

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail. Please cite the original version.

Risk Prediction of Prostate Cancer with Single Nucleotide

Polymorphisms (SNPs) and Prostate-Specific Antigen (PSA)

Sam Li-Sheng Chen¹, Jean Ching-Yuan Fann², Csilla Sipeky³, Teng-Kai Yang^{4,5}, Sherry Yueh-Hsia Chiu⁶, Amy Ming-Fang Yen¹, Virpi Laitinen⁷, Teuvo LJ Tammela⁸, Ulf-Håkan Stenman^{9,10}, Anssi Auvinen¹¹, Johanna Schleutker^{3,12}, Hsiu-Hsi Chen⁵

¹ School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

² Department of Health Industry Management, School of Healthcare Management, Kainan University, Tao-Yuan, Taiwan

³ Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Finland

⁴ Department of Urology, National Taiwan University Hospital

⁵ Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

⁶ Department and Graduate Institute of Health Care Management, College of Management, Chang Gung University, Tao-Yuan, Taiwan

⁷ Institute of Biomedical Technology/BioMediTech, University of Tampere, Tampere, Finland

⁸ Department of Urology, Tampere University Hospital, and Faculty of Medicine and Biosciences, University of Tampere, Tampere, Finland

⁹ Department of Clinical Chemistry, Helsinki University Central Hospital, Helsinki, Finland

¹⁰ Faculty of Medicine, University of Helsinki, Helsinki, Finland

¹¹ Faculty of Social Sciences/Section of Health Sciences, University of Tampere, Tampere, Finland

¹² Department of Medical Genetics, Division of Laboratory, Turku University Hospital, Turku, Finland

Running head: Prediction for Prostate Cancer by PSA and Genetic Polymorphisms

Correspondence & reprint requests to: Hsiu-Hsi Chen, Ph.D., Division of Biostatistics, Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 521, 5F, No. 17 Suchow Road, Taipei 100, Taiwan
Tel: +886-2-33228033, Fax: +886-2-23587707
E-mail: chenlin@ntu.edu.tw

Abstract

Purpose: Combined information on single nucleotide polymorphisms (SNPs) and prostate-specific antigen (PSA) offers opportunities for improving the performance of screening by risk stratification. We aim to predict the risk of prostate cancer (PrCa) based on PSA together with SNPs information.

Materials and Methods: Prospective study of 20,575 men with PSA test from the Finnish population-based screening trial for PrCa during 1996-2007 and 5,269 samples on seven SNPs from the Finnish PrCa family study during 1994-2013. Bayesian predictive model was built for estimating the risk of PrCa by sequentially combining genetic information with PSA. Receiver operating characteristic curve was used to evaluate the optimal cutoffs with PSA alone and the combined PSA and genetic variants with the conventional cutoff of PSA from 4 ng/mL onward.

Results: The posterior odds for PrCa based on the seven SNPs together with the PSA level ranged from 3.7 at PSA 4 ng/mL, 14.2 at 6 ng/mL, 40.7 at PSA 8 ng/mL, to 98.2 at PSA 10 ng/mL. The areas under curves increased from 70.1% (95% CI: 69.6%-70.7%) using PSA alone to 95.8% (95% CI: 95.3%-96.4%) for PSA in combination with seven SNPs for subjects limited to PSA \geq 4ng/mL.

Conclusions: Expedient use of multiple genetic variants together with information on PSA levels better predicts the risk of PrCa than PSA alone and allows higher PSA cutoffs. Combined information also provides a basis for risk stratification that can be used for optimizing the performance of PrCa screening.

Introduction

Several international collaborative genome-wide association studies have been conducted to identify genetic factors in association with hereditary predisposition to prostate cancer (PrCa). A constellation of >120 single-nucleotide polymorphisms (SNPs) have been revealed with several located in five chromosomal regions-three at 8q24 and one each at 17q12 and 17q24.3.¹⁻⁵ Although the effect of each of the SNPs on the risk for prostate cancer is small to moderate (in general ORs of 1.2-1.5), a strong cumulative association has been demonstrated by using several SNPs in combination.⁶ Multiple prostate cancer-specific multigene panels have been evaluated for detection of prostate cancer.⁷ Use of the major SNPs offers an opportunity to identify sub-groups of men with PrCa risk substantially below and above the population average.

