Published in final edited form as:
Eur Urol. 2012 March ; 61(3): 471–477. doi:10.1016/j.eururo.2011.10.047.

Evaluation of Multiple Risk-Associated Single Nucleotide Polymorphisms Versus Prostate-Specific Antigen at Baseline to Predict Prostate Cancer in Unscreened Men

Robert J. Klein[1,*], **Christer Hallden**[2,†], **Amit Gupta**[3,†], **Caroline J. Savage**[4], **Anders Dahlin**[5], **Anders Bjartell**[6], **Jonas Manjer**[7], **Peter T. Scardino**[3], **David Ulmert**[3,6], **Peter Wallström**[8], **Andrew J. Vickers**[4], and **Hans Lilja**[2,3,9]

[1]Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

[2]Department of Laboratory Medicine, Lund University, Skåne University Hospital, Malmö, Sweden

[3]Department of Surgery (Urology), Memorial Sloan-Kettering Cancer Center, New York, NY, USA

[4]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

[5]Department of Clinical Sciences (Medicine), Lund University, Skåne University Hospital, Malmö, Sweden

[6]Department of Clinical Sciences (Urological Cancers), Lund University, Skåne University Hospital, Malmö, Sweden

[7]Department of Clinical Sciences (Surgery), Lund University, Skåne University Hospital, Malmö, Sweden

[8]Department of Clinical Sciences (Nutritional Epidemiology), Lund University, Skåne University Hospital, Malmö, Sweden

[*]**Corresponding author**: Memorial Sloan-Kettering Cancer Center 1275 York Ave., Box 337 New York, NY 10065 Tel.: 646-888-2525 Fax: 646-422-0197 kleinr@mskcc.org.
[†]These authors contributed equally to this work.

Author contributions: Robert J. Klein had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Klein, Vickers, Lilja.
Acquisition of data: Hallden, Dahlin, Bjartell, Manjer, Ulmert, Lilja.
Analysis and interpretation of data: Klein, Savage, Gupta, Vickers, Lilja.
Drafting of the manuscript: Klein, Gupta, Vickers, Lilja.
Critical revision of the manuscript for important intellectual content: None.
Statistical analysis: Klein, Savage, Gupta, Vickers.
Obtaining funding: Scardino, Lilja.
Administrative, technical, or material support: None.
Supervision: Klein, Vickers, Lilja.
Other (specify): None.

Financial disclosures: I certify that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/ affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Hans Lilja holds patents for free PSA, intact PSA, and hK2 assays. Peter T Scardino holds stock in Claros. Robert J. Klein receives research funding from Sanofi-Aventis. The other authors have nothing to disclose.

[9]Departments of Laboratory Medicine and Medicine (GU-Oncology), Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Abstract

Background—Although case-control studies have identified numerous single nucleotide polymorphisms (SNPs) associated with prostate cancer, the clinical role of these SNPs remains unclear.

Objective—Evaluate previously identified SNPs for association with prostate cancer and accuracy in predicting prostate cancer in a large prospective population-based cohort of unscreened men.

Design, setting, and participants—This study used a nested case-control design based on the Malmö Diet and Cancer cohort with 943 men diagnosed with prostate cancer and 2829 matched controls. Blood samples were collected between 1991 and 1996, and follow-up lasted through 2005.

Measurements—We genotyped 50 SNPs, analyzed prostate-specific antigen (PSA) in blood from baseline, and tested for association with prostate cancer using the Cochran-Mantel-Haenszel test. We further developed a predictive model using SNPs nominally significant in univariate analysis and determined its accuracy to predict prostate cancer.

Results and limitations—Eighteen SNPs at 10 independent loci were associated with prostate cancer. Four independent SNPs at four independent loci remained significant after multiple test correction ($p < 0.001$). Seven SNPs at five independent loci were associated with advanced prostate cancer defined as clinical stage ≥T3 or evidence of metastasis at diagnosis. Four independent SNPs were associated with advanced or aggressive cancer defined as stage ≥T3, metastasis, Gleason score ≥8, or World Health Organization grade 3 at diagnosis. Prostate cancer risk prediction with SNPs alone was less accurate than with PSA at baseline (area under the curve of 0.57 vs 0.79), with no benefit from combining SNPs with PSA. This study is limited by our reliance on clinical diagnosis of prostate cancer; there are likely undiagnosed cases among our control group.

