

ARTICLE

# Genetic heterogeneity in Finnish hereditary prostate cancer using ordered subset analysis

Claire L Simpson<sup>1,7</sup>, Cheryl D Cropp<sup>1,7</sup>, Tiina Wahlfors<sup>2</sup>, Asha George<sup>1,3</sup>, MaryPat S Jones<sup>4</sup>, Ursula Harper<sup>4</sup>, Damaris Ponciano-Jackson<sup>4</sup>, Teuvo Tammela<sup>5</sup>, Johanna Schleutker<sup>2,6</sup> and Joan E Bailey-Wilson<sup>\*,1</sup>

Prostate cancer (PrCa) is the most common male cancer in developed countries and the second most common cause of cancer death after lung cancer. We recently reported a genome-wide linkage scan in 69 Finnish hereditary PrCa (HPC) families, which replicated the *HPC9* locus on 17q21-q22 and identified a locus on 2q37. The aim of this study was to identify and to detect other loci linked to HPC. Here we used ordered subset analysis (OSA), conditioned on nonparametric linkage to these loci to detect other loci linked to HPC in subsets of families, but not the overall sample. We analyzed the families based on their evidence for linkage to chromosome 2, chromosome 17 and a maximum score using the strongest evidence of linkage from either of the two loci. Significant linkage to a 5-cM linkage interval with a peak OSA nonparametric allele-sharing LOD score of 4.876 on Xq26.3-q27 ( $\Delta\text{LOD} = 3.193$ , empirical  $P = 0.009$ ) was observed in a subset of 41 families weakly linked to 2q37, overlapping the *HPCX1* locus. Two peaks that were novel to the analysis combining linkage evidence from both primary loci were identified; 18q12.1-q12.2 (OSA LOD = 2.541,  $\Delta\text{LOD} = 1.651$ ,  $P = 0.03$ ) and 22q11.1-q11.21 (OSA LOD = 2.395,  $\Delta\text{LOD} = 2.36$ ,  $P = 0.006$ ), which is close to *HPC6*. Using OSA allows us to find additional loci linked to HPC in subsets of families, and underlines the complex genetic heterogeneity of HPC even in highly aggregated families.

*European Journal of Human Genetics* (2013) 21, 437–443; doi:10.1038/ejhg.2012.185; published online 5 September 2012

**Keywords:** linkage analysis; ordered subset analysis; prostate cancer

## INTRODUCTION

In most Western countries, prostate cancer (PrCa) is one of the most commonly diagnosed cancers in men. In Finland in 2009, there were over 4500 cases diagnosed and 785 deaths,<sup>1</sup> making it the most commonly diagnosed cancer and the second most common cause of cancer death. There has been a concerted effort over many years to try and understand the mechanisms that drive the development and progression of prostate tumors as part of an attempt to improve detection and therapeutic interventions.

Family history has long been known to be a major risk factor<sup>2–6</sup> and many genetic studies have attempted to identify genetic variants that predispose men to development of the disease or contribute to the aggressiveness of the tumor. Many different studies consistently report PrCa as highly heritable.<sup>5,7–10</sup> At least 15 different loci have been linked to hereditary PrCa (HPC) through linkage analysis in highly aggregated families. So far, a few genes have been positively identified in the search for high-penetrance PrCa susceptibility loci, but the evidence does not suggest that risk alleles in these genes explain large proportions of either HPC or nonhereditary PrCa because the risk alleles at this loci appear to be quite rare. *HPC1/RNASEL* (1q23–25),<sup>11</sup> *HPC2/ELAC2* (17p)<sup>12</sup> and *MSRI* (8p22–23)<sup>13</sup> are genes with rare, high-penetrance risk alleles that have been found through sequencing under linkage peaks in HPC families. Multiple other risk loci, such as

*PCAP* (1q42.2–43),<sup>14,15</sup> *CAPB* (1p36),<sup>15,16</sup> *MYC* (8q24),<sup>17–19</sup> *HPC20* (20q13)<sup>20</sup> and *HPCX1* (Xq27–28),<sup>21</sup> have been implicated through linkage analysis in HPC families.

Common, low-penetrance polymorphisms on 3q, 8q, 10q, 11q, 17q, 19q and Xp have been consistently and repeatedly detected in genome-wide association studies (GWASs).<sup>17,22–32</sup> However, more work remains to be done to fully elucidate the role of heredity in the complex disease of PrCa. The causative genes responsible for these associations are still unknown.

Ordered subset analysis (OSA) is a widely used technique to address the genetic heterogeneity of many complex diseases and traits and to allow detection of gene–gene interactions.<sup>33</sup> We have previously reported genome-wide linkage analysis of hereditary PrCa in 69 highly aggregated Finnish families,<sup>34</sup> where we replicated a previously reported linkage to *HPC9* on 17q21-q22<sup>35,36</sup> and identified a novel locus on 2q37.2. Here we describe a secondary analysis of these data using OSA to condition on linkage to either or both linkage peaks to attempt to identify additional loci linked to this disease. Recently, functional variants in the *HOXB13* gene<sup>37</sup> have been proposed as candidates for the causal risk alleles in the 17q21-q22 region, making it quite important to condition on linkage to this region in the search for additional loci important to familial PrCa risk.