In parallel with these genome-wide studies, the effectiveness of population-based screening for prostate cancer with prostate specific-antigen (PSA) has been intensively researched. However, the effectiveness of screening in reducing mortality is still debatable due to conflicting results of the two major randomized trials and the balance between benefits and harms remains uncertain (add refs ERSPC and PLCO). To enhance the efficiency and reduce the harm, i.e. overdiagnosis caused by screening, combining genetic information together with PSA given age and genetic variant holds promise for more accurate identification of high-risk men with potential for large screening benefits.

The purpose of this study was to develop a Bayesian algorithm to predict the risk of prostate cancer based on PSA data from the Finnish population-based screening trial for prostate cancer together with the SNPs identified from the Finnish PrCa family study.

Methods

Study Subjects

To estimate the risk of PrCa based on PSA and selected SNPs, we combined two Finnish datasets, one from the population-based randomized screening trial (20,575 men enrolled) and an unselected patient series from Tampere University hospital. The details of study design and preliminary results for the former have been published previously^{8,9,10} and the mortality results have been published also as a part of the ERSPC trial.¹¹ The dataset included DNA samples genotyped for SNPs rs4242382, rs10486567, rs16901979, rs6983267, rs138213197, rs1447295, and rs1859962 collected from 2,959 individuals who participated in the Finnish screening trial (518 prostate cancers and 2,441 prostate cancer-free subjects) plus 2310 prostate cancer patients from the Tampere University Hospital. The rationale for using the two datasets is their complimentary nature, as the genetic dataset included wide-scale genetic information, but with incomplete PSA data, while the situation was reverse for the screening trial. Figure 1 gives a summary of the estimates of interest, the use of model and distribution, and data sources. Age-specific incidence rates of PrCa were obtained from the Finnish Cancer Registry. Information on age-dependent PSA level was obtained from the screening trial.

Study design on the ascertainment of PrCa with various cutoffs

In clinical practice, the PSA threshold for referral to biopsy is often 4 ng/mL, but other cut-offs are also used, including age-specific values. It is therefore reasonable to obtain sensitivity and specificity for various cutoffs and to assess the performance of PSA screening with receiver operating characteristic (ROC) curve.

Genetic polymorphisms

To incorporate information on SNPs in association with PrCa, we assessed the combined effects of seven SNPs, rs4242382, rs6983267, rs1601979, and rs1447295 at 8q24, rs104865677 at 7p15.2, rs138213197 and rs1859962 at 17q21. The risk allele A of rs424238 at 8q24 has been previously reported to be associated with PrCa and aggressive PrCa. The risk allele G of rs10486567 at 7p, the intron 2 of the JAZF zinc finger1 gene (*JAZF1*) is commonly seen in Europeans.¹² The association between rs138213197 in HOXB13 and the risk of hereditary prostate cancer has also been addressed,¹³ and the effect has been shown to be especially strong in the Finnish population.¹⁴

Statistical Analysis

We classified PSA into 13 categories with an increment of 0.5 ng/mL. To fit the normal distribution, the PSA concentrations were transformed into logarithms. The distribution of PSA in men with and without PrCa is given in the Appendix table. To incorporate information on SNPs, we assessed the joint effects of seven SNPs. The optimal cutoff of PSA based on ROC curve could be calculated by the largest value of the formula, Sensitivity + Specificity – 1, from each PSA cut-off.

As the genetic variants associated with PCa are heterogeneous, it is necessary to make a comparison across different ethnic groups or populations by using the information on the proportion of each SNP in population and the effect of each SNP to PrCa risk. We used results from the previous Zheng's study⁶ for external validation of developed model.

The details of the algorithm developed with Bayesian underpinning are given in the Appendix. Note that the risk of PrCa in terms of prior, likelihood, and posterior is denoted by the odds (probability/(1-probability)). Data analysis was performed with SAS 9.4 and Winbugs software.

Results

Estimates of the risk for PrCa (Posterior Odds) by different levels of PSA

Table 1 shows the likelihood ratios for log(PSA) and the SNPs, as well as the posterior odds by PSA levels given the prior odds (1: 2.78) for the risk of PrCa for men aged 60 years or younger at baseline. Our model was used to discern the PrCa cases from 4 ng/mL upward given the posterior odds by combining PSA and seven SNPs, increasing from 3.7 (95% CI:1.6-10) at 4 ng/mL of PSA to 98.2 (95% CI:27.3-437.5) at 10 ng/mL of PSA. The likelihood ratio based on the presence of the risk alleles of the seven SNPs was 2.8 considering the weighted distribution (the proportion of each SNP in population) contributed from each SNP (see the footnote of Table 1).