Conclusions—Only a few previously reported SNPs were associated with prostate cancer risk in the large prospective Diet and Cancer cohort in Malmö, Sweden. SNPs were less useful in predicting prostate cancer risk than PSA at baseline.

Keywords

Prostate cancer; Biomarkers; SNPs; PSA; Sensitivity and specificity

1. Introduction

Prostate cancer risk is higher among men with a family history of disease [1]. Recent genomewide association studies (GWAS) identified numerous single nucleotide polymorphisms (SNPs) associated with prostate cancer predisposition [2]. Additional independent SNPs at GWAS-identified loci were also associated with prostate cancer risk [3–6]. SNPs were also identified that influence levels of the prostate-secreted proteins prostate-specific antigen (PSA), hexokinase (hK) 2, and β-microseminoprotein (β-MSP); some of these SNPs were also associated with prostate cancer risk [7–9].

Given the maturity of the literature on prostate cancer risk SNPs, an obvious question concerns the clinical utility of these SNPs. For example, might knowledge of SNPs help a clinician decide whether a man should be advised to undergo regular screening or prescribed a chemopreventive? We considered five criteria for a study on the clinical usefulness of

prostate cancer SNPs. First, the study would focus on independent replication. Rather than attempting to discover new SNPs associated with prostate cancer risk, with attendant problems of false discovery, we would evaluate SNPs previously reported to be associated with prostate cancer risk. Second, we would study a large prospective unscreened population-based cohort rather than the conventional case-control cohort in which inclusion of men subject to PSA screening may identify SNPs associated with screen detection (eg, PSA levels) rather than prostate cancer risk.

We also considered that any clinically useful study must include an evaluation of predictive accuracy. Rather than asking whether there is a statistical association between a SNP and prostate cancer risk, we need to know how well SNPs predict prostate cancer. Given that a clinician can use PSA level to help predict long-term risk of prostate cancer [10], a key question concerns whether SNPs can add predictive value to PSA. Finally, we specified that any predictive model based on SNPs and PSA should be subject to full cross-validation of both selection of SNPs for the model as well as the weights given to each SNP in the model. In this study, we aimed to evaluate the clinical utility of SNPs for prostate cancer risk using data from a Swedish cohort followed for many years without PSA screening.

2. Patients and methods

2.1. Study cohort

The Malmö Diet and Cancer cohort was described previously [11]. The study was approved by the local ethics committee; all subjects gave written informed consent. Details of the cohort are in the Appendix.

We used a nested case-control design and matched each prostate cancer case with three controls selected at random from men with date of birth and date of baseline venipuncture within 3 mo of the case who were alive and without a prostate cancer diagnosis at the follow-up time at which the index case was diagnosed. Overall there were 943 cases and 2829 controls (Table 1). To determine whether any of the 50 SNPs correlate with unquestionably significant disease that is likely to lead to severe morbidity if left untreated, we used three definitions based on tumor grade and clinical stage at diagnosis (Table 5). For each definition, cases matching that definition were rematched to three controls with the caveat that only samples included in the initial case-control analysis could be matched.

2.2. Genotyping and biomarker measurements

We selected 50 SNPs for genotyping (Table 2). A total of 49 SNPs were selected on the basis of previous reports of association with prostate cancer or levels of prostate-produced proteins; the 50th SNP (rs2347867) is a candidate SNP involved in breast cancer [12]. Genotypes were determined using the Sequenom MassARRAY matrix-assisted laser desorption/ionization (MALDI-TOF) system. The levels of PSA in archived ethylenediaminetetraacetic acid (EDTA) anticoagulated blood plasma obtained at baseline were determined using the dual-label DELFIA ProStatus assay (PerkinElmer, Turku, Finland). Genotype calls and PSA measurements were independently made by an investigator blind to the clinical outcome of the participants.

2.3. Statistical analysis

All association analyses were performed using PLINK v1.06, as detailed in the Appendix. For each of the 37 SNPs previously reported to be associated with prostate cancer, we used the reported odds ratio along with the minor allele frequency we observed to compute power for a significance level of $\alpha = 0.05$ [13]. We performed 10 000 iterations of a simulation in

which each SNP was randomly assigned “significant” or “not significant” based on the power, and the total number of “significant” SNPs was counted.

