

## Utility of Incorporating Genetic Variants for the Early Detection of Prostate Cancer

Robert K. Nam,<sup>1,9</sup> William W. Zhang,<sup>1</sup> John Trachtenberg,<sup>6</sup> Arun Seth,<sup>2</sup> Laurence H. Klotz,<sup>1</sup> Aleksandra Stanimirovic,<sup>1</sup> Sanoj Punnen,<sup>1</sup> Vasundara Venkateswaran,<sup>1</sup> Ants Toi,<sup>7</sup> D. Andrew Loblaw,<sup>3</sup> Linda Sugar,<sup>4</sup> Katherine A. Siminovitch,<sup>8</sup> and Steven A. Narod<sup>5</sup>

**Abstract Purpose:** Several single nucleotide polymorphisms (SNP) have been associated with the risk of prostate cancer. The clinical utility of using SNPs in the early detection of prostate cancer has not been evaluated.

**Experimental Design:** We examined a panel of 25 SNPs from candidate genes and chromosomal regions in 3,004 unselected men who were screened for prostate cancer using serum prostate-specific antigen (PSA) and digital rectal examination. All underwent a prostate biopsy. We evaluated the ability of these SNPs to help predict the presence of prostate cancer at biopsy.

**Results:** Of the 3,004 patients, 1,389 (46.2%) were found to have prostate cancer. Fifteen of the 25 SNPs studied were significantly associated with prostate cancer ( $P = 0.02-7 \times 10^{-8}$ ). We selected a combination of 4 SNPs with the best predictive value for further study. After adjusting for other predictive factors, the odds ratio for patients with all four of the variant genotypes compared with men with no variant genotype was 5.1 (95% confidence interval, 1.6-16.5;  $P = 0.006$ ). When incorporated into a nomogram, genotype status contributed more significantly than PSA, family history, ethnicity, urinary symptoms, and digital rectal examination (area under the curve = 0.74). The positive predictive value of the PSA test ranged from 42% to 94% depending on the number of variant genotypes carried ( $P = 1 \times 10^{-15}$ ).

**Conclusions:** SNP genotyping can be used in a clinical setting for the early detection of prostate cancer in a nomogram approach and by improving the positive predictive value of the PSA test.

Recently, several genome-wide association studies have identified associations between alleles of specific single nucleotide polymorphisms (SNP) and the risk of prostate cancer (1-4). Three independent studies identified important variants at chromosome 8q24 (2-4). Zheng et al. studied 2,893 prostate cancer cases and 1,781 controls from Sweden (1). The effect of each individual SNP was small [odds ratios (OR) between 1.1 and 1.5], but men who had five genetic variants had an OR of 4.5 for having prostate cancer compared with men with no genetic variant. We and others have used a candidate gene

approach to find additional SNPs that are significantly associated with prostate cancer (5-7).

These studies raise the possibility that multigenic SNP models may be clinically relevant and may help to categorize men into various levels of cancer risk. This assignment could be beneficial in choosing men for surveillance and potentially for chemoprevention (8). It is also possible that, among men who undergo prostate-specific antigen (PSA) screening, SNP genotyping could provide important predictive information beyond that of the PSA level itself. However, the clinical significance of incorporating SNP genotyping in a clinical setting is largely unknown. Furthermore, the study by Zheng et al. included only Swedish men and it is not clear to what extent their results are applicable to other ethnic groups (1).

To determine whether SNPs found previously to be associated with prostate cancer are useful in a clinical setting in North America, we genotyped 3,004 men who underwent a prostate biopsy for a panel of 25 SNPs.

### Materials and Methods

**Study subjects.** Patients were drawn from a sample of 3,261 eligible men who were referred to the prostate centers of the University of Toronto (Sunnybrook & Women's College Health Sciences Centre and University Health Network) between June 1999 and June 2007 (9). No patient had a past history of prostate cancer. Patients were included in the study if they had an abnormal PSA value ( $\geq 4.0$  ng/mL) or an abnormal digital rectal examination (DRE). All patients underwent one or more transrectal ultrasonography-guided needle core biopsies.

**Authors' Affiliations:** Divisions of <sup>1</sup>Urology and <sup>2</sup>Molecular and Cellular Biology, Sunnybrook Research Institute; Departments of <sup>3</sup>Radiation Oncology and <sup>4</sup>Pathology, Sunnybrook Health Sciences Centre; Departments of <sup>5</sup>Public Health Sciences and <sup>6</sup>Division of Urology and Departments of <sup>7</sup>Medical Imaging and <sup>8</sup>Medicine, Immunology, and Molecular Genetics, University Health Network; and <sup>9</sup>Department of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

Received 6/27/08; revised 10/20/08; accepted 11/14/08; published OnlineFirst 02/17/2009.