Table 1 also shows the posterior odds for PrCa among men 60 years or younger by updating the prior with the likelihood ratios based on genetic polymorphisms together with the seven SNPs. Table 1 shows similar findings for the men aged 63-71 at baseline, with substantially higher risk levels, but equally large posterior odds related to the genetic risk determinants.

The posterior probability of prostate cancer by age and PSA level taking seven SNPs into account from Finnish study was simulated and the results are shown in Figure 2.

ROC curves limited to men with PSA \geq 4 ng/mL

Figure 3 shows ROC curves limited only to men with PSA \geq 4ng/mL, using information on PSA alone and the combined information on both PSA and the selected SNPs. Compared with Figure 3, it can be clearly seen that adding SNP information to this risk group substantially enhanced the performance of risk prediction for PrCA as the area under curve (AUC) increased from 70.1% based on PSA only to 95.8% with PSA combined with seven SNPs.

External Validation

The proposed predictive model was further extended to incorporate five SNPs from the Zheng's study.⁶ Considering the five SNPs, the odds of PrCa was 2.4 at 4 ng/mL compared with 2 at 0.6 ng/mL (Table 2). The optimal cutoff was 9.9 ng/mL when using PSA plus the five SNPs. The corresponding AUC was 86.8% (95% CI: 86.6%-87.0%).

The external validation based on four common SNPs (rs1859962, rs16901979, rs6983267, and rs1447295) from the Finnish and Zheng's studies was also conducted. The predicted ROC curve was built by applying the regression coefficients of the four SNPs obtained from the Zheng's study to the empirical Finnish PSA data. The comparison between the externally predicted ROC and the observed ROC of the Finnish PSA data is shown in Figure 4. We found that AUC 81.7% (95% CI: 81.5%-82.0%) for PSA in combination with the four SNPs in the Zheng's study was slightly lower than the 85.3% (95% CI: 85.1%-85.5%) for PSA combined with four SNPs in Finnish data ($P < 0.0001$). The statistical significant difference suggests the results are not compatible even if the difference in the ROC values is not large.

Discussion

The proposed clinical prediction algorithm with Bayesian underpinning provides a feasible approach for PrCa risk stratification by combining information on PSA multiple genetic variants identified from genome-wide studies. The merits of our proposed method are several. First, the large contrast in PrCa risk between high and low-risk groups when using both sources of information has the potential to aid physicians and patients in early detection of PrCa by virtue of better discrimination of the risk. It also provides opportunities for individually tailored screening such as commencing screening from younger age and considering shorter inter-screening interval for the men with higher posterior PrCa risk.¹⁰ Second, the combined use of information on PSA and the SNPs may also reduce false negative cases missed at PSA screen (such as interval cancer), as some men with low PSA levels may nevertheless have an increased risk of prostate cancer if they carry one or more high-risk alleles. The posterior odds was 4-fold higher than the prior based on PSA alone at 4 ng/ml if all seven risk SNPs were present. Third, the proposed method may also reduce false positive results. The optimal cut-off was raised from 9.1 to 10.7 when information on the seven SNPs is added, which is likely to reduce the frequency of screen-positive findings (Among men with ≥ 4 ng/mL, 17.5 % of men had PSA > 9.1 ng/mL in our screening data). It should be noted that the optimal cut-off was selected by using the maximum efficacy of both sensitivity and specificity. The utility of sensitivity and specificity was equally treated. Fourth, enhanced risk stratification in screening may improve cost-effectiveness, because it not only reduces the false negative results but also false positive findings. Although the incorporation of

genetic information may involve substantial costs, improved performance in early detection can outweigh the cost incurred by the genetic testing. The unit cost of such genetic testing at population level may also be reduced due to a large scale. However, this requires a formal cost-effectiveness analysis for the evaluation of the net balance between costs from genetic testing and benefits from early detection.

From the viewpoint of translational research, our proposed approach has the potential to enhance the efficiency of mass screening for PrCa. The results obtained from our approach may enable modification of screening algorithms for PrCa screening modalities, including PSA cut-off level, inter-screening interval, and age of starting screen by risk profile. The higher the risk predicted by the proposed model, the more costly screening method, the shorter inter-screening interval, and the earlier age of commencing screening should be considered. Again, the balance of benefits and harms from such risk-based screening approach needs to be confirmed by further research using cost-effectiveness analysis.