<sup>1</sup>Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA; <sup>2</sup>Institute of Biomedical Technology/BioMediTech, University of Tampere, Tampere, Finland and Finlab Laboratories, Tampere, Finland; <sup>3</sup>Fox Chase Cancer Center, Philadelphia, PA, USA; <sup>4</sup>Genomics Core/Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Rockville, MD, USA; <sup>5</sup>Department of Urology, Tampere University Hospital, University of Tampere, Tampere, Finland; <sup>6</sup>Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland

\*Correspondence: Dr JE Bailey-Wilson, Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, 333 Cassell Drive, Baltimore, MD 21224, USA. Tel: 443 740 2921; Fax: 443 740 2165; E-mail: jebw@nhgri.nih.gov

<sup>7</sup>These authors contributed equally to this work.

Received 30 March 2012; revised 25 June 2012; accepted 17 July 2012; published online 5 September 2012

## MATERIALS AND METHODS

### Families and genotyping

Sixty-nine multiplex Finnish families, all Caucasian, were included in the study. All of the 69 families had at least three confirmed cases of PrCa and 6 out of 69 families had  $\geq 5$  affected members. The detailed description of the families, our sample collection protocol and confirmation of diagnoses are presented elsewhere.<sup>38,39</sup> A total of 54 families were genotyped with microsatellite markers, 44 of these families from the microsatellite study, plus an additional 15 previously ungenotyped families were genotyped with SNP markers. Details regarding the DNA preparation, PCR conditions and allele-scoring techniques for the markers are described elsewhere.<sup>38</sup>

### Linkage analyses

Primary, nonparametric linkage (NPL) analyses were performed in GENEHUNTER-PLUS.<sup>40,41</sup> The X-chromosome version of GENEHUNTER-PLUS (v1.3) was used in X-chromosome analyses. The nonparametric affecteds-only linkage analyses included NPL scores from GENEHUNTER-PLUS using the 'all' option and allele-sharing LOD scores as developed by Kong and Cox<sup>40</sup> (performed by the ASM program in conjunction with GENEHUNTER-PLUS).

### OSA

To address the apparent genetic heterogeneity observed by many studies of HPC,<sup>34</sup> we used OSA to condition on NPL scores at one or both of the two linked loci. NPL methods are powerful to detect loci that contribute to risk in a large proportion of families, but less powerful when the proportion of linked families is small. By conditioning on one or both of our two already identified loci, we reduce heterogeneity and so increase power to detect linkage to other loci. Multipoint NPL scores calculated with GENEHUNTER-PLUS software were utilized as the ranking covariate in OSA to take advantage of the extended pedigree structure, ranking families by familywise NPL score. The OSA program first arranged the familywise NPL scores in an ascending order (low to high) and later in a descending order (high to low) to find an appropriate subset of families that maximized the evidence of linkage to other regions of the genome.<sup>33</sup> The 'optimal slice' that gave the maximum OSA LOD score determined a subset of 'adjacent' families in the covariate rankings (not necessarily including the end points), thus allowing families with extremely low or high covariate scores to be excluded from the linked subset. The OSA program generates this subset by taking the most significant of the two ordered ranking subsets and then sequentially dropping families from the top of the subset to see if an even more significant subset can be found, which does not necessarily include the tails of the covariate distribution.<sup>33</sup> The OSA results were then graphically compared with the multipoint GH results and the empirical *P*-values for the change in the LOD score were examined. An empiric probability was calculated to assess whether the observed OSA LOD score (based on ordering the families by the covariate) was significantly increased over the OSA LOD obtained from a random reordering of families. This empiric probability was computed using a permutation test, which randomized the order of the families 10000 times to determine the proportion of times the randomly ordered families gave an OSA LOD score greater than the observed OSA LOD score. This permutation test examines the hypothesis that the covariate-defined subset yields a higher LOD score than in the randomly assigned family subsets.<sup>33</sup> The OSA method maximizes the OSA LOD score over the subsets. Therefore, the maximum LOD score will always be at least as large as the overall LOD score in all families. This means that the distribution of the maximum OSA LOD score cannot be the same as the distribution of a LOD score, which has not been maximized. Evaluation of the OSA LOD scores for a subset has to be done in the context of the evidence for linkage in the entire sample and must account for the selection of a subset of the data. By evaluating the change in evidence in favor of linkage, we can account for both the baseline evidence and the nonrandom subsetting of the data.

We analyzed three variables, individual family NPL scores for the peak on 2q37, individual family NPL scores for the peak at 17q21-q22 and a maximum score, which was calculated as the larger of the two NPL scores. A *P*-value was considered significant at  $P \leq 0.025$ , to account for the two opposing ranking

methods for the covariate.<sup>33</sup> Family-based association analyses were also performed (Supplementary Methods).

## RESULTS

### Conditioning on linkage to chromosome 2

Conditioning on linkage to chromosome 2q37 yielded six loci with OSA LOD scores  $> 2$  and significant  $\Delta$ LOD scores by permutation testing (Table 1 and Figure 1). The highest OSA LOD score of 4.88 ( $\Delta$ LOD = 3.193,  $P = 0.009$ ) was on chromosome Xq26.3-q27.1 in an optimal-slice subset of 41 families with weak to moderate evidence of linkage to chromosome 2. Of these families, 18 had evidence of male-to-male transmission and 23 had no evidence of male-to-male transmission.