To evaluate if SNPs added to PSA in predicting prostate cancer, we assessed predictive accuracy by the area under the curve (AUC) using the conventional 10-fold cross-validation method. We randomly split the data set into 10 subsamples, matching on prostate cancer outcomes so that each subsample included a similar number of events. We iteratively removed one of the 10 subsamples, using the remaining 9 as the training set. Using the training set, we identified SNPs that were nominally significant ($p < 0.05$) on univariate analysis and included these SNPs in a multivariable logistic regression model. The model predictions were then applied to the subsample that was not included in the model building. The entire process was repeated 10 times with each of the 10 subsamples used once as the validation set. All aspects of SNP selection and model building were repeated independently for each of the 10 training sets.

3. Results

We genotyped 943 prostate cancer cases and 2829 controls. Of these, we excluded 52 cases and 159 controls from the study due to missing genotype calls for at least 20% of the SNPs for that sample. An additional 149 controls were removed because they were matched with the 52 removed cases. This resulted in 891 cases and 2521 controls available for analysis. Of the 891 cases, 751 were still matched to three controls, 128 were matched with two controls, and 12 were matched with only one control. Except for one SNP (rs13252298) that had 10.9% of the data missing, all other SNPs had <6% of the data missing. None of these SNPs showed significant deviation from the Hardy-Weinberg equilibrium in controls ($p > 0.001$ for each SNP). Therefore, we retained all 50 SNPs in the analysis.

We first assessed whether any of the 50 SNPs were associated with prostate cancer. Using conditional logistic regression analysis, we found that 18 of the 50 SNPs were nominally associated with prostate cancer ($p < 0.05$; Table 2). Of these, six SNPs remained significant after accounting for multiple testing with the Bonferroni correction (nominal $p < 0.001$; Table 2). Because some of the significant SNPs were in linkage disequilibrium (LD), we determined whether they represented the same association signal or tagged independent loci. Of the 18 SNPs showing significant association with prostate cancer at the 5% level, 10 SNPs at four loci were in LD with another significant SNP ($r^2 > 0.20$; Table 3). For each set of SNPs in LD, after conditioning on the most significant SNP, the other SNPs no longer showed association with prostate cancer. This resulted in 12 SNPs at 10 separate loci independently associated with prostate cancer in this cohort (Table 4). Of these, 4 SNPs at 4 loci were significant after correcting for all 50 tests ($p < 0.001$; Table 4).

To determine if our nonsignificant results were expected for the 37 SNPs selected for genotyping based on previous reports of association with prostate cancer, we computed our power to detect association at $p < 0.05$ (Table 2). We observed nominal association at only one of five SNPs for which we had >99% power to detect association at previously reported odds ratios. Further, we had >50% power to detect 26 of 37 SNPs and >80% power to detect 16 of them. To determine whether we observed fewer significant SNPs than expected by chance, we simulated counts of significant SNPs based on the reported power. Of the 37 SNPs, 14 were significant with $p < 0.05$. This is significantly fewer than we would expect given the computed power ($p < 0.0001$). We also determined whether any SNP was associated with advanced or aggressive prostate cancer, using three separate definitions (Table 5). No SNPs were significant after correcting for multiple testing under any definition. Seven SNPs were nominally significant ($p < 0.05$) for association under definition 1, two SNPs were nominally significant under definition 2, and four SNPs were

nominally significant under definition 3 (Table 5 and Supplementary Table). For all these SNPs, the 95% confidence interval (CI) of the odds ratio overlapped with the 95% CI for overall risk.

Finally, we compared the discriminatory value of SNPs alone or PSA at baseline to a model that included PSA and nominally significant SNPs to predict prostate cancer diagnosis. SNPs alone predicted prostate cancer risk poorly with an AUC of only 0.571 (95% CI, 0.548–0.594) for the outcome of any diagnosis of prostate cancer, and SNPs alone were no better than chance at predicting risk of advanced or aggressive prostate cancer (Table 6). The addition of SNPs to PSA did not significantly enhance the ability to predict prostate cancer beyond that of PSA alone; indeed, a multivariable model including both PSA and SNPs had lower discrimination. For example, the AUC for advanced cancer (≥cT3 or metastasis) was reduced from 0.800 (95% CI, 0.771–0.830) for PSA to 0.788 (95% CI, 0.757–0.818) for PSA plus SNPs.

4. Discussion

We set out to determine whether previously reported prostate cancer risk SNPs might have a clinically relevant role in risk stratification for early detection or prevention. We were unable to replicate many SNPs despite adequate statistical power to detect many of these associations; we found evidence of nominally significant association at only 14 of 37 SNPs at 10 previously reported independent loci. We found that the predictive value of SNPs was poor, both when considered alone or as an adjunct to PSA at baseline. It is difficult to envision that markers that failed to improve prediction could have important clinical value alone.

