Dense genome-wide SNP linkage scan in 301 hereditary prostate cancer families identifies multiple regions with suggestive evidence for linkage

Janet L. Stanford1,2,*, Liesel M. FitzGerald1,†, Shannon K. McDonnell3,‡, Erin E. Carlson3, Laura M. McIntosh1, Kerry Deutsch4, Lee Hood4, Elaine A. Ostrander5 and Daniel J. Schaid3

1Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, 2Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA 98195, USA, 3Division of Biostatistics, Mayo Clinic, Rochester, NY 55905, USA, 4Institute for Systems Biology, Seattle, WA 98103, USA and 5Cancer Genetics Branch, NHGRI, National Institutes of Health, Bethesda, MD 20892, USA

Received December 3, 2008; Revised and Accepted February 25, 2009

The search for susceptibility loci in hereditary prostate cancer (HPC) has proven challenging due to genetic and disease heterogeneity. Multiple risk loci have been identified to date, however few loci have been replicated across independent linkage studies. In addition, most previous analyses have been hampered by the relatively poor information content provided by microsatellite scans. To overcome these issues, we have performed linkage analyses on members of 301 HPC families genotyped using the Illumina SNP linkage panel IVb. The information content for this panel, averaged over all pedigrees and all chromosomes, was 86% (range 83–87% over chromosomes). Analyses were also stratified on families according to disease aggressiveness, age at diagnosis and number of affected individuals to achieve more genetically homogeneous subsets. Suggestive evidence for linkage was identified at 7q21 (HLOD = 1.87), 8q22 (KCLOD = 1.88) and 15q13–q14 (HLOD = 1.99) in 289 Caucasian families, and nominal evidence for linkage was identified at 2q24 (LOD = 1.73) in 12 African American families. Analysis of more aggressive prostate cancer phenotypes provided evidence for linkage to 11q25 (KCLOD = 2.02), 15q26 (HLOD = 1.99) and 17p12 (HLOD = 2.13). Subset analyses according to age at diagnosis and number of affected individuals also identified several regions with suggestive evidence for linkage, including a KCLOD of 2.82 at 15q13–q14 in 128 Caucasian families with younger ages at diagnosis. The results presented here provide further evidence for a prostate cancer susceptibility locus on chromosome 15q and demonstrate the power of utilizing high information content SNP scans in combination with homogenous collections of large prostate cancer pedigrees.

INTRODUCTION

In 2008, an estimated 186,320 US men were diagnosed with prostate cancer (1), and ~5% of these patients will have developed their disease due to an inherited mutation(s) in one or more hereditary prostate cancer (HPC) gene(s) (2). Over the past decade, research efforts have focused on mapping chromosomal regions that may harbor genes responsible for HPC, with anticipation that such discoveries will provide insight into the molecular genetics of both the inherited and sporadic forms of the disease. Family-based genome-wide linkage studies have highlighted over two dozen putative loci for HPC with significant or suggestive signals, although not all have been confirmed...
and most have not yet yielded clear evidence of specific genes or mutations, highlighting the challenges of finding HPC genes (reviewed in 3–5). The fact that there is both locus and disease heterogeneity makes the search for HPC loci difficult. Recently, analytic strategies aimed at developing homogenous subsets as well as refined phenotypic definitions have proven helpful. A family-based linkage study followed by fine-mapping and then association-based analyses is one strategy that led to the recent discoveries of prostate cancer susceptibility loci on chromosomes 8q and 17q (6, 7), with strong confirmatory evidence (8–13).

The approach of focusing on patients diagnosed with more clinically significant prostate cancer has also proven useful as a way to reduce disease heterogeneity and potentially improve power. Several studies have searched for genetic susceptibility loci by focusing on comparatively more aggressive disease phenotypes (14–24). Gleason score alone was used to define aggressive prostate cancer in most of these studies, except for three in which other clinical characteristics were included in the definition (14, 19, 23). The ability to stratify affected patients based on clinical features has led to the identification of a number of putative loci highlighted by more than one study, e.g. 4p (16,21), 7q (18,24), 19q (20,21,24) and 22q (14,23).