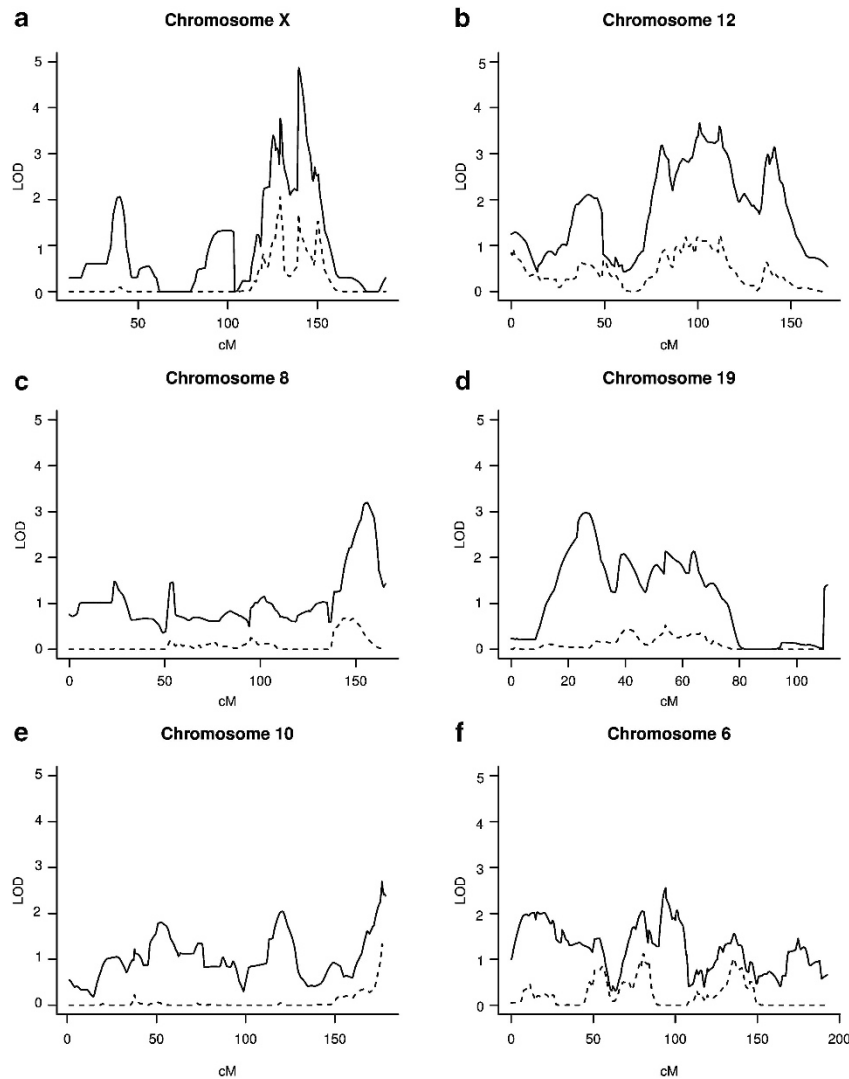
Other loci with high OSA LOD scores were on 12q21.1-q23.3 (OSA LOD = 3.05,  $\Delta$ LOD = 1.835,  $P = 0.02$ ) in a subset of 17 families unlinked to chromosome 2 and 8q24.22-q24.3 (OSA LOD = 3.195,  $\Delta$ LOD = 2.963,  $P = 0.02$ ) in an optimal-slice subset of 15 families with weak evidence of linkage to chromosome 2.

### Conditioning on linkage to chromosome 17

Conditioning on linkage to chromosome 17q21-q22 yielded four loci with OSA LOD scores  $> 2$  and significant  $\Delta$ LOD scores by permutation testing (Table 1 and Figure 2). The highest OSA LOD score was on 3q26.31-q27.1 (OSA LOD = 3.49,  $\Delta$ LOD = 2.39,  $P = 0.02$ ) in a subset of 47 families with moderate to no linkage to chromosome 17. The second strongest locus was on 12q14.2-q21.31 (OSA LOD = 3.23,  $\Delta$ LOD = 2.33,  $P = 0.02$ ) in a subset of 34 families with weak to no linkage to chromosome 17.

**Table 1** Maximum OSA LOD scores in descending order in the 69 families conditioning on linkage to chromosome 2, chromosome 17 or both

Chromosome	Position (cm)	Maximum OSA LOD	$\Delta$ LOD	<i>P</i> -value	Proximal-flanking marker	Distal-flanking marker
<i>Ascending by linkage to chromosome 2</i>						
12	111.98	3.05	1.83	0.02	rs1484828	rs1862032
10	116.50	2.04	2.01	0.01	rs2039305	rs1034178
<i>Optimal slice by linkage to chromosome 2</i>						
X	139.56	4.88	3.19	0.01	rs2254857	rs2485729
8	156.18	3.20	2.96	0.02	rs894358	rs879429
19	26.0	2.98	2.95	0.02	rs753842	rs892011
6	94.08	2.56	2.56	0.00	rs1213353	rs846773
<i>Descending by linkage to chromosome 17</i>						
8	141.68	2.57	2.01	0.02	rs755520	rs2318351
<i>Optimal slice by linkage to chromosome 17</i>						
3	185.03	3.49	2.39	0.02	rs2046350	rs2054172
12	82.35	3.23	2.33	0.02	rs1695105	rs1163016
5	127.50	2.02	2.02	0.02	rs890785	rs258825
<i>Ascending by linkage to chromosomes 2 and 17</i>						
12	82.35	2.66	1.75	0.02	rs1695105	rs1484828
18	55.0	2.54	1.65	0.03	rs2011596	rs1789501
<i>Optimal slice by linkage to chromosomes 2 and 17</i>						
22	1.0	2.40	2.36	0.01	rs7288876	rs374225
6	94.08	2.14	2.14	0.01	rs1213353	rs500204



**Figure 1** Nonparametric LOD score plot by chromosome for the most significant OSA LOD scores when conditioning on linkage to chromosome 2. LOD scores for the overall sample (dashed line) and the subset with maximum NPL score (solid line) at (a) 140 (b) 156 (c) 112 (d) 26 (e) 94 and (f) 117 cM.