Several prior papers support our findings. Failure to replicate associations between SNPs and prostate cancer is not unusual: Examination of several previously reported replication studies in diverse populations showed that only 25–60% of previously reported SNPs were associated with prostate cancer risk [14–17], whereas some associations were weak, particularly with respect to aggressive disease [18].

Our finding that SNPs had poor predictive value on their own (AUC: 0.57) is similar to several other reports with AUCs from 0.61 to 0.64 [19–22]. The predictive value of PSA may vary due to the composition of the study cohort [23] or selection of outcome such as long-term risk of prostate cancer [24], death from prostate cancer [10], or evidence of prostate cancer at biopsy [25]. Whereas Aly et al [22] suggested that SNPs improve predictive accuracy on top of PSA alone, Nam et al [14] and Johansson et al [26] found only small AUC increases (around 0.01). In all three studies the estimates of improvement may have been overly optimistic because these models were not fully cross-validated. These results suggest that in a PSA-screened population, adding SNPs to the predictive model will not improve the accuracy of the predictive test. Others have been unable to find an association between prostate cancer susceptibility variants and prostate cancer mortality [27,28].

Although our goal was to evaluate whether SNPs can predict who will develop prostate cancer, in reality we could only test if SNPs can predict who will be *diagnosed* with prostate cancer. Many older men have undiagnosed prostate cancer, either due to a lack of signs or symptoms indicating the need for a prostate biopsy or a false-negative biopsy due to the random nature of needle placements. Such undetected cases undoubtedly were present in our control population. There remains a dearth of evidence that GWAS-derived risk markers have clinical value for prostate cancer risk prediction [2]. In our evaluation of 50 SNPs previously reported in the literature, we found that most had no significant association with

aggressive prostate cancer, and adding SNPs to PSA did not improve predictive accuracy. As such, it is worth speculating whether there may be alternative clinical uses of SNPs. One possible novel approach is to determine whether SNPs can provide information to help interpret marker levels. Several SNPs associated with prostate cancer influence the function and/or production of a prostate cancer marker. For example, rs198977 in *KLK2* may influence hK2 function and reduce hK2 levels in blood [8]. The SNP rs10993994 in *MSMB* reduces levels of β-MSP and is associated with increased levels of PSA in the blood or semen of healthy young men [9]. A recent GWAS reported that rs10788160, located near *FGFR2*, is associated with increased PSA levels and is less common in men with a positive prostate biopsy compared with men with a benign biopsy [29]. The use of SNPs such as these would therefore not be to determine risk of prostate cancer at the population level but to help determine a biopsy decision by adjusting biomarker values based on SNP genotype. Considerable further research will be needed to investigate this possibility.

5. Conclusions

Although we could replicate many known prostate cancer risk SNPs, the numbers replicated were fewer than expected by chance. More important, SNPs do not add to the accuracy of predictive models for prostate cancer risk either alone or in addition to PSA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding/Support and role of the sponsor: This investigation was supported by the National Cancer Institute (grant numbers R33 CA 127768-02, P50-CA92629); Swedish Cancer Society (3455); Swedish Research Council (Medicine-20095; MTBH: 2006-7600); the Sidney Kimmel Center for Prostate and Urologic Cancers; David H. Koch through the Prostate Cancer Foundation; Fundación Federico SA; the AUA Foundation Research Scholars Program.

References

[1]. Gronberg H, Damber L, Damber JE. Familial prostate cancer in Sweden. A nationwide register cohort study. Cancer. 1996; 77:138–43. [PubMed: 8630920]

[2]. Stadler ZK, Thom P, Robson ME, et al. Genome-wide association studies of cancer. J Clin Oncol. 2010; 28:4255–67. [PubMed: 20585100]

[3]. 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; 99:1525–33. [PubMed: 17925536]

[4]. 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; 17:1203–13. [PubMed: 18483343]

[5]. Sun J, Zheng SL, Wiklund F, et al. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat Genet. 2008; 40:1153–5. [PubMed: 18758462]

[6]. Rafnar T, Sulem P, Stacey SN, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet. 2009; 41:221–7. [PubMed: 19151717]

[7]. Severi G, Hayes VM, Neufing P, et al. Variants in the prostate-specific antigen (PSA) gene and prostate cancer risk, survival, and circulating PSA. Cancer Epidemiol Biomarkers Prev. 2006; 15:1142–7. [PubMed: 16775173]