The results described above highlight some of the challenges and recent successes in finding prostate cancer predisposition loci. Similar to most previously published linkage studies, our prior genome-wide scan of 254 HPC families was based on simple tandem repeat polymorphisms and was limited by the overall low information content of markers (61%) and the sparse genome coverage, with an average marker spacing of 8.1 cM (25). It has previously been demonstrated that by utilizing the greater density single nucleotide polymorphism (SNP) arrays, investigators can detect linkage peaks otherwise missed by microsatellite scans (26). In order to take advantage of the higher information content of these arrays (an increase from 61 to 86% in our data) and to try and reduce genetic heterogeneity, we have completed a genome-wide linkage scan with 5867 SNPs in 2072 and reduce genetic heterogeneity, we have completed a genome-wide linkage scan with 5867 SNPs in 2072 arrays (an increase from 61 to 86% in our data) and to try and reduce genetic heterogeneity, we have completed a genome-wide linkage scan with 5867 SNPs in 2072 individuals from 307 families, incorporating clinical features of prostate cancer in affected men to create a more refined disease phenotype for linkage analyses.

RESULTS

Of the 307 families, Native American and Asian/Pacific Islander families (n = 3), as well as families that were uninformative for linkage (n = 3) were excluded. Hence, the linkage analyses were focused on 301 pedigrees, which are summarized in Table 1. A total of 2072 individuals were genotyped: 994 affected men, 581 unaffected men and 497 women. Also, we were able to construct genotypes for 174 deceased affected men. The 301 families included Caucasian (n = 289, including three Hispanic) and African American (n = 12) families. Twenty of the Caucasian families were thought to be Ashkenazi Jewish as the majority of family members reported their religious affiliation as Jewish and that their parents and/or grandparents were from Central or Eastern Europe. Among the Caucasian pedigrees, 126 had at least two affected men with an aggressive phenotype, 128 had a younger age of disease onset (<65 years) and 124 had five or more affected men (Table 1).

Figure 1 presents the LOD scores for Caucasian families when considering any prostate cancer as the phenotype. LOD scores of at least 1.86 were considered ‘suggestive’ for linkage (27), and these are summarized in Table 2. The most striking linkage signal was detected on chromosome 15q13–q14, with a dominant HLOD of 1.99 between SNPs rs343913 and rs2033610 (31.3–31.5 Mb). There was also suggestive evidence for linkage at chromosome 8q22, with the KCLOD of 1.88 between SNPs rs1449233 and rs1483457 (93.6–94.2 Mb). In addition, a recessive HLOD of 1.87 was detected at 7q21 under Model 2 between SNPs rs2030711 and rs1029847 (80.1–81.4 Mb). Figure 2 presents LOD scores for African American families for the any prostate cancer phenotype. The strongest linkage signal was a recessive HLOD of 1.73 on chromosome 2q24, between rs1990760 and rs1424937 (162.8–169.3 Mb). No other chromosome attained LOD scores > 1.5.

Figure 3 presents the LOD scores for African American families for the aggressive prostate cancer phenotype. LOD scores of at least 1.86 are presented in Table 2. The strongest linkage signal was a recessive HLOD of 2.13 at chromosome 17p12, between SNPs rs1984661 and rs2215054 (12.6–12.8 Mb). Suggestive evidence for linkage was also provided by a KCLOD of 2.02 on chromosome 11q25, between SNPs rs1630675 and rs471982 (132.6–133.5 Mb). Finally, a dominant HLOD of 1.99 was found on chromosome 15q26, between SNPs rs6598554 and rs2045112 (97.2–98.1 Mb).

Subset analyses were performed on the Caucasian families, first considering any prostate cancer as the phenotype and then considering only an aggressive prostate cancer phenotype. Results from these stratified analyses are summarized in Table 3. Because of multiple testing, only subsets with a LOD score ≥2.5 are shown (see Materials and Methods). For the any prostate cancer phenotype, the strongest signal was a KCLOD of 2.82 on chromosome 15q13–q14 in pedigrees with a younger age at diagnosis. But, there was no evidence for linkage heterogeneity across age strata (heterogeneity P = 0.09). There were two hints of linkage in families with an older age at diagnosis. Under the recessive model, an HLOD of 2.63 was detected on chromosome 2q21 (heterogeneity P = 0.02), and an HLOD of 2.48 was detected on chromosome 7q21 (heterogeneity P = 0.09). There was no suggestive evidence for linkage in the subset analyses defined by the number of affected men within a pedigree.