### Conditioning on linkage to chromosomes 2 and 17

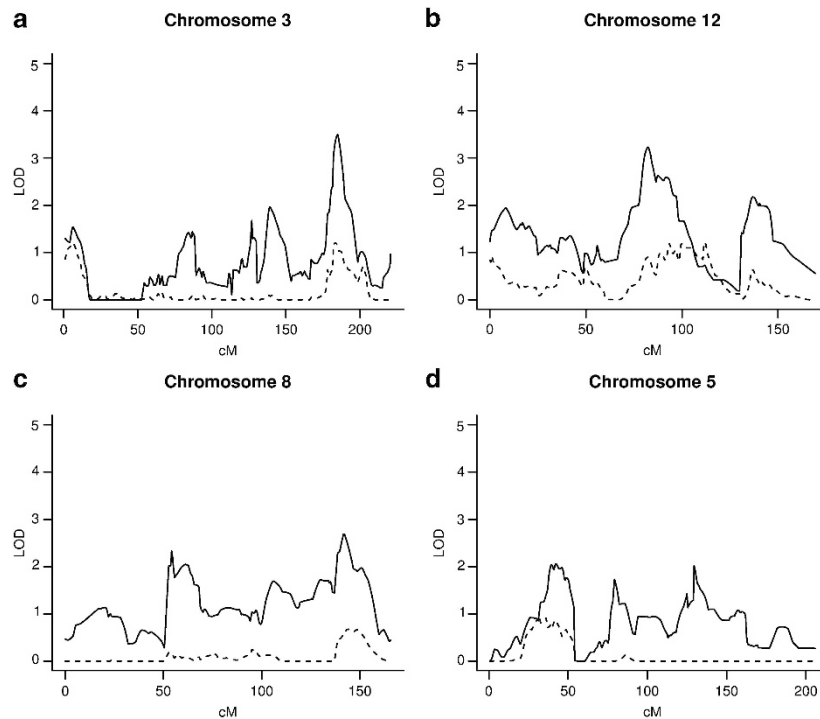
We then calculated the maximum family LOD score for chromosomes 2 and 17, so we could condition on linkage to either chromosome 2 or 17. Four loci had OSA LOD scores  $>2$  and had significant  $\Delta$ LOD scores by permutation testing (Table 1 and Figure 3). None were as strong as the scores found when conditioning on just one of the chromosomes, however, two loci were completely novel to this third analysis; 18q12.1-q12.2, which was not quite significant (OSA LOD = 2.54,  $\Delta$ LOD = 1.65,  $P = 0.03$ ) in the 38 families not linked to either locus, and 22q11.1-q11.21 (OSA LOD = 2.40,  $\Delta$ LOD = 2.36,  $P = 0.006$ ) in an optimal-slice subset of 12 families with weak evidence of linkage to either chromosome 2 or 17. Family-based association analysis results can be found in the Supplementary Results and Supplementary Table 1.

### DISCUSSION

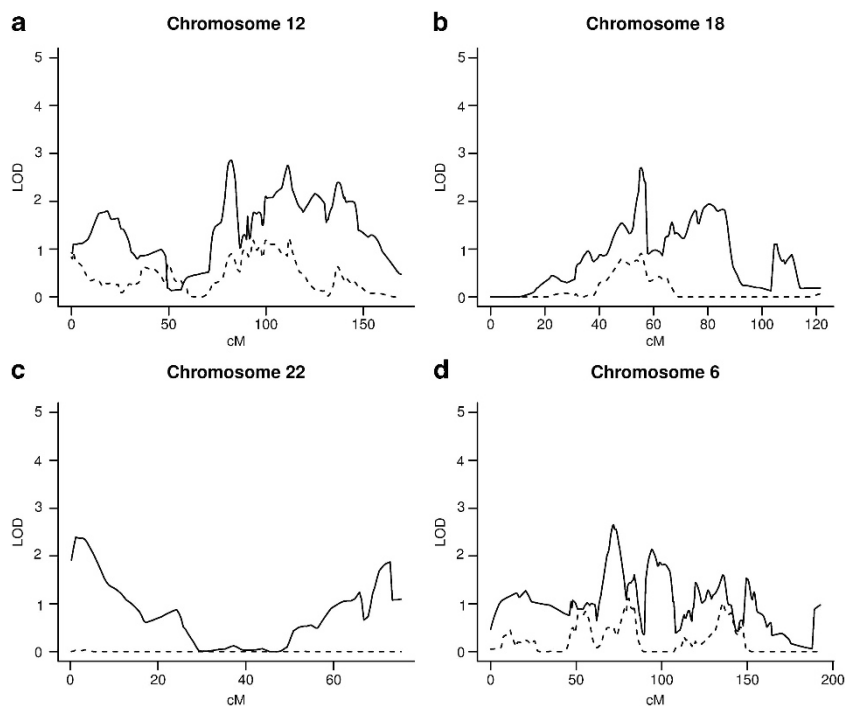
The considerable genetic heterogeneity of PrCa makes it difficult to identify the various genetic factors that contribute to the risk of developing the disease. Here we have used ordered subset linkage analysis in order to find additional loci that cannot be found with

other linkage methods. Inflation in the false-positive rate in these OSA analyses (induced by examining multiple family subsets for a given covariate) is controlled using a permutation test. Simulation studies performed by Hauser *et al*<sup>33</sup> show that the type 1 error rate is adequately controlled by the permutation procedure. We have not controlled for OSA comparisons over the different conditioning loci because some correlation may well exist between the loci. A Bonferroni correction of such a  $P$ -value would be very conservative. These are exploratory analyses and as such should be seen as a hypothesis-generating approach that requires follow-up and cross comparison with other studies. This is similar to the procedure of Cox *et al*,<sup>42</sup> who proposed the idea of examining the difference between overall and conditional LOD scores to evaluate the effects of epistasis or genetic heterogeneity, and Hauser *et al*<sup>33</sup> showed that this idea could also be applied to OSA.