A key methodological concern is the assumption that the SNPs are independent of PSA level, which remains imperfectly verified. It could be debated whether such an assumption is reasonable. A previous study demonstrated that the five prostate cancer associated SNPs were independent of PSA levels⁶. There is no significant association between SNPs (rs4242382, rs10486567, rs16901979, rs6983267, rs1447295, rs1859962, and rs138213197) and PSA concentration in patient samples^{15,16}. Accordingly, the joint effect of PSA and these seven SNPs can be easily decomposed into the product of their independent effects. Although this assumption is supported by Zheng et al., it should be

empirically verified before applying our PrCa risk stratification algorithm for screening. Moreover, the seven-SNP model slightly out-performed the five-SNP model.

Another concern is the variation in genetic risk prediction across populations, i.e. population stratification. The genetic determinants of PrCa risk from different populations are not highly consistent, suggesting that the genetic factors underlying hereditary susceptibility may vary between populations. The validation was not well fitted in our analysis of external validation. This could mean that different SNPs will need to be incorporated in different populations (but a similar underlying methodology could still be used). It is still unclear to what extent the proposed model can be applied in populations other than where it has been developed (possible overfitting). In our proposed model, the area under ROC was as high as 95.8%, with substantial improvement from PSA alone (AUC: 70.1%). It can be expected that the predictive validity would be further enhanced when more SNPs are added. The contribution of additional SNPs depends on their frequency, effect size and independence of the already incorporated SNPs.

Note that predicting the risk for PrCa with PSA often only focuses on those limited to $PSA \geq 4$ ng/mL¹⁷⁻¹⁹ and few studies cover those with PSA below 4 ng/mL as in our study. The AUC was 61% for total PSA in a study of 1051 men with PSA between 4 to 10 ng/mL who were diagnosed with benign prostatic hyperplasia after biopsy.¹⁷ The AUC for total PSA was 71% in a similar study targeted at 150 patients with benign prostatic hyperplasia.¹⁸ Punglia et al. reported 62-69% of AUC for young (age < 60 years) and old men (age \geq 60 years), respectively, based on biopsy-positive men and men with interval cancers during 18-month follow-up as subjects missed at earlier assessment.¹⁹ As

biopsy was more likely for those who had high PSA level compared with low PSA level, AUC of PSA tests was raised to 72%-86% after the correction of this verification bias. The AUC for PSA testing was also high in two population-based case-control studies.²⁰
²¹ These previous findings were compatible to our results based on the limited ranges of PSA higher than 4 ng/mL. The AUCs of PSA test were 70.1% in our Bayesian analysis and 68% in traditional logistic regression analysis (Appendix Figure 1), respectively. The higher AUC in ROC analysis might be arguable. However, the cross-validation by randomly selecting two-third training data and one-third validation data from combined two Finnish datasets was therefore performed. In Appendix Figure 2, the 68.4% of AUC for training data was close to 67.1% for validation data (P=0.15) in those PSA level ≥ 4 ng/mL suggesting reliability of the prediction model.

Adding SNPs information to subjects with PSA > 4 ng/mL increased predictive ability of AUC substantially (by 25 percentage points) in our analysis. It should be also noted that although augmenting PrCa risk prediction by genetic data improved AUC by enhancing sensitivity, it could result in more false positive cases. However, the elevation of sensitivity was larger than reduction in false positives in the subjects with PSA > 4 ng/mL. The AUC values are calculated with similar weighting for sensitivity and specificity, implicitly assuming that they are of equal importance. To assess this would require information on consequences of missed cases (false negatives) and unnecessary biopsy referrals (false positives).

In conclusion, the expedient use of multiple genetic variants in seven chromosomal regions associated with PrCa risk together with information on PSA through a Bayesian reasoning algorithm improves risk stratification, i.e. classification of men into different

risk levels, which could provide the basis for risk-adapted PrCa screening to maximize its benefits and minimize the harms.