In the second set of stratified analyses, where the phenotype was defined as more aggressive prostate cancer, the largest signal was a KCLOD of 2.90 on chromosome 17q25 in families with a younger age of diagnosis (Table 3). But there was no evidence of heterogeneity across age strata for either parametric model. There were two observations of suggestive linkage in aggressive pedigrees with an older age at diagnosis. The first was on chromosome 18q21 with a KCLOD of 2.52 between SNPs rs869224 and rs1145315 (49.2–49.9 Mb; heterogeneity P = 0.01). The second observation was on chromosome 8q11 with a KCLOD of 2.46 between SNPs rs898520 and rs718251 (51.3–52.9 Mb; heterogeneity P = 0.04). Under a recessive model, an HLOD of 2.83
on chromosome 17p12 in pedigrees with at least five affected men was observed between SNPs rs1984661 and rs2215054 (12.6–12.8 Mb; heterogeneity $P = 0.07$).

A few other signals of note were observed (data not shown). In the subset analyses for any prostate cancer phenotype, pedigrees with five or more affected men had a KCLOD of 2.16 on chromosome 8q24 (flanking SNPs: rs2833 and rs4870888, 124.1–125.2 Mb). In the subset analyses for the aggressive phenotype, a recessive HLOD of 2.16 was observed for chromosome 22q12 in pedigrees with fewer than five affected men (flanking SNPs: rs714027 and rs4444, 28.9–29.5 Mb, heterogeneity $P = 0.01$). Finally, analysis of families of Jewish heritage ($n = 20$) highlighted two linkage peaks on chromosome 3p26 (KCLOD = 2.44) and 7q21 (KCLOD = 2.18).

### DISCUSSION

Previously, we reported on a genome-wide microsatellite scan of 254 HPC families that identified regions of linkage on chromosomes 6p, 7q and 17p (25). However, evidence for linkage was modest in these regions and also in previously implicated HPC regions. In order to increase power and decrease disease heterogeneity, the present study involved a SNP genome-wide linkage scan in 289 Caucasian HPC families using both any prostate cancer and a more aggressive prostate cancer phenotype, and in 12 African American HPC families using an any prostate cancer phenotype. Caucasian HPC families were further stratified according to age at diagnosis and number of affected men within each family. In the any prostate cancer analysis, suggestive evidence for linkage was observed at 7q21, 8q22 and 15q13–q14, and a further locus at 2q21 was highlighted in families with a median age of diagnosis $\geq 65$ years. Among Caucasian families, the analysis of 126 families with a more aggressive phenotype detected three loci, 11q25, 15q26 and 17p12 that were distinct from the any prostate cancer analysis, suggesting that these loci may play a specific role in the development of more clinically significant disease. A further three loci at 8q11, 17q25 and 18q21 were highlighted in the more aggressive subset analyses.

Replication of findings is the gold standard in the search for genes involved in common complex diseases such as prostate cancer. In that regard, the results presented here stand up to this criterion with loci detected both in any and aggressive prostate cancer analyses replicating several previous findings by our own and other research groups. The linkage peak at 7q21 corresponds with earlier results published by this group (25,28) and others (16,19,24). As previous results suggested that this region may harbor a genetic mutation that is particularly pertinent to prostate cancer susceptibility in families of Jewish heritage (28), we analyzed these families separately. A slightly stronger linkage signal was identified on chromosome 7q just centromeric to the peak in the overall analysis (KCLOD 2.18), as well as a signal at 3p26.1 (KCLOD 2.44), which was not observed in the previous study (28). Currently, fine-mapping at the 7p21 locus is ongoing in an effort to characterize the underlying susceptibility variant.

The linkage peak identified at 8q22 in the any prostate cancer analysis is broad and will require fine-mapping to determine whether it includes multiple susceptibility loci or a single risk locus. This peak is quite a distance centromeric to the 8q24 region, which is currently generating a great deal of interest not only in prostate cancer studies but also in studies of numerous other cancers (29–31). However, in our analysis of families with five or more affected men, a signal at 8q24 (KCLOD 2.2) was identified $\sim 3$ Mb centro-
meric to the previously described Region 1 (29). These findings are consistent with the recent suggestion that there are likely multiple risk loci across chromosome 8q (10,32); and hence, we are currently genotyping SNPs in the 8q24 region previously found to be associated with prostate cancer risk in the PROGRESS families, in order to determine their contribution to hereditary forms of the disease.