Our OSA analysis conditioning on linkage to chromosome 2 revealed a subset of families with strong evidence of linkage to chromosome X. The OSA LOD score peak of 4.88 on chromosome Xq26.3-q27.1 was found in families with weak to moderate evidence of linkage to chromosome 2. The first documented X-linked PrCa



**Figure 2** Nonparametric LOD score plot by chromosome for the most significant OSA LOD scores when conditioning on linkage to chromosome 17. LOD scores for the overall sample (dashed line) and the subset with maximum NPL score (solid line) at (a) 185 (b) 82 (c) 142 and (d) 128 cM.



**Figure 3** Nonparametric LOD score plot by chromosome for the most significant OSA LOD scores when conditioning on linkage to chromosome 2 or 17. LOD scores for the overall sample (dashed line) and the subset with maximum NPL score (solid line) at (a) 82 (b) 55 (c) 1 and (d) 94 cM.

susceptibility locus in the region Xq27-q28 (*HPCX*) was identified in 1998 in a study of 360 families with multiple cases of PrCa.<sup>21</sup> From this study, a subset of Finnish families with no male-to-male transmission and late age of onset of PrCa (> 65 years) demonstrated

stronger evidence for linkage to Xq27-q28 than other families belonging to the complete data set.<sup>39</sup> In the current data set, there was no significant overall evidence for linkage to Xq26-q28, but this region was significant in the subset of families that showed weak to



moderate evidence of linkage to chromosome 2. Thus, this OSA procedure is successfully detecting an X-linked subset in the presence of strong heterogeneity. Our results for chromosome X lie within the vicinity of *HPCX* and the genomic coordinates for *SPANXC* (Xq27.1), a member of the *SPANX* family, which encode for differentially expressed testis-specific proteins that localize to various subcellular compartments. Originally, the *SPANX* family of genes was proposed as putative candidates for the *HPCX*-susceptibility locus.<sup>43</sup> However, recent studies have not identified mutations in any of these genes that definitely account for PrCa risk in X-linked HPC families.<sup>44</sup>

Also found by conditioning on linkage to chromosome 2 was a significant peak on 12q21-q23, with an OSA LOD of 3.67 in families unlinked to chromosome 2. Nominal but nonsignificant evidence of linkage to 12q was detected in the complete data set with a heterogeneity LOD (HLOD) of 2.34<sup>34</sup> on 12q22-q23, which overlaps with the present region. Again, the OSA technique was able to identify the subset of families strongly linked to this region. This locus spans 30 Mb and is close to a locus reported by the International Consortium for Prostate Cancer Genetics (ICPCG) when performing OSA using 406 marker loci (distributed across the genome) as the OSA-ranking covariate in 426 PrCa families from Johns Hopkins Hospital, University of Michigan, University of Umeå and University of Tampere.<sup>45</sup> Some of the families included here were part of this ICPCG study. This 12q locus also overlaps with the locus recently found in African Americans.<sup>46</sup> Gain and loss of 12q has been reported in many studies of prostate tumors.<sup>47,48</sup>

A third significant locus found by conditioning on linkage to chromosome 2 was a peak on 8q24.22-q24.3 in families with weak evidence of linkage to chromosome 2. There was not even a suggestive evidence of linkage to this region in the overall data set, but OSA was able to identify the 15 families showing linkage to 8q24.22-q24.3. Amplification of 8q24 in PrCa has been consistently reported.<sup>49-51</sup> This 7.5-Mb locus overlaps with the 8q24 region, which has been reported in linkage and GWASs of PrCa<sup>17,23,24,29,32</sup> and has been replicated in many studied populations.<sup>22,25,26,52-68</sup> The prostate stem cell antigen maps to this region<sup>69</sup> as does the oncogene *MYC*. The region has been extensively studied and is relatively gene-poor. A number of functional enhancers have been identified in the region<sup>70,71</sup> and some of these enhancers have been shown to regulate *MYC*.

The strongest signal found by conditioning on linkage to chromosome 17 was on 3q26.3-q27.1 in families with moderate to no linkage to chromosome 17. This region did not exhibit even a nominal evidence of linkage in the complete data set. The locus spans 8 Mb and contains a number of genes known to be upregulated in cancer cells. Amplification of this region is associated with PrCa<sup>72-75</sup> and it contains the gene *PIK3CA*, which is associated with other tumors including breast, ovarian, colorectal and gastric cancers.<sup>76-80</sup>

Some of these loci were found in subsets that already had some evidence of linkage to chromosome 2 or 17. These loci may not have been detected in the original linkage analyses because the effect on risk due to these loci is smaller than the effect on risk of the chromosome 2 or 17 loci, leading to insufficient power. Although several of the loci detected by this OSA analysis have been seen in other studies, they were not significant in these data even when using HLOD scores. The two novel loci on 18q and 22q may only be detectable after conditioning on other loci with larger effects on PrCa risk. There may also be a statistical interaction on risk between these loci and the conditioning loci but the current analyses cannot test explicitly for this. Current thinking suggests that for many common cancers, multiple loci may be acting together to account for the high

risk of developing the disease in highly aggregated families. If the additional loci found here are truly affecting risk of developing PrCa, then they may be having a modifying role on the complex process of neoplastic transformation, perhaps through multiple different mechanisms. Alternatively, because these are highly selected families, it is possible that each locus has independent, noninteracting effects on PrCa risk, and risk genotypes are segregating for multiple loci in some of the same families because of the mode of ascertainment.