[8]. Klein RJ, Hallden C, Cronin AM, et al. Blood biomarker levels to aid discovery of cancer-related single-nucleotide polymorphisms: kallikreins and prostate cancer. Cancer Prev Res (Phila). 2010; 3:611–9. [PubMed: 20424135]

[9]. Xu X, Valtonen-Andre C, Savblom C, Hallden C, Lilja H, Klein RJ. Polymorphisms at the microseminoprotein-beta locus associated with physiologic variation in beta-microseminoprotein and prostate-specific antigen levels. Cancer Epidemiol Biomarkers Prev. 2010; 19:2035–42. [PubMed: 20696662]

[10]. Vickers AJ, Cronin AM, Bjork T, et al. Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. BMJ. 2010; 341:c4521. [PubMed: 20843935]

[11]. Manjer J, Carlsson S, Elmstahl S, et al. The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. Eur J Cancer Prev. 2001; 10:489–99. [PubMed: 11916347]

[12]. Sonestedt E, Ivarsson MI, Harlid S, et al. The protective association of high plasma enterolactone with breast cancer is reasonably robust in women with polymorphisms in the estrogen receptor alpha and beta genes. J Nutr. 2009; 139:993–1001. [PubMed: 19321582]

[13]. Klein RJ. Power analysis for genome-wide association studies. BMC Genet. 2007; 8:58. [PubMed: 17725844]

[14]. Nam RK, Zhang WW, Trachtenberg J, et al. Utility of incorporating genetic variants for the early detection of prostate cancer. Clin Cancer Res. 2009; 15:1787–93. [PubMed: 19223501]

[15]. 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; 70:270–5. [PubMed: 19902474]

[16]. Waters KM, Le Marchand L, Kolonel LN, et al. Generalizability of associations from prostate cancer genome-wide association studies in multiple populations. Cancer Epidemiol Biomarkers Prev. 2009; 18:1285–9. [PubMed: 19318432]

[17]. Yamada H, Penney KL, Takahashi H, et al. Replication of prostate cancer risk loci in a Japanese case-control association study. J Natl Cancer Inst. 2009; 101:1330–6. [PubMed: 19726753]

[18]. Fitzgerald LM, Kwon EM, Koopmeiners JS, Salinas CA, Stanford JL, Ostrander EA. Analysis of recently identified prostate cancer susceptibility loci in a population-based study: associations with family history and clinical features. Clin Cancer Res. 2009; 15:3231–7. [PubMed: 19366831]

[19]. Hsu FC, Sun J, Zhu Y, et al. Comparison of two methods for estimating absolute risk of prostate cancer based on single nucleotide polymorphisms and family history. Cancer Epidemiol Biomarkers Prev. 2010; 19:1083–8. [PubMed: 20332264]

[20]. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. N Engl J Med. 2008; 358:910–9. [PubMed: 18199855]

[21]. Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet. 2010; 42:570–5. [PubMed: 20562874]

[22]. Aly M, Wiklund F, Xu J, et al. Polygenic risk score improves prostate cancer risk prediction: results from the stockholm-1 cohort study. Eur Urol. 2011; 60:21–8. [PubMed: 21295399]

[23]. Vickers AJ, Cronin AM, Roobol MJ, et al. The relationship between prostate-specific antigen and prostate cancer risk: the Prostate Biopsy Collaborative Group. Clin Cancer Res. 2010; 16:4374–81. [PubMed: 20736330]

[24]. Lilja H, Cronin AM, Dahlin A, et al. Prediction of significant prostate cancer diagnosed 20 to 30 years later with a single measure of prostate-specific antigen at or before age 50. Cancer. 2011; 117:1210–9. [PubMed: 20960520]

[25]. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med. 2004; 350:2239–46. [PubMed: 15163773]

[26]. Johansson M, Holmstrom B, Hinchliffe SR, et al. Combining 33 genetic variants with prostate specific antigen for prediction of prostate cancer: Longitudinal study. Int J Cancer. In press. DOI: 10.1002/ijc.25986.