The most striking linkage signal in the any prostate cancer analysis was observed at chromosome 15q13–q14, with evidence from several sequential SNPs. This region was also highlighted in the subset analysis of patients with a younger age at diagnosis, strengthening the likelihood that this may be a genetically inherited susceptibility locus, as an earlier age of onset is a hallmark of genetically inherited cancer (33). This locus has also been highlighted in two previous prostate cancer linkage analyses of multiplex sibships from the USA (34) and families from Germany (35), as well as a recent linkage study of colorectal cancer (36), suggesting a possible pleiotrophic effect similar to that of the 8q24 region. Interestingly, different markers delineate the peak linkage signal at 15q13–q14 in our study and the two previous prostate cancer microsatellite linkage analyses; however, they all fall within the same gene, FMN1. Three additional microsatellite scans and one Affymetrix 10K SNP scan have also presented nominal evidence for linkage to regions in close proximity. A peak was previously identified at 15q13 in Caucasian prostate cancer families (37), at 15q12 in aggressive prostate cancer families (17), at 15q11 in families from the International Consortium for Prostate Cancer Genetics (38) and at 15q21 in a large Caucasian pedigree (39).

It is interesting to note that a signal at 15q13 was not observed in our previous genome-wide microsatellite scan (25), and that the strongest linkage signal detected in that study, at 6p22, was only of nominal significance in this analysis. Although the original linkage peak at 6p22 could have been a false positive, there is still a weak signal in this same region in the present SNP analysis. It may be that the 6p22 region contains a susceptibility locus that is responsible for disease in a proportion of the original 254 families but with the additional power provided in the current study, the contribution of this locus appears diminished. In a study by Schaid et al. (26) comparing the performance of microsatellite markers to SNPs, it was also found that the greater information content provided by the SNP array allowed identification of linkage peaks that would have otherwise gone undetected. In addition to the enhanced density and greater information content, since the initial PROGRESS genome-wide scan of 254 families, the study has been expanded considerably with an additional 47 families and 141 affected men included in the current analyses. Therefore, our power to detect suscepti-
bility loci has been substantially enhanced since the initial PROGRESS linkage study.

To increase genetic homogeneity, analyses were repeated in only those families with at least two cases of aggressive prostate cancer. Suggestive evidence for linkage (LOD ≥ 1.86) was found on chromosomes 11q25, 15q26 and 17p12 in these families. Although linkage to 11q25 has been previously observed in families with prostate cancer (35), to the best of our knowledge, prior studies have not reported evidence for linkage to 11q25 or 15q26 in analyses of more aggressive prostate cancer families. Evidence for linkage to 17p12 was originally presented by Tavtigian et al. (40) but since then, subsequent linkage scans have failed to replicate this finding. The linkage peak presented here falls just telomeric of ELAC2 and encompasses another interesting candidate gene, RICH2.

### Table 2. Suggestive LOD scores (LOD ≥ 1.86) for the any prostate cancer and the more aggressive prostate cancer phenotypes in Caucasian families

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>cM</th>
<th>Left flanking SNP (bp)</th>
<th>Right flanking SNP (bp)</th>
<th>Modela</th>
<th>KCLODb</th>
<th>Dominant HLOD</th>
<th>Recessive HLOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any prostate cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7q21</td>
<td>93.27</td>
<td>rs2030711 (80057878)</td>
<td>rs1029847 (81431156)</td>
<td>1</td>
<td>0.89</td>
<td>1.31</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8q22</td>
<td>100.26</td>
<td>rs1449233 (93622634)</td>
<td>rs1483457 (94214382)</td>
<td>1</td>
<td>1.88</td>
<td>1.02</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.99</td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td>15q13–q14</td>
<td>25.70</td>
<td>rs343913 (31133168)</td>
<td>rs2033610 (31542203)</td>
<td>1</td>
<td>1.50</td>
<td>1.99</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1.94</td>
<td></td>
<td>0.69</td>
</tr>
<tr>
<td>Aggressive prostate cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11q25</td>
<td>158.63</td>
<td>rs1630675 (132601535)</td>
<td>rs470982 (133472588)</td>
<td>1</td>
<td>2.02</td>
<td>1.02</td>
<td>1.05</td>
</tr>
<tr>
<td>15q26</td>
<td>123.51</td>
<td>rs6598554 (97172788)</td>
<td>rs2045112 (98120495)</td>
<td>1</td>
<td>1.23</td>
<td>1.99</td>
<td>0.74</td>
</tr>
<tr>
<td>17p12</td>
<td>38.99</td>
<td>rs1984661 (12591502)</td>
<td>rs2215054 (12779734)</td>
<td>1</td>
<td>0.72</td>
<td>0.61</td>
<td>2.13</td>
</tr>
</tbody>
</table>