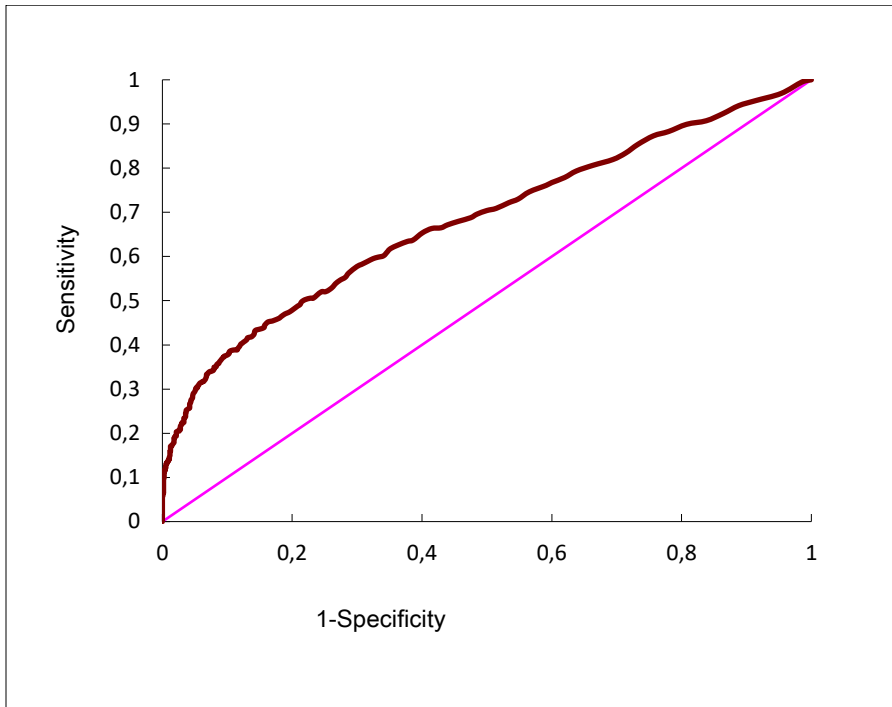
Appendix

Appendix Table: The distribution of PSA with log transformation among men with and without prostate cancer in the Finnish prostate cancer screening trial

Age	Free of Prostate Cancer		Prostate Cancer		p-Value*
	N	log(PSA), Mean(SD)	N	log(PSA), Mean(SD)	
Age 55-59	399	1.742(0.325)	227	2.156 (0.705)	<0.0001
Age 63-71	719	1.801(0.346)	388	2.251(0.801)	<0.0001
Overall	1118	1.780(0.340)	615	2.216(0.768)	<0.0001

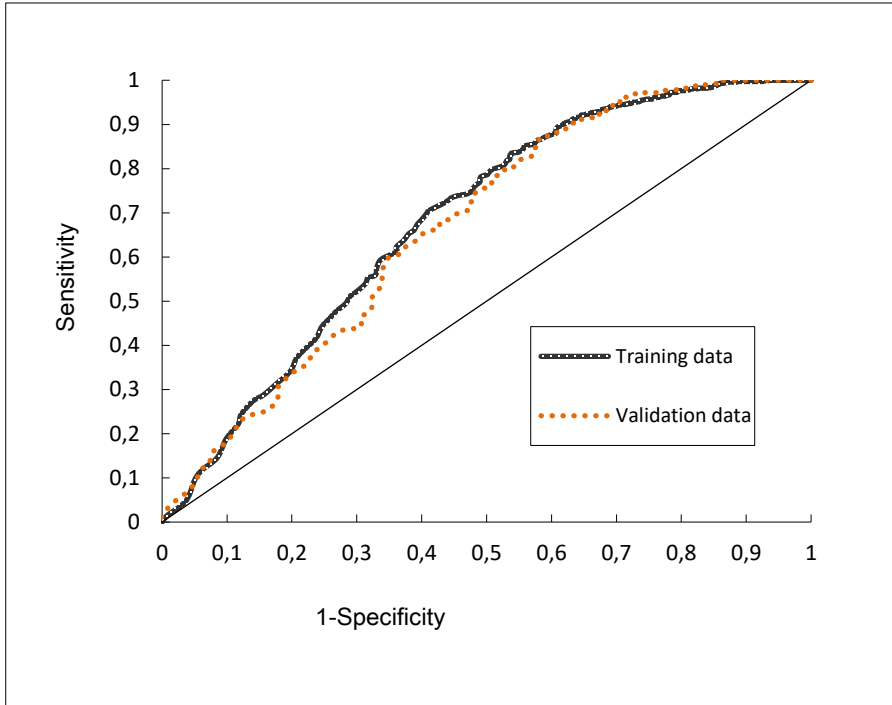
* Adjusting for age

Appendix Figure 1: Receive Operating Characteristic Curves for Prostate Cancer using traditional logistic regression analysis.