**Grant support:** National Cancer Institute grant 010294 and Canadian Institute of Health Research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Requests for reprints:** Robert K. Nam, 2075 Bayview Avenue, MG-406, Toronto, Ontario, Canada M4N 3M5. Phone: 416-480-5075; Fax: 416-480-6121; E-mail: Robert.nam@utoronto.ca.

© 2009 American Association for Cancer Research.  
doi:10.1158/1078-0432.CCR-08-1593

### Translational Relevance

New single nucleotide polymorphism (SNP) variants have been found to be associated with prostate cancer risk. This study examines how various combinations of SNP variants can be used in a clinical setting and develops a clinical tool (nomogram) to help integrate the information into practice. A panel of 19 SNPs was genotyped in 3,004 patients who underwent a prostate biopsy. Their predictive value, alone and in combination, was compared with other methods of diagnosing prostate cancer. We found that a combination of four SNP variants could be used in a clinical setting to predict prostate cancer and incorporated these into a nomogram-based prostate cancer risk calculator. This is the first study that translates the findings of many SNP discovery studies into a nomogram-based prostate cancer risk calculator in diagnosing prostate cancer in the context of prostate cancer screening program.

Patients were excluded if their PSA was >50 ng/mL (where the decision to biopsy would be considered unequivocal;  $n = 54$ ), if they were not capable of giving consent to participate in a research study ( $n = 46$ ), or if they could not provide sufficient baseline information ( $n = 53$ ). From 6 to 15 ultrasound-guided needle core biopsies were done (median, 8) using an 18-gauge spring loaded biopsy device. Samples were obtained using a systematic pattern and additional targeted samples were obtained from suspicious areas. The primary endpoint was the histologic presence of adenocarcinoma of the prostate in the biopsy specimen. All grading was based on the Gleason scoring system (10). Of the 3,108 men, 3,004 (97%) had sufficient leukocyte DNA available for SNP analysis.

**Baseline data information and primary endpoint.** A urologic voiding history (American Urological Association symptom score; ref. 11), DRE results, serum PSA level, family history of prostate cancer information, and ethnic background were obtained by research personnel through questionnaire administration and medical record review. All data were stored within a centralized database. Stored serum (at  $-70^{\circ}\text{C}$ ) was obtained for free:total PSA ratio measurements for each patient using standardized commercial kits (Beckman-Coulter).

**Selection of SNPs.** We examined a panel of 25 SNPs; 15 of the SNPs were reported by Zheng et al. from chromosomal regions 8q24 and 17q (1). We also examined 10 other SNPs that have been associated with prostate cancer; these included SNPs from the KLK2, TNF, HOGG, 9p22, and ETV1-rs2348763 and ETV1-rs13225697 genes and from the locus of HPC1 on chromosome 1q24 (5–7). We included 2 SNPs from within the ERG transcription genes [we and others have shown that TMPRSS2:ERG gene fusion products are associated with higher rates of prostate cancer progression (12)].

Genotyping was conducted using mass spectrometry-based genotyping analysis and matrix-assisted laser desorption ionization-time of flight (MassArray System; Sequenom) following the manufacturer's instructions. Details of the SNPs regarding location, allele type, location, and primers have been detailed elsewhere (1, 5–7, 12). A standard protocol for multiplex homogeneous mass extend assay developed by Sequenom was used and modified according to designed primers. For quality control, we assigned negative controls for each test plate (Microseal TM 384 version 2.0).

Successful genotyping assays were defined as those for which  $\geq 90\%$  of all possible genotyping calls were obtained. We analyzed a total of 25 SNPs (15 from Zheng et al. and 10 from our panel) that had call rates of  $>90\%$  (average, 95%) from mass spectrometry SNP analysis.

To validate these SNPs and include them in further analyses, we tested for Hardy-Weinberg equilibrium among the controls. Departures

from Hardy-Weinberg equilibrium could be due to ethnic heterogeneity or to errors in genotyping. Six of the 25 SNPs (rs983085, rs6983561, rs7214479, rs6501455, and rs4242382 and ETV1) were not in Hardy-Weinberg equilibrium ( $P < 0.01$ ) and were excluded from further analyses. Thus, a total of 19 SNPs was