*aModel 1 = affecteds only (see text), Model 2 = Smith model (two liability classes; see text).

bKCLOD = Kong and Cox.

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Genome-wide linkage results for the any prostate cancer phenotype in 12 African American Families. DomHLOD (red) and RecHLOD (blue) are the dominant and recessive heterogeneity LOD scores for model 1 (affecteds only). KCLOD (black) is the Kong and Cox exponential allele sharing model LOD score.
note that the association with this locus is greater in families with five or more affected men (HLOD 2.83).

Previous linkage analyses have identified numerous regions linked to aggressive prostate cancer, several of which have been confirmed in subsequent studies [5q31–q35 (20,21,24), 6p22 (15,17,19), 7q31–q33 (18,24,41,42), 9q22 (15,42), 19q12–q13 (20,21,42) and 22q11 (14,23)]. In the aggressive phenotype analyses presented here, we only found nominal evidence for two previously highlighted loci, 9q22 (dominant HLOD 1.09) and 22q12 (dominant HLOD 1.21). Subset analyses of men diagnosed at a younger age (KCLOD 2.33) and families with fewer than five affected men (recessive HLOD 2.16) provided stronger evidence for linkage to 22q12, although the LODs were below our cut-point of 2.5. The large number of loci highlighted in aggressive prostate cancer linkage studies to date, and the fact that only a few loci have been replicated here and in other aggressive prostate cancer linkage studies, raises the question as to whether stratifying by an aggressive phenotype is, in fact, achieving the goal of disease homogeneity. However, different criteria to define a more aggressive phenotype have been used across studies. The majority of earlier genome-wide scans used only Gleason score and identified loci at 5q, 7q and 19q (20,21,24,41,42). More recent scans like ours used a number of clinical characteristics in addition to Gleason score, including tumor stage, pretreatment PSA level and/or death from metastatic prostate cancer and consistently identified regions of linkage to 6p and 22q (14,15,17,19,23). A subtle difference in this composite clinical definition is that the study presented here used a Gleason score of 7 (4 + 3) to define a more aggressive phenotype, whereas others have simply included all tumors with a Gleason score of 7. As outcomes (recurrence and prostate cancer-specific mortality) are worse for patients with a Gleason of 7 (4 + 3) score versus those with 7 (3 + 4) tumors (43–45), the distinct loci observed to be linked with aggressive disease in the current analyses could be due to our more precise criteria for defining a more aggressive prostate cancer phenotype.

In summary, three loci are highlighted as candidate prostate cancer susceptibility loci in this dense SNP genome-wide scan, 7q21, 8q22 and 15q13–q14. Chromosome 8q clearly contains at least one susceptibility locus for prostate cancer and a number of groups, including our own (10), have consistently confirmed an association between multiple genetic variants in the 8q24 region and risk of prostate cancer. The most interesting result from this study is the region of suggestive linkage on chromosome 15q13–q14, which was the strongest linkage signal overall and in the early onset disease subgroup. It is possible that an underlying susceptibility gene in this region may have a pleiotropic effect, as this is also a candidate region for colon cancer (36). Several plausible candidate genes reside in the 15q13–q14 linkage region.
including FMN1, GREM1 and AVEN, however, fine-mapping will be needed to narrow the region in anticipation of exploring specific genes and regulatory regions.

This study also highlights the utility of using a more clinically refined phenotype of prostate cancer that incorporates Gleason score and other clinical features of disease. It is interesting to note that the evidence for linkage was stronger for some loci in stratified analyses, which may indicate that the stratifications are successfully producing more homogeneous subsets of families. These three regions (11q25, 15q26 and 17p12) warrant further follow-up in datasets that have comparable clinical data.