[27]. Wiklund FE, Adami HO, Zheng SL, et al. Established prostate cancer susceptibility variants are not associated with disease outcome. Cancer Epidemiol Biomarkers Prev. 2009; 18:1659–62. [PubMed: 19423541]

[28]. Penney KL, Pyne S, Schumacher FR, et al. Genome-wide association study of prostate cancer mortality. Cancer Epidemiol Biomarkers Prev. 2010; 19:2869–76. [PubMed: 20978177]

[29]. Gudmundsson J, Besenbacher S, Sulem P, et al. Genetic correction of PSA values using sequence variants associated with PSA levels. Sci Transl Med. 2010; 2:62–92.

[30]. Vickers AJ, Gupta A, Savage CJ, et al. A panel of kallikrein marker predicts prostate cancer in a large, population-based cohort followed for 15 years without screening. Cancer Epidemiol Biomarkers Prev. 2011; 20:255–61. [PubMed: 21148123]

Take-home message

Although numerous single nucleotide polymorphisms (SNPs) associated with prostate cancer risk have been identified and can be consistently replicated, these SNPs neither can predict who will develop prostate cancer alone nor do they enhance the predictive power of prostate-specific antigen testing.

Table 1

Characteristics of cancer cases*

PSA at diagnosis (*n* = 847)	10.7 (6.1–22)
Clinical T stage† (%)	
T1	170 (19.1)
T1a	19 (2.1)
T1b	14 (1.6)
T1c	160 (18)
T2	282 (31.7)
T2c	3 (0.3)
T3	200 (22.5)
T4	22 (2.5)
Missing	21 (2.4%)
WHO/Gleason Grade (%)	
WHO 1 or Gleason ≤6	149 (16.7)
WHO 2 or Gleason 7	469 (52.6)
WHO 3 or Gleason ≥8	149 (16.7)
Missing	124 (13.9)
Metastasis at presentation (%)	
No	487 (54.7)
Yes	65 (7.3)
Missing	339 (38.1)
Year of diagnosis (%)	
1998 or earlier	202 (22.7)
1999–2001	243 (27.3)
2002–2003	218 (24.5)
2004–2006	228 (25.6)

PSA = prostate-specific antigen; WHO = World Health Organization.

*All values are median (quartiles) or frequency (proportion).

†Based on Union Internationale Contre le Cancer 2001 staging.

Table 2

Association of 50 single nucleotide polymorphisms with prostate cancer diagnosis in the Malmö Diet and Cancer cohort[†]

SNP	Chr.	Position	Min/Maj allele	MAF	OR (95% CI)	p	Power
rs1465618	2	43407452	A/G	0.25	1.09 (0.96–1.23)	0.19	0.24
rs721048	2	62985234	A/G	0.18	1.10 (0.96–1.27)	0.16	0.54
rs12621278	2	173019798	G/A	0.06	1.03 (0.83–1.29)	0.77	0.65
rs2660753	3	87193363	T/C	0.09	1.20 (1.00–1.44)	0.051	0.47
rs10934853	3	129521062	A/C	0.26	1.11 (0.98–1.25)	0.11	0.48
rs12500426	4	95733631	A/C	0.46	1.03 (0.92–1.15)	0.63	0.30
rs17021918	4	95781899	T/C	0.34	0.94 (0.84–1.06)	0.33	0.46
rs7679673	4	106280982	A/C	0.40	0.98 (0.87–1.109)	0.68	0.41
rs2736098	5	1347085	A/G	0.26	1.06 (0.94–1.20)	0.37	0.54
rs401681	5	1375086	A/G	0.45	1.01 (0.90–1.12)	0.89	0.24
rs2347867	6	152271542	G/A	0.34	1.02 (0.91–1.15)	0.71	
rs9364554	6	160753653	T/C	0.32	1.01 (0.90–1.14)	0.87	0.80
rs10486567	7	27943087	T/C	0.22	0.79 (0.69–0.90)	**0.00065** *	0.45
rs6465657	7	97654262	C/T	0.48	1.08 (0.96–1.20)	0.19	0.57
rs2928679	8	23494919	A/G	0.43	1.00 (0.89–1.11)	0.96	0.15
rs12543663	8	127993840	C/A	0.30	1.05 (0.93–1.18)	0.40	0.27
rs10086908	8	128081118	C/T	0.31	0.94 (0.84–1.06)	0.33	0.99
rs1016343	8	128162478	T/C	0.23	1.16 (1.03–1.32)	**0.019**	0.99
rs13252298	8	128164337	G/A	0.27	0.89 (0.78–1.02)	0.093	0.81
rs16902094	8	128389527	G/A	0.13	0.99 (0.84–1.17)	0.91	0.73
rs445114	8	128392362	C/T	0.35	0.85 (0.75–0.95)	**0.0052**	0.64
rs620861	8	128404854	A/G	0.35	0.84 (0.75–0.94)	**0.0028**	0.79
rs6983267	8	128482486	T/G	0.48	0.90 (0.81–1.00)	0.052	0.99
rs4242382	8	128586754	A/G	0.12	1.06 (0.90–1.26)	0.46	0.99
rs10763588	10	51209767	G/T	0.42	1.16 (1.04–1.29)	**0.0083**	
rs7098889	10	51214480	C/T	0.42	1.15 (1.03–1.28)	**0.013**	
rs10993994	10	51219501	T/C	0.36	1.19 (1.07–1.33)	**0.0021**	0.78