For OSA to have sufficient power, there must be adequate correlation between evidence of linkage and the levels of the OSA covariate. In our analysis, the covariate was itself a linkage signal, and so the extent to which this varies between families will be closely correlated with the overall genetic heterogeneity in the sample.

In order to further investigate these loci, useful approaches to refining the signal and identifying causative variants would be targeted sequencing of the region in selected individuals or whole-genome sequencing (WGS). Linkage results such as these were helpful in identifying the candidate causal mutations in the *HOXB13* gene on chromosome 17q21.<sup>37</sup> Thus, these linkage results may be very useful in WGS studies of individuals with a family history of PrCa, because they will allow us to prioritize these regions in the search for rare variants with major effects on PrCa risk.<sup>81-83</sup>

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We would like to express our gratitude to the families who participated in this study. This work was supported in part by the Intramural Program of the National Human Genome Research Institute, National Institutes of Health, the Academy of Finland (116437, 251074), the Sigrid Juselius Foundation, the Finnish Cancer Organizations and the Competitive Research Funding of the Tampere University Hospital (9M094). CDC is the recipient of an NHGRI Health Disparities Research Fellowship. The International Consortium for Prostate Cancer Genetics (ICPCG), including the consortium's Data Coordinating Center (DCC), is made possible by a grant from the National Institutes of Health U01 CA89600 (to William B Isaacs). SNP genotyping services were provided by the Center for Inherited Disease Research (NIH Contract: N01-HG-65403).

- 1 Finnish Cancer Registry. Cancer incidence and mortality in Finland: Cancer Statistics. 2009.
- 2 Woolf CM: An investigation of the familial aspects of carcinoma of the prostate. *Cancer* 1960; **13**: 739-744.
- 3 Meikle AW, Smith JA, West DW: Familial factors affecting prostatic cancer risk and plasma sex-steroid levels. *Prostate* 1985; **6**: 121-128.
- 4 Steinberg GD, Carter BS, Beaty TH, Childs B, Walsh PC: Family history and the risk of prostate cancer. *Prostate* 1990; **17**: 337-347.
- 5 Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC: Mendelian inheritance of familial prostate cancer. *Proc Natl Acad Sci USA* 1992; **89**: 3367-3371.
- 6 Narod SA, Dupont A, Cusan L *et al*: The impact of family history on early detection of prostate cancer. *Nat Med* 1995; **1**: 99-101.
- 7 Page WF, Braun MM, Partin AW, Caporaso N, Walsh P: Heredity and prostate cancer: a study of World War II veteran twins. *Prostate* 1997; **33**: 240-245.
- 8 Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN: Evidence for autosomal dominant inheritance of prostate cancer. *Am J Hum Genet* 1998; **62**: 1425-1438.
- 9 Suarez BK, Lin J, Burmester JK *et al*: A genome screen of multiplex sibships with prostate cancer. *Am J Hum Genet* 2000; **66**: 933-944.
- 10 Verhage BA, Baffoe-Bonnie AB, Baglietto L *et al*: Autosomal dominant inheritance of prostate cancer: a confirmatory study. *Urology* 2001; **57**: 97-101.
- 11 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**: 181-184.
- 12 Tavtigian SV, Simard J, Teng DH *et al*: A candidate prostate cancer susceptibility gene at chromosome 17 p. *Nat Genet* 2001; **27**: 172-180.
- 13 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**: 321-325.