SNP genome-wide association scans have proven successful in identifying common variants associated with prostate cancer in recent case–control studies. However, to detect rarer susceptibility loci, large pedigrees remain a powerful resource. The SNP genome-wide linkage scans reported here and previously (26,39), demonstrate the increase in information content and ability to identify susceptibility loci that the combination of dense SNP scans and large pedigrees provides. It is our hope that fine-mapping will provide further evidence of linkage to some of the regions highlighted in this study.

\section*{MATERIALS AND METHODS}

\section*{Hereditary prostate cancer families}

Families included in these analyses are participating in the Prostate Cancer Genetic Research Study (\textit{PROGRESS}), a large collection of HPC families ascertained from across North America. To date, 307 families have been enrolled with an average of seven members per family participating. Of these 307 families, 301 were included in the final analyses presented here (see Results section) and 254 of these same families were included in an earlier linkage analysis study based on microsatellite markers (25). There are an additional 47 families and 141 affected men included in the present analyses. All prostate cancer survivors, men without prostate cancer aged 40 and older, and selected women expected to be informative for linkage were invited to join \textit{PROGRESS}. Ascertainment and eligibility criteria for the study have been described previously (25). Initial data collection involved:

\begin{enumerate}
\item completion of a baseline participant survey that incorporated questions about demographics, medical history, family structure and cancer history and lifestyle factors,
\item providing a blood sample for genotyping and
\item consent to obtain medical records for prostate cancer patients. In addition to the baseline survey, two follow-up surveys have been completed by participating family members to update medical and family cancer history, including new diagnoses of prostate cancer. There are currently 1483 (living and deceased) men with prostate cancer in the 307 families, and 1085 (73\%) of these have been confirmed by medical records or death certificates. Most unconfirmed diagnoses represent deceased men from older generations for whom medical records and death certificates are not available. On the basis of review of medical records obtained to date (n = 961), 100\% have confirmed the prostate cancer diagnosis.
\end{enumerate}

\section*{Clinical data}

Medical records obtained for 961 participating men diagnosed with prostate cancer were used to abstract data on Gleason grade and score, stage of disease (localized, regional, distant) and serum prostate-specific antigen (PSA) level at diagnosis; data were coded according to protocols developed by the Surveillance, Epidemiology and End Results cancer registry program (46). For deceased men diagnosed with prostate cancer, death certificates were collected to confirm underlying cause (prostate cancer-specific death or death due to another cause), date and age at death. These clinical data have been used to define a comparatively more aggressive prostate cancer phenotype, which includes one or more of the following criteria: regional or distant stage at diagnosis, Gleason score of 7 (4 + 3) or 8–10, diagnostic PSA level of 20 ng/ml or higher, or death from metastatic prostate cancer before age 65 years (23).

\section*{Genotyping and quality control}

DNA samples from 2154 individuals in 307 HPC families were prepared and plated at Fred Hutchinson Cancer Research Center and then shipped to the Center for Inherited Disease Research (CIDR) for genotyping using the Illumina Linkage

