. Under this model, linkage at locus 1 would increase among families that are strongly linked to a locus 2, and vice versa. The

different strengths in supporting the epistatic interaction between the two reversible directions may be due to fluctuations of LOD scores at these two regions. LOD scores that are calculated in regions of linkage for multifactorial-susceptibility genes are highly influenced by sampling and multiple confounding factors such as phenocopies and incomplete penetrance. Alternatively, the observation of stronger evidence for an interaction in one direction of two involved loci may suggest that one of the two loci is a major gene while the other is a modifying gene. Under this model, linkage at a modifying gene (locus 1) would increase among families that are strongly linked to a major gene (locus 2); however, evidence for an interaction in the reverse direction may not be as strong.

We did not report the findings of searching for gene-gene interactions under a heterogeneity model using the OSA analysis in this study. We reasoned that the OSA method has less power to detect heterogeneity interactions. When implementing the OSA method to model a heterogeneity interaction, families would be ranked based on linkage evidence at the reference region, from smallest (most negative) to highest. The notion is that if a family is not linked at the reference region then, the family would be more likely to be linked at the target region. This is likely not an optimal approach to assess heterogeneity interactions in complex diseases. In linkage analysis of complex diseases, while a positive LOD score of a region within a specific family suggests the disease is likely linked to the region; a negative LOD score in a family could be observed due to a variety of reasons, including incomplete penetrance, phenocopies, and other genes (besides those at the target region).

The purpose of this study is to mine valuable data that were generated as part of the largest genome-wide screen for prostate cancer susceptibility genes reported to date. Our rationale to explore for gene-gene interactions is justified because interactions of multiple genes are widely hypothesized to influence risk for prostate cancer. The development of new analytical methods makes it feasible to systematically explore genome-wide interactions. While it is difficult to determine the true

statistical significance of these findings, results of our study have generated specific hypotheses that can be tested in follow-up studies. In particular, we have identified six sets of loci with interactions and two regions, 4q35 and 11q14, that appear to interact with 17q22. In the future, such hypotheses can be readily tested among a set of over 1,200 prostate cancer families currently being assembled by the International Consortium for Prostate Cancer Genetics.