PSA ≥ 4 , AUC : 0.68 (0.65-0.71)

Appendix Figure 2: Receive Operating Characteristic Curves for Prostate Cancer using training and cross-validation data.



AUC (Training data): 68.4% (95% CI: 65.1%-71.7%)
AUC (Validation data): 67.1% (95% CI: 62.0%-72.1%)
P=0.1486

Prediction of the risk of PrCa with SNPs and PSA levels by using Bayesian clinical reasoning

We adopted a Bayesian clinical reasoning to estimate PSA- and SNP-based posterior odds for PrCa by updating the baseline risk of PrCa (prior) with the likelihood ratios between PrCa positive and negative men formed by the two corresponding distributions of PSA and the other likelihood ratio based on SNPs contribution, which is equivalent to the ratio of sensitivity to false positive, yielding the ROC curve. The posterior odds of developing prostate cancer given a specific PSA level and the SNPs of interests by the Bayesian algorithm considering different scenarios are derived as follows:

(1) With seven SNPs

$$\begin{aligned} & \frac{P(D|SNP_1^+, SNP_2^+, \dots, SNP_7^+, PSA_D)}{P(\bar{D}|SNP_1^+, SNP_2^+, \dots, SNP_7^+, PSA_{\bar{D}})} \\ &= \frac{P(D)}{P(\bar{D})} \times \frac{P(PSA_D | D, SNP_1^+, SNP_2^+, \dots, SNP_7^+)}{P(PSA_{\bar{D}} | \bar{D}, SNP_1^+, SNP_2^+, \dots, SNP_7^+)} \times \frac{P(SNP_1^+, SNP_2^+, \dots, SNP_7^+ | D)}{P(SNP_1^+, SNP_2^+, \dots, SNP_7^+ | \bar{D})} \end{aligned}$$

, where D represents the event of prostate cancer, and \bar{D} is the complement of D (non-disease). P(D) is prior probability of prostate cancer and P(\bar{D}) is prior probability of being free of prostate cancer.

Assume PSA level is the conditionally independent of SNP once the disease status is determined. The formula can be simplified as $\frac{P(PSA_D|D)}{P(PSA_{\bar{D}}|\bar{D})}$

Let PSA_D and $PSA_{\bar{D}}$ denote PSA in men with and without prostate cancer. Both follow the two normal distributions, indicated by $N(u_D, \sigma_D^2)$ and $N(u_{\bar{D}}, \sigma_{\bar{D}}^2)$; the likelihood ratio then becomes

$$\frac{P(PSA_D | D, SNP_1^+, SNP_2^+, \dots, SNP_7^+)}{P(PSA_{\bar{D}} | \bar{D}, SNP_1^+, SNP_2^+, \dots, SNP_7^+)} = \sqrt{\frac{\sigma_{\bar{D}}}{\sigma_D}} \times \exp \left\{ -\frac{1}{2} \left[\left(\frac{PSA_D - u_D}{\sigma_D} \right)^2 - \left(\frac{PSA_{\bar{D}} - u_{\bar{D}}}{\sigma_{\bar{D}}} \right)^2 \right] \right\}$$

u_D : the average estimate of PSA for prostate cancer cases

σ_D : standard deviation of PSA for prostate cancer cases

$u_{\bar{D}}$: average PSA for prostate cancer free men

$\sigma_{\bar{D}}$: standard deviation of PSA for prostate cancer free men

Reference

1. 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
2. Haiman CA, Patterson N, Freedman ML, et al: Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007;39:638-644
3. 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
4. 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
5. 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.
6. Zheng SL, Sun J, Wiklund F, et al: Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008; 358:910-919.
7. Little J, Wilson B, Carter R, et al. Multigene panels in prostate cancer risk assessment: a systematic review. *Genet Med.* 2016;18:535-544.
8. Kilpeläinen T, Tammela T, Malila N, et al: Prostate cancer mortality in the Finnish randomized screening trial. *J Natl Cancer Inst* 2013;105:719-25.
9. Finne P, Stenman UH., Määttänen L et al: The Finnish trial of prostate cancer screening: where are we now? *BJU Int* 2003; 92:22-26.
10. Wu GH, Auvinen A, Yen AMF, et al: A Stochastic Model for Survival of Early Prostate Cancer with Adjustments for Leadtime, Length Bias, and Over-detection. *Biom J* 2012; 54:20-44.
11. Schröder FH, Hugosson J, Roobol MJ, et al: Prostate-cancer mortality at 11 years of follow-up. *N Engl J Med* 2012; 366:981-990.
12. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genomewide association study of prostate cancer. *Nat Genet* 2008;40: 310-315.
13. Ewing CM, Ray AM, Lange EM, et al. Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 2012;12;366:141-149.