\begin{table}[h]
\centering
\caption{Summary of subset linkage results with LOD scores \(\geq 2.5\) in at least one stratum in Caucasian families}
\begin{tabular}{lllllllll}
\hline
Chromosome & Strata & cM & Left flanking SNP (bp) & Right flanking SNP (bp) & Model\textsuperscript{a} & KCLOD & Dom HLOD & Rec HLOD \\
& & & & & & & (Het P-value)\textsuperscript{b} & (Het P-value)\textsuperscript{b} \\
\hline
\textbf{Any prostate cancer} & & & & & & & & \\
2q21 & Age \(\geq 65\) & 146.26 & rs1836669 (133958941) & rs20349 (134292719) & 1 & 1.31 & \(-0.01\) (0.87) & \textbf{2.58} (0.02) \\
& & & & & & & & \\
7q21 & Age \(\geq 65\) & 93.27 & rs2030711 (80057878) & rs1029847 (81431156) & 2 & 0.85 & 1.85 (0.08) & \textbf{2.48} (0.09) \\
15q13–q14 & Age \(< 65\) & 25.70 & rs343913 (31133168) & rs2033610 (31542203) & 1 & \textbf{2.82} & \textbf{2.22} (0.10) & 1.86 (0.27) \\
& & & & & & & & \\
\textbf{Aggressive prostate cancer} & & & & & & & & \\
8q11 & Age \(\geq 65\) & 64.28 & rs898520 (51251445) & rs718251 (52877076) & 1 & \textbf{2.46} & 0.92 (0.11) & \textbf{1.84} (0.04) \\
17q25 & Age \(< 65\) & 117.94 & rs1000821 (71945598) & rs717571 (72132418) & 1 & \textbf{2.90} & 1.00 (0.61) & 2.25 (0.45) \\
17p12 & 5+ PC & 38.99 & rs1984661 (12591502) & rs2215054 (12779734) & 1 & 1.17 & 0.93 (0.58) & \textbf{2.83} (0.07) \\
18q21 & Age \(\geq 65\) & 78.54 & rs869224 (49231796) & rs1145315 (49942953) & 1 & \textbf{2.52} & \textbf{2.38} (0.01) & 1.99 (0.11) \\
\hline
\end{tabular}
\textsuperscript{aModel 1 = affecteds only (see text), Model 2 = Smith model (two liability classes; see text).}
\textsuperscript{bHeterogeneity P-value for likelihood ratio statistic testing for difference between the two subsets.}
\end{table}
Panel IVb. An overall error rate of 0.01% was observed in the 98 blind duplicate pairs submitted to CIDR. Data were not analyzed for four of the individuals due to poor sample performance. The remaining 2150 genotyped individuals all had call rates >98%. SNPs were excluded from analyses if they had: (1) a SNP call rate <95%, (2) minor allele frequency (MAF) <0.05, (3) departure from Hardy–Weinberg equilibrium $P < 0.001$, (4) a single allele or (5) insufficient map information. In addition, 16 markers on the Y chromosome and 27 markers in the pseudo-autosomal regions on the X chromosome were not analyzed. Of the original 5867 SNPs genotyped, 5742 were available for analyses.

Four pedigrees had excessive Mendelian inheritance errors of the SNPs, apparently from misspecified relationships; eight subjects from these pedigrees were problematic. To determine agreement between putative and inferred relationships, PREST software was used (47). Exclusion of one or more members from these four pedigrees resolved these inconsistencies. To further evaluate potential misspecified relationships in all pedigrees, PREST (47) was run again. Results were filtered ($P < 0.01$) to highlight any relative pairs that deviated from the expected IBD sharing. This resulted in the exclusion of 12 subjects from nine families. Subsequently, checks for Mendelian consistency using PEDCHECK (48) were performed. A total of 146 Mendelian errors were detected in 119 subjects from 82 families and removed from further analysis.

To verify self-reported race and to evaluate differences in allele frequencies among the major racial groups, we created principal components using tag-SNPs. Plots of the first two principal components demonstrated clear cluster differences for Caucasians and African Americans, as well as for Native American and Asian/Pacific Islander families ($n = 3$; data not shown). Clusters for Caucasian and Hispanic families were close. Owing to these differences, linkage analyses were performed for the pool of self-reported Caucasian and Hispanic families ($n = 289$, henceforth referred to as Caucasian) and separately for African American families ($n = 12$). The remaining three families were excluded from linkage analyses.

**Linkage analysis**

Both parametric and non-parametric allele-sharing linkage analyses were performed using Merlin software (49). The parametric LOD scores were computed using Merlin software (49). The parametric LOD scores were calculated using the Kong and Cox exponential allele sharing model score (KCLOD). Linkage information content was estimated using the program Merlin, by the use of the entropy information described by Kruglyak et al. (51).

Although Merlin can analyze SNPs that are in linkage disequilibrium (LD) by treating them as multi-allelic markers and employing the EM algorithm (52), this option requires an excessive amount of memory and analytical time for large pedigrees. For this reason, we accounted for LD by first choosing tag-SNPs, and then analyzed the tag-SNPs without LD. For the Caucasian families, the maximal set of independent subjects (490 subjects) was used to calculate LD. Tag-SNPs were identified using LD select (53) with a low $r^2$ threshold ($r^2 \leq 0.10$); tags for each bin were chosen considering Illumina QC measures, MAF and SNP call rate. A total of 4743 tag-SNPs were selected with a median intermarker distance over all chromosomes of 0.60 cM (range: 0.001–5.97 cM). The median MAF among all tag-SNPs was 0.40 (range: 0.05–0.50), whereas the median call rate was 99% (range: 97–100%). The overall information content was excellent, with a median of 86% (range: 63–90%). Marker allele frequencies were estimated across the pool of all subjects, ignoring genetic relationships.

For the 12 African American families, genotype data were available on only 32 subjects. To determine more reliable allele frequencies and measures of LD, we used the 60 HapMap Yoruba founders. Using a threshold of $r^2 \leq 0.10$, 1884 tag-SNPs were selected with a median intermarker distance of 1.34 cM (range: 0–14.45 cM) across all chromosomes. The median MAF among all selected tag-SNPs was 0.31 (range: 0.01–0.50), whereas the median call rate was 99% (range: 97–100%). The number of selected tag-SNPs was smaller than that used for Caucasian families; this may be due to the smaller sample size used to estimate LD, which could in turn inflate LD estimates and result in fewer LD bins. The information content for the African American families had a median of 63% (range: 17–90%).

To ‘fit’ pedigree data into the memory limits of Merlin software, trimming of family members was conducted in an iterative fashion. Uninformative subjects were first removed. Next, genotyped terminal subjects with unknown phenotype were removed if both parents were available and genotyped. Finally, subjects were iteratively removed by preferentially removing: (1) subjects with unknown affection status, (2) unaffacted subjects or (3) affected men if necessary. Trimming was performed on each pedigree to obtain a maximum bit size of 24. If more than one affected informative man had to be trimmed, then the pedigree was split into two sub-pedigrees for analyses instead of trimming.

Linkage analyses were first performed considering any prostate cancer as the phenotype. To search for loci associated with a more clinically aggressive prostate cancer phenotype, analyses were also performed in 126 Caucasian families with two or more men with aggressive disease. For these analyses, men with aggressive prostate cancer were coded as affected and men with non-aggressive prostate cancer were re-coded as being of unknown affection status. Analyses involving the phenotype of aggressive prostate cancer were run using only Model 1 in order to avoid assigning non-aggressive affected men the same penetrance values as those for men with...
unknown affection status. Because of the small number of African American families, these families were not analyzed for the aggressive prostate cancer phenotype.

Caucasian pedigrees were also stratified into subsets according to the following characteristics: (1) age at diagnosis (<65 versus ≥65 years, with 65 years being the median age at diagnosis) and (2) number of men affected with prostate cancer (<5 versus ≥5). These subsets were analyzed for ‘any prostate cancer’ and ‘more aggressive prostate cancer’ phenotypes. To evaluate whether there was statistically significant heterogeneous cancer and more aggressive prostate cancer’ phenotypes. Work by Weeks et al. analyzes a more stringent LOD score threshold was applied. Stratification of linkage should not depend on the stratification factor. If there is no significant difference in linkage across subsets, then this suggests that the interpretation of linkage should not depend on the stratification factor. Owing to the multiple testing present in these stratified analyses, a more stringent LOD score threshold was applied. Work by Weeks et al. (54) suggests that a penalty of 0.1–1.0 should be added when examining different genetic models and different phenotype classifications. We took the mid-range (0.5), and simplified this by rounding up to a LOD score threshold of 2.5 (1.86 + 0.5 = 2.36). We view this threshold as a way to highlight regions worth further study, but not formal statistical control for multiple testing.

ACKNOWLEDGEMENTS

We thank all the men and women who are participating in the PROGRESS study for their time, effort and cooperation. We also thank the study staff for help with ongoing data collection and processing. We acknowledge the Prostate Cancer Foundation and the Intramural Program of the National Human Genome Research Institute.

Conflict of Interest statement. The authors declare that they have no conflicts of interest.

FUNDING

This work was supported by the National Cancer Institute [grant numbers RO1 CA080122 and P50-CA097186] with additional support from the Fred Hutchinson Cancer Research Center. Genotyping services provided by the Center for Inherited Disease Research at Johns Hopkins University were supported by the National Institutes of Health [contract number N01-HG-65403].

REFERENCES