References

- Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A et al (1994) Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* 368:258–261
- Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC (1992) Mendelian inheritance of familial prostate cancer. *Proc Natl Acad Sci USA* 89:3367–3371
- Conlon EM, Goode EL, Gibbs M, Stanford JL, Badzioch M, Janer M, Kolb S, Hood L, Ostrander EA, Jarvik GP, Wijsman EM (2003) Oligogenic segregation analysis of hereditary prostate cancer pedigrees: evidence for multiple loci affecting age at onset. *Int J Cancer* 105:630–635
- Cui J, Staples MP, Hopper JL, English DR, McCredie MR, Giles GG (2001) Segregation analyses of 1,476 population-based Australian families affected by prostate cancer. *Am J Hum Genet* 68:1207–1218
- Cunningham JM, McDonnell SK, Marks A, Hebbing S, Anderson SA, Peterson BJ, Slager S, French A, Blute ML, Schaid DJ, Thibodeau SN (2003) Genome linkage screen for prostate cancer susceptibility loci: results from the Mayo Clinic Familial Prostate Cancer Study. *Prostate* 57:335–346
- Edwards S, Meitz J, Eles R, Evans C, Easton D, Hopper J, Giles G, Foulkes WD, Narod S, Simard J, Badzioch M, Mahle L (2003) Results of a genome-wide linkage analysis in prostate cancer families ascertained through the ACTANE consortium. *Prostate* 57:270–279
- Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R (1993) The human mutator gene homolog MSH2 and its association with hereditary non-polyposis colon cancer. *Cell* 75:1027–1038
- Gillanders EM, Xu J, Chang BL, Lange EM, Wiklund F, Bailey-Wilson JE, Baffoe-Bonnie A, Jones M, Gildea D, Riedesel E, Albertus J, Isaacs SD, Wiley KE, Mohai CE, Matikainen MP, Tammela TL, Zheng SL, Brown WM, Rokman A, Carpten JD, Meyers DA, Walsh PC, Schleutker J, Gronberg H, Cooney KA, Isaacs WB, Trent JM (2004) Combined genome-wide scan for prostate cancer susceptibility genes. *J Natl Cancer Inst* 96:1240–1247
- Gong G, Oakley-Girvan I, Wu AH, Kolonel LN, John EM, West DW et al (2002) Segregation analysis of prostate cancer in 1,719 white, African-American and Asian-American families in the United States and Canada. *Cancer Causes Control* 13:471–482
- Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M et al (1991) Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 66:589–600
- Gronberg H, Bergh A, Damber JE, Emanuelsson M (2000) Cancer risk in families with hereditary prostate carcinoma. *Cancer* 89:315–321
- Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M (2004) Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol* 27:53–63
- Hsieh C-L, Oakley-Girvan I, Balise RR, Halpern J, Gallagher RP, Wu AH, Kolonel LN, O'Brien LE, Lin IG, Van Den Berg DJ, Teh CZ, West DW, Whittemore AS (2001) A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. *Am J Hum Genet* 69:148–158
- Isaacs WB, Xu J, Walsh PC (2001) Hereditary prostate cancer. In: Chung LWK, Isaacs WB, Simons JW (eds) *Prostate cancer, biology, genetics, and the new therapeutics*. Humana Press, Totowa, NJ, pp 13–28
- Janer M, Friedrichsen DM, Stanford JL, Badzioch MD, Kolb S, Deutsch K, Peters MA, Goode EL, Welti R, DeFrance HB, Iwasaki L, Li S, Hood L, Ostrander EA, Jarvik GP (2003) Genomic scan of 254 hereditary prostate cancer families. *Prostate* 57:309–319
- Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D et al (1991) Identification of FAP locus genes from chromosome 5q21. *Science* 253:661–665
- Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179–1188
- Land H, Parada LF, Weinberg RA (1983) Cellular oncogenes and multistep carcinogenesis. *Science* 222:771–778
- Lange EM, Gillanders EM, Davis CC, Brown WM, Campbell JK, Jones M, Gildea D, Riedesel E, Albertus J, Freas-Lutz D, Markey C, Giri V, Dimmer JB, Montie JE, Trent JM, Cooney KA (2003) Genome-wide scan for prostate cancer susceptibility genes using families from the University of Michigan prostate cancer genetics project finds evidence for linkage on chromosome 17 near BRCA1. *Prostate* 57:326–334
- Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, Sistonen P, Aaltonen LA, Nystrom-Lahti M et al (1993) Mutations of a mutS homolog in hereditary non-polyposis colorectal cancer. *Cell* 75:1215–1225
- Maier C, Herkommer K, Hoegel J, Vogel W, Paiss T (2005) A genomewide linkage analysis for prostate cancer susceptibility genes in families from Germany. *Eur J Hum Genet* 13:352–360
- Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W et al (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266:66–71
- Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, Mori T (1997) Germline mutation of MSH6 as the cause of hereditary non-polyposis colorectal cancer. *Nat Genet* 17:271–272
- Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM et al (1994) Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 371:75–80
- Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN (1998) Evidence for autosomal dominant inheritance of prostate cancer. *Am J Hum Genet* 62:1425–1438
- Schleutker J, Baffoe-Bonnie AB, Gillanders E, Kainu T, Jones MP, Freas-Lutz Jr, Markey C, Gildea D, Riedesel E, Albertus J, Gibbs KD Jr, Matikainen M, Koivisto PA, Tammela T, Bailey-Wilson JE, Trent JM, Kallioniemi OP (2003) Genome-wide scan for linkage in Finnish hereditary prostate cancer (HPC) families identifies novel susceptibility loci at 11q14 and 3p25-26. *Prostate* 57:280–289
- Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD et al (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* 274:1371–1374
- Verhage BA, Baffoe-Bonnie AB, Baglietto L, Smith DS, Bailey-Wilson JE, Beaty T et al (2001) Autosomal dominant inheritance of prostate cancer: a confirmatory study. *Urology* 57:97–101
- Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. *Biometrics* 50:118–127
- Wiklund F, Gillanders EM, Albertus JA, Bergh A, Damber JE, Emanuelsson M, Freas-Lutz DL, Gildea DE, Goransson I, Jones MS, Jonsson BA, Lindmark F, Markey CJ, Riedesel EL, Stenman E, Trent JM, Grönberg H (2003) Genome-wide scan of Swedish families with hereditary prostate cancer: suggestive evidence of linkage at 5q11.2 and 19p13.3. *Prostate* 57:290–297

- Witte J, Goddard K, Conti D, Elston R, Lin J, Suarez B, Broman K, Burmester J, Weber J, Catalona W (2000) Genomewide scan for prostate cancer-aggressiveness loci. *Am J Hum Genet* 67:92–99
- Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G (1995) Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378:789–792
- Xu J, Gillanders EM, Isaacs SD, Chang BL, Wiley KE, Zheng SL, Jones M, Gildea D, Riedesel E, Albertus J, Freas-Lutz D, Markey C, Meyers DA, Walsh PC, Trent JM, Isaacs WB (2003) Genome-wide scan for prostate cancer susceptibility genes in the Johns Hopkins hereditary prostate cancer families. *Prostate* 57:320–325
- Xu J, Langefeld CD, Zheng SL, Gillanders EM, Chang BL, Isaacs SD, Williams AH, Wiley KE, Dimitrov L, Meyers DA, Walsh PC, Trent JM, Isaacs WB (2004) Interaction effect of PTEN and CDKN1B chromosomal regions on prostate cancer linkage. *Hum Genet* 115:255–262