- 14 Berthon P, Valeri A, Cohen-Akenine A *et al*: Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2-43. *Am J Hum Genet* 1998; **62**: 1416–1424.
- 15 Berry R, Schaid DJ, Smith JR *et al*: Linkage analyses at the chromosome 1 loci 1q24-25 (HPC1), 1q42.2-43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer. *Am J Hum Genet* 2000; **66**: 539–546.
- 16 Gibbs M, Stanford JL, McIndoe RA *et al*: Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. *Am J Hum Genet* 1999; **64**: 776–787.
- 17 Amundadottir LT, Sulem P, Gudmundsson J *et al*: A common variant associated with prostate cancer in European and African populations. *Nat Genet* 2006; **38**: 652–658.
- 18 Camp NJ, Farnham JM, Allen-Brady K, Cannon-Albright LA: Statistical recombination mapping in extended high-risk Utah pedigrees narrows the 8q24 prostate cancer locus to 2.0 Mb. *Prostate* 2007; **67**: 1456–1464.
- 19 Lu L, Cancel-Tassin G, Valeri A *et al*: Chromosomes 4 and 8 implicated in a genome wide SNP linkage scan of 762 prostate cancer families collected by the ICPG. *Prostate* 2011.
- 20 Berry R, Schroeder JJ, French AJ *et al*: Evidence for a prostate cancer-susceptibility locus on chromosome 20. *Am J Hum Genet* 2000; **67**: 82–91.
- 21 Xu J, Meyers D, Freije D *et al*: Evidence for a prostate cancer susceptibility locus on the X chromosome. *Nat Genet* 1998; **20**: 175–179.
- 22 Eeles RA, Kote-Jarai Z, Giles GG *et al*: Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 2008; **40**: 316–321.
- 23 Yeager M, Orr N, Hayes RB *et al*: Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007; **39**: 645–649.
- 24 Freedman ML, Haiman CA, Patterson N *et al*: Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA* 2006; **103**: 14068–14073.
- 25 Al Olama AA, Kote-Jarai Z, Giles GG *et al*: Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat Genet* 2009; **41**: 1058–1060.
- 26 Gudmundsson J, Sulem P, Gudbjartsson DF *et al*: Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* 2009; **41**: 1122–1126.
- 27 Yeager M, Chatterjee N, Ciampa J *et al*: Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2009; **41**: 1055–1057.
- 28 Eeles RA, Kote-Jarai Z, Al Olama AA *et al*: Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* 2009; **41**: 1116–1121.
- 29 Gudmundsson J, Sulem P, Manolescu A *et al*: Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007; **39**: 631–637.
- 30 Gudmundsson J, Sulem P, Steinthorsdottir V *et al*: Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007; **39**: 977–983.
- 31 Gudmundsson J, Sulem P, Rafnar T *et al*: Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 2008; **40**: 281–283.
- 32 Haiman CA, Patterson N, Freedman ML *et al*: Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007; **39**: 638–644.
- 33 Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M: Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol* 2004; **27**: 53–63.
- 34 Cropp CD, Simpson CL, Wahlfors T *et al*: Genome-wide linkage scan for prostate cancer susceptibility in Finland: evidence for a novel locus on 2q37.3 and confirmation of signal on 17q21-q22. *Int J Cancer* 2011; **129**: 2400–2407.
- 35 Lange EM, Gillanders EM, Davis CC *et al*: Genome-wide scan for prostate cancer susceptibility genes using families from the University of Michigan prostate cancer genetics project finds evidence for linkage on chromosome 17 near BRCA1. *Prostate* 2003; **57**: 326–334.
- 36 Lange EM, Robbins CM, Gillanders EM *et al*: Fine-mapping the putative chromosome 17q21-22 prostate cancer susceptibility gene to a 10 cm region based on linkage analysis. *Hum Genet* 2007; **121**: 49–55.
- 37 Ewing CM, Ray AM, Lange EM *et al*: Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 2012; **366**: 141–149.
- 38 Schleutker J, Baffoe-Bonnie AB, Gillanders E *et al*: Genome-wide scan for linkage in Finnish hereditary prostate cancer (HPC) families identifies novel susceptibility loci at 11q14 and 3p25-26. *Prostate* 2003; **57**: 280–289.
- 39 Schleutker J, Matikainen M, Smith J *et al*: A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: frequent HPC linkage in families with late-onset disease. *Clin Cancer Res* 2000; **6**: 4810–4815.
- 40 Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–1188.
- 41 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- 42 Cox NJ, Frigge M, Nicolae DL *et al*: Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nat Genet* 1999; **21**: 213–215.
- 43 Kouprina N, Pavlicek A, Noskov VN *et al*: Dynamic structure of the SPANX gene cluster mapped to the prostate cancer susceptibility locus HPCX at Xq27. *Genome Res* 2005; **15**: 1477–1486.
- 44 Kouprina N, Noskov VN, Solomon G *et al*: Mutational analysis of SPANX genes in families with X-linked prostate cancer. *Prostate* 2007; **67**: 820–828.
- 45 Chang BL, Lange EM, Dimitrov L *et al*: Two-locus genome-wide linkage scan for prostate cancer susceptibility genes with an interaction effect. *Hum Genet* 2006; **118**: 716–724.
- 46 Bonilla C, Hooker S, Mason T, Bock CH, Kittles RA: Prostate cancer susceptibility Loci identified on chromosome 12 in African Americans. *PLoS One* 2011; **6**: e16044.
- 47 Jiang M, Li M, Fu X *et al*: Simultaneously detection of genomic and expression alterations in prostate cancer using cDNA microarray. *Prostate* 2008; **68**: 1496–1509.
- 48 Kasahara K, Taguchi T, Yamasaki I, Kamada M, Yuri K, Shuin T: Detection of genetic alterations in advanced prostate cancer by comparative genomic hybridization. *Cancer Genet Cytogenet* 2002; **137**: 59–63.
- 49 El Gedaily A, Bubendorf L, Willi N *et al*: Discovery of new DNA amplification loci in prostate cancer by comparative genomic hybridization. *Prostate* 2001; **46**: 184–190.
- 50 Tsuchiya N, Kondo Y, Takahashi A *et al*: Mapping and gene expression profile of the minimally overrepresented 8q24 region in prostate cancer. *Am J Pathol* 2002; **160**: 1799–1806.
- 51 van Dekken H, Alers JC, Damen IA *et al*: Genetic evaluation of localized prostate cancer in a cohort of forty patients: gain of distal 8q discriminates between progressors and nonprogressors. *Lab Invest* 2003; **83**: 789–796.
- 52 Severi G, Hayes VM, Padilla EJ *et al*: The common variant rs1447295 on chromosome 8q24 and prostate cancer risk: results from an Australian population-based case-control study. *Cancer Epidemiol Biomarkers Prev* 2007; **16**: 610–612.
- 53 Wang L, McDonnell SK, Slusser JP *et al*: Two common chromosome 8q24 variants are associated with increased risk for prostate cancer. *Cancer Res* 2007; **67**: 2944–2950.
- 54 Schumacher FR, Feigelson HS, Cox DG *et al*: A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer Res* 2007; **67**: 2951–2956.
- 55 Zheng SL, Sun J, Cheng Y *et al*: Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. *J Natl Cancer Inst* 2007; **99**: 1525–1533.
- 56 Robbins C, Torres JB, Hooker S *et al*: Confirmation study of prostate cancer risk variants at 8q24 in African Americans identifies a novel risk locus. *Genome Res* 2007; **17**: 1717–1722.
- 57 Sun J, Lange EM, Isaacs SD *et al*: Chromosome 8q24 risk variants in hereditary and non-hereditary prostate cancer patients. *Prostate* 2008; **68**: 489–497.
- 58 Cheng I, Plummer SJ, Jorgenson E *et al*: 8q24 and prostate cancer: association with advanced disease and meta-analysis. *Eur J Hum Genet* 2008; **16**: 496–505.
- 59 Beebe-Dimmer JL, Levin AM, Ray AM *et al*: Chromosome 8q24 markers: risk of early-onset and familial prostate cancer. *Int J Cancer* 2008; **122**: 2876–2879.
- 60 Salinas CA, Kwon E, Carlson CS *et al*: Multiple independent genetic variants in the 8q24 region are associated with prostate cancer risk. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 1203–1213.
- 61 Tan YC, Zeigler-Johnson C, Mittal RD *et al*: Common 8q24 sequence variations are associated with Asian Indian advanced prostate cancer risk. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 2431–2435.
- 62 Beuten J, Gelfond JA, Martinez-Fierro ML *et al*: Association of chromosome 8q variants with prostate cancer risk in Caucasian and Hispanic men. *Carcinogenesis* 2009; **30**: 1372–1379.
- 63 Pal P, Xi H, Guha S *et al*: Common variants in 8q24 are associated with risk for prostate cancer and tumor aggressiveness in men of European ancestry. *Prostate* 2009; **69**: 1548–1556.
- 64 Liu M, Kurosaki T, Suzuki M *et al*: Significance of common variants on human chromosome 8q24 in relation to the risk of prostate cancer in native Japanese men. *BMC Genet* 2009; **10**: 37.
- 65 Zheng SL, Hsing AW, Sun J *et al*: Association of 17 prostate cancer susceptibility loci with prostate cancer risk in Chinese men. *Prostate* 2010; **70**: 425–432.
- 66 Hooker S, Hernandez W, Chen H *et al*: Replication of prostate cancer risk loci on 8q24, 11q13, 17q12, 19q33, and Xp11 in African Americans. *Prostate* 2010; **70**: 270–275.
- 67 Chen M, Huang YC, Yang S *et al*: Common variants at 8q24 are associated with prostate cancer risk in Taiwanese men. *Prostate* 2010; **70**: 502–507.
- 68 Wokolorczyk D, Gliniewicz B, Stojewski M *et al*: The rs1447295 and DG8S737 markers on chromosome 8q24 and cancer risk in the Polish population. *Eur J Cancer Prev* 2010; **19**: 167–171.
- 69 Reiter RE, Gu Z, Watabe T *et al*: Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer. *Proc Natl Acad Sci USA* 1998; **95**: 1735–1740.
- 70 Jia L, Landan G, Pomerantz M *et al*: Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet* 2009; **5**: e1000597.
- 71 Sotelo J, Esposito D, Duhagon MA *et al*: Long-range enhancers on 8q24 regulate c-Myc. *Proc Natl Acad Sci USA* 2010; **107**: 3001–3005.
- 72 Sattler HP, Rohde V, Bonkhoff H, Zwergel T, Wullich B: Comparative genomic hybridization reveals DNA copy number gains to frequently occur in human prostate cancer. *Prostate* 1999; **39**: 79–86.
- 73 Jung V, Kindich R, Kamradt J *et al*: Genomic and expression analysis of the 3q25-q26 amplification unit reveals TLOC1/SEC62 as a probable target gene in prostate cancer. *Mol Cancer Res* 2006; **4**: 169–176.
- 74 Herness EA, Naz RK: A novel human prostate-specific gene-1 (HPG-1): molecular cloning, sequencing, and its potential involvement in prostate carcinogenesis. *Cancer Res* 2003; **63**: 329–336.
- 75 Sattler HP, Lensch R, Rohde V *et al*: Novel amplification unit at chromosome 3q25-q27 in human prostate cancer. *Prostate* 2000; **45**: 207–215.
- 76 Shayesteh L, Lu Y, Kuo WL *et al*: PIK3CA is implicated as an oncogene in ovarian cancer. *Nat Genet* 1999; **21**: 99–102.
- 77 Ma YY, Wei SJ, Lin YC *et al*: PIK3CA as an oncogene in cervical cancer. *Oncogene* 2000; **19**: 2739–2744.