14. Laitinen VH, Wahlfors T, Saaristo L, et al. HOXB13 G84E mutation in Finland: population-based analysis of prostate, breast, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 2013; 22:452-460.
15. Bao BY, Pao JB, Lin VC, et al. Individual and cumulative association of prostate cancer susceptibility variants with clinicopathologic characteristics of the disease. *Clin Chim Acta*. 2010 6;411(17-18):1232-1237
16. Kote-Jarai Z, Mikropoulos C, Leongamornlert DA, et al. Prevalence of the HOXB13 G84E germline mutation in British men and correlation with prostate cancer risk, tumour characteristics and clinical outcomes. *Ann Oncol*. 2015;26:756-61
17. Djavan B, Zlotta A, Remzi M, et al: Optimal predictors of prostate cancer on repeat prostate biopsy: a prospective study of 1,051 men. *J Urol* 2000; 163:1144-1148.
18. Hara I, Miyake H, Hara S, et al: Significance of prostate-specific antigen--alpha(1)-antichymotrypsin complex for diagnosis and staging of prostate cancer. *Jpn J Clin Oncol* 2001; 31:506-509.
19. Punglia RS, D'Amico AV, Catalona WJ, et al: Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. *N Engl J Med* 2003; 349:335-342.
20. Jacobsen SJ, Bergstralh EJ, Guess HA, et al. Predictive properties of serum-prostate-specific antigen testing in a community-based setting. *Arch Intern Med* 1996; 156:2462-2468.
21. Morgan TO, Jacobsen SJ, McCarthy WF, et al. Age-specific reference ranges for prostate-specific antigen in black men. *N Engl J Med* 1996; 335:304-310.

FIGURE LEGENDS

Figure 1. Summary of the estimates of interest, the use of model and distribution, and data sources.

Figure 2. Posterior odds of prostate cancer by age with or without considering seven SNPs Finnish Study.

Figure 3. Receiver operating characteristic curves for prostate cancer based on PSA alone, PSA and genetic data (seven SNPs).

Figure 4. Receiver operating characteristic curves for external validation based on four common SNPs(rs1859962, rs16901979, rs6983267, and rs1447295).

Table 1. Posterior odds of prostate cancer by PSA level based on seven SNPs, the Finnish prostate cancer screening trial

PSA Level	Men younger 60 years				Men aged 63-71 years			
	$\frac{P(\text{PSA} D)}{P(\text{PSA} \bar{D})}$ Likelihood Ratio given PSA (A)	$\frac{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_7^+ D)}{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_7^+ \bar{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 7 SNPs (C)		$\frac{P(\text{PSA} D)}{P(\text{PSA} \bar{D})}$ Likelihood Ratio given PSA (A)	$\frac{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_7^+ D)}{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_7^+ \bar{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 7 SNPs (C)	
	Estimate	Estimate	Estimate	95%CI	Estimate	Estimate	Estimate	95%CI
4.0	3.8	2.8	3.7	(1.6-10)	1.39	2.8	1.3	(0.6-3)
4.5	5.5	2.8	5.4	(2.2-15.3)	1.86	2.8	1.8	(0.8-4.2)
5.0	7.7	2.8	7.6	(3-23.3)	2.42	2.8	2.3	(1.1-5.6)
5.5	10.7	2.8	10.5	(3.9-33.6)	3.10	2.8	2.9	(1.3-7.5)
6.0	14.5	2.8	14.2	(5.2-47.5)	3.89	2.8	3.7	(1.6-10)
6.5	19.3	2.8	18.8	(6.5-65.7)	4.82	2.8	4.6	(2-12.7)
7.0	25.3	2.8	24.7	(8.4-89.1)	5.90	2.8	5.6	(2.4-16)
7.5	32.7	2.8	31.9	(10.3-122.6)	7.16	2.8	6.8	(2.8-20.1)
8.0	41.8	2.8	40.7	(12.9-157.4)	8.60	2.8	8.2	(3.2-24.7)
8.5	52.9	2.8	51.5	(15.6-207)	10.25	2.8	9.7	(3.8-30.2)
9.0	66.1	2.8	64.9	(18.9-267)	12.14	2.8	11.5	(4.4-36.7)
9.5	81.9	2.8	79.8	(22.6-348.4)	14.27	2.8	13.6	(5-44.5)
10.0	100.8	2.8	98.2	(27.28-437.5)	16.64	2.8	15.7	(5.7-54)