SNP	Chr.	Position	Min/Maj allele	MAF	OR (95% CI)	p	Power
rs17178655	10	51231804	A/G	0.22	1.01 (0.88–1.15)	0.91	
rs7127900	11	2190149	A/G	0.21	1.16 (1.01–1.32)	**0.031**	0.88
rs11228565	11	68735155	A/G	0.20	1.35 (1.18–1.54)	**0.000011** *	0.90
rs7931342	11	68751072	G/T	0.50	1.20 (1.08–1.34)	**0.0012**	0.91
rs10896449	11	68751242	G/A	0.49	1.21 (1.08–1.35)	**0.00084** *	0.43
rs11649743	17	33149091	T/C	0.21	0.91 (0.80–1.04)	0.17	0.95
rs4430796	17	33172152	C/T	0.44	0.89 (0.80–1.00)	**0.040**	0.96
rs7501939	17	33175268	A/G	0.37	0.95 (0.84–1.06)	0.33	0.87
rs1859962	17	66620347	T/G	0.49	0.90 (0.81–1.00)	0.054	0.93
rs8102476	19	43427452	A/G	0.44	0.88 (0.79–0.98)	**0.024**	0.56
rs925013	19	56045412	G/A	0.22	1.09 (0.96–1.24)	0.20	
rs266882	19	56049824	G/A	0.48	1.10 (0.98–1.23)	0.095	
rs2271094	19	56051308	C/T	0.40	1.16 (1.03–1.29)	**0.011**	
rs61752561	19	56053194	A/G	0.045	1.10 (0.85–1.43)	0.46	
rs17632542	19	56053568	C/T	0.071	0.64 (0.51–0.81)	**0.00019** *	
rs2735839	19	56056434	A/G	0.14	0.75 (0.64–0.89)	**0.00067** *	0.64
rs198977	19	56073588	T/C	0.23	0.92 (0.81–1.05)	0.22	
rs198978	19	56074883	T/G	0.34	0.92 (0.82–1.04)	0.18	
rs80050017	19	56075012	T/G	0.078	0.93 (0.76–1.15)	0.51	
rs1654553	19	56102927	G/A	0.48	1.00 (0.90–1.12)	0.94	
rs7291691	22	38784014	G/T	0.20	1.00 (0.87–1.14)	1.0	0.99
rs5759167	22	41830155	T/G	0.49	0.80 (0.72–0.90)	**0.00010** *	0.81
rs5945619	X	51258411	G/A	0.38	1.30 (1.11–1.52)	**0.0015**	0.89

Chr. = chromosome; Ref. = reference from which the SNP was chosen; Min/Maj allele = minor and major alleles; MAF = minor allele frequency; OR (95% CI) = odds ratio and 95% confidence interval; *p* = *p* value for association test. Power is the probability of detecting an association with $p < 0.05$ assuming the minor allele frequency we observe in the controls along with the OR from the literature.

† Nominally associated SNPs ($p < 0.05$) are shown in bold.

* An asterisk indicates statistical significance after correcting for 50 SNPs ($p < 0.001$).

Table 3

Test for independence of associated single nucleotide polymorphisms in linkage disequilibrium

Chr.	Main SNP	Secondary SNP	r^2	p
8	rs620861	rs445114	0.97	ND
10	rs10993994	rs10763588	0.71	0.94
		rs7098889	0.72	0.79
11	rs11228565	rs10896449	0.26	0.31
		rs7931342	0.26	0.33
19	rs17632542	rs2735839	0.48	0.23

Chr. = chromosome; SNP = single nucleotide polymorphism; ND = not determined.

Table 4

Set of single nucleotide polymorphisms independently associated with prostate cancer risk in the Malmö Diet and Cancer study

SNP	Chromosomal band	Nearby gene(s) of interest	p
rs10486567	7p15.2	*JAZF1*	0.00065*
rs1016343	8q24.21	*MYC*	0.019
rs620861	8q24.21	*MYC*	0.028
rs10993994	10q11.23	*MSMB; NCOA4*	0.0021
rs7127900	11p15.5	*IGF2; INS; TH*	0.031
rs11228565	11q13.2	*TPCN2; MYEOV*	0.000011*
rs4430796	17q12	*HNF1B*	0.040
rs8102476	19q13.2	*DPF1; PPP1R14A; SPINT2*	0.024
rs2271094	19q13.33	*KLK3*	0.011
rs17632542	19q13.33	*KLK3*	0.00019*
rs5759167	22q13.2	*TTLL1;BIK*	0.00010*
rs5945619	Xp11.22	*NUDT11*	0.0015

SNP = single nucleotide polymorphism.

* *p* value reaches statistical significance after correcting for 50 SNPs ($p < 0.001$). Genes of interest are based on genes in a 500-kb window in the University of California, Santa Cruz, genome browser. When evidence suggests a particular gene is involved in the locus, we omitted other genes in the window.

Table 5

Significant associations with aggressive prostate cancer risk in the Malmö Diet and Cancer Study

SNP	Chr.	Nearby genes of interest	Clinical stage ≥T3 or evidence of metastasis at time of diagnosis (Definition 1; *n* = 239)		WHO 3 or Gleason score ≥8 at time of diagnosis (Definition 2; *n* = 167)		Clinical stage ≥T3, evidence of metastasis, WHO 3, or Gleason score ≥8 at time of diagnosis (Definition 3; *n* = 319)	
			OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
rs10486567	7	*JAZF1*	0.80 (0.61–1.04)	0.091	0.66 (0.47–0.92)	0.013	0.75 (0.59–0.96)	0.021
rs1016343	8	*MYC*	1.30 (1.02–1.66)	0.035	1.34 (1.01–1.79)	0.043	1.28 (1.03–1.58)	0.024
rs10763588	10	*MSMB; NCOA4*	1.28 (1.04–1.59)	0.020	1.25 (0.97–1.62)	0.081	1.18 (0.98–1.42)	0.077
rs7098889	10	*MSMB; NCOA4*	1.29 (1.04–1.59)	0.018	1.22 (0.95–1.57)	0.12	1.17 (0.97–1.41)	0.092
rs10993994	10	*MSMB; NCOA4*	1.30 (1.05–1.62)	0.015	1.27 (0.98–1.64)	0.072	1.19 (0.99–1.44)	0.065
rs11228565	11	*TPCN2; MYEOV*	1.46 (1.14–1.88)	0.0029	1.29 (0.95–1.75)	0.096	1.38 (1.10–1.74)	0.0049
rs11649743	17	*HNF1B*	0.76 (0.59–1.00)	0.044	0.73 (0.52–1.01)	0.061	0.76 (0.60–0.96)	0.022
rs17632542	19	*KLK3*	0.57 (0.35–0.92)	0.021	1.07 (0.65–1.75)	0.8	0.74 (0.50–1.10)	0.13

SNP = single nucleotide polymorphism; Chr. = chromosome; WHO = World Health Organization; OR (95% CI): odds ratio and 95% confidence interval; *p* = *p* value for association test.

Table 6

Areas under the receiving operating curve (area under the curve) for models built using prostate-specific antigen (PSA) alone, PSA, and single nucleotide polymorphisms (SNPs) or SNPs alone in predicting risk of a prostate cancer diagnosis[†]

	Any prostate cancer	Aggressive or advanced prostate cancer (clinical stage ≥T3, evidence of metastasis, WHO grade 3, or Gleason stage ≥8 at diagnosis)	Advanced prostate cancer (clinical stage ≥T3 or evidence of metastasis at diagnosis)
PSA alone	0.792 (0.774–0.810)	0.823 (0.792–0.855)	0.800 (0.771–0.830)
PSA plus SNPs	0.791 (0.773–0.809)	0.811 (0.777–0.844)	0.788 (0.757–0.818)
SNPs alone	0.571 (0.548–0.594)	0.498 (0.455–0.541)	0.499 (0.460–0.538)

WHO = World Health Organization; SMP = single nucleotide polymorphism.

[†]All estimates have been corrected for overfit using 10-fold repeated cross-validation and are reported as area under the curve (95% confidence interval).