- 78 Katso R, Okkenhaug K, Ahmadi K, White S, Timms J, Waterfield MD: Cellular function of phosphoinositide 3-kinases: implications for development, homeostasis, and cancer. *Annu Rev Cell Dev Biol* 2001; **17**: 615–675.
- 79 Mizoguchi M, Nutt CL, Mohapatra G, Louis DN: Genetic alterations of phosphoinositide 3-kinase subunit genes in human glioblastomas. *Brain Pathol* 2004; **14**: 372–377.
- 80 Pedrero JM, Carracedo DG, Pinto CM *et al*: Frequent genetic and biochemical alterations of the PI 3-K/AKT/PTEN pathway in head and neck squamous cell carcinoma. *Int J Cancer* 2005; **114**: 242–248.
- 81 Bailey-Wilson JE, Wilson AF: Linkage analysis in the next-generation sequencing era. *Hum Hered* 2011; **72**: 228–236.
- 82 Simpson CL, Justice CM, Krishnan M *et al*: Old lessons learned anew: family-based methods for detecting genes responsible for quantitative and qualitative traits in the Genetic Analysis Workshop 17 mini-exome sequence data. *BMC Proc* 2011; **5**(Suppl 9): S83.
- 83 Hinrichs AL, Suarez BK: Incorporating linkage information into a common disease/rare variant framework. *Genet Epidemiol* 2011; **35**(Suppl 1): S74–S79.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)