Considering seven SNPs (rs4242382 & rs10486567 & rs16901979 & rs6983267 & rs138213197 & rs1447295 & rs1859962) from the Finnish DNA study
 The likelihood ratios: 1.88 (95% CI:1.42-2.49) for rs4242382, 1.68 (95% CI:1.35-2.09) for rs10486567, 1.45 (95% CI:1.18-1.77) for rs1601979, 1.54 (95%
 CI:1.36-1.74) for rs6983267, 8.98 (95% CI:5.51-14.65) for rs138213197, 1.93 (95% CI:1.46-2.56) for rs1447295, and 1.42 (95% CI:1.25-1.61) for rs1859962,

$$\frac{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_7^+ | D)}{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_7^+ | \bar{D})} \cdot \exp(0.0438 \cdot \log(1.88) + 0.933 \cdot \log(1.68) + 0.0818 \cdot \log(1.45) + 0.285 \cdot \log(1.54) + 0.0367 \cdot \log(8.98) + 0.0441 \cdot \log(1.93) + 0.7558 \cdot \log(1.42)) = 2.8$$

$$(C) = \frac{P(D)}{P(\bar{D})} \times (A) \times (B)$$

Table 2. Posterior odds of prostate cancer by PSA level based on five SNPs data from Zheng's study

PSA Level	Men younger 60 years				Men aged 63-71 years			
	$\frac{P(\text{PSA} D)}{P(\text{PSA} \bar{D})}$ Likelihood Ratio given PSA (A)	$\frac{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_5^+ D)}{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_5^+ \bar{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 5 SNPs (C)		$\frac{P(\text{PSA} D)}{P(\text{PSA} \bar{D})}$ Likelihood Ratio given PSA (A)	$\frac{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_5^+ D)}{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_5^+ \bar{D})}$ SNP-specific risk (B)	Posterior Odds by combing PSA and 5 SNPs (C)	
	Estimate	Estimate	Estimate	95%CI	Estimate	Estimate	Estimate	95%CI
4.0	3.8	1.7	2.3	(1-6.1)	1.39	1.7	0.8	(0.4-1.8)
4.5	5.5	1.7	3.4	(1.4-9.5)	1.86	1.7	1.1	(0.6-2.6)
5.0	7.7	1.7	4.8	(1.9-14.4)	2.42	1.7	1.5	(0.7-3.5)
5.5	10.7	1.7	6.7	(2.6-21)	3.10	1.7	1.9	(0.9-4.6)
6.0	14.5	1.7	9.0	(3.4-29)	3.89	1.7	2.3	(1.1-6.1)
6.5	19.3	1.7	11.9	(4.3-41.4)	4.82	1.7	2.9	(1.3-7.8)
7.0	25.3	1.7	15.7	(5.4-55.8)	5.90	1.7	3.5	(1.5-9.9)
7.5	32.7	1.7	20.3	(6.7-75.5)	7.16	1.7	4.3	(1.8-12.5)
8.0	41.8	1.7	26.0	(8.3-100.2)	8.60	1.7	5.2	(2.1-15.4)
8.5	52.9	1.7	32.7	(10.1-129.9)	10.25	1.7	6.2	(2.5-18.7)
9.0	66.1	1.7	40.8	(12.3-168.2)	12.14	1.7	7.3	(2.8-23.1)
9.5	81.9	1.7	50.7	(14.7-217.1)	14.27	1.7	8.6	(3.3-27.5)
10.0	100.8	1.7	62.4	(17.74-271)	16.64	1.7	10.0	(3.7-33.1)

Considering five SNPs (rs4430796, rs1859962, rs16901979, rs6983267, and rs1447295) from Zheng's study

$$\frac{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_5^+|D)}{P(\text{SNP}_1^+, \text{SNP}_2^+, \dots, \text{SNP}_5^+|\bar{D})} \cdot \exp(0.56 \cdot \log(1.38) + 0.5 \cdot \log(1.28) + 0.03 \cdot \log(1.53) + 0.51 \cdot \log(1.37) + 0.14 \cdot \log(1.22)) = 1.7$$

$$(C) = \frac{P(D)}{P(\bar{D})} \times (A) \times (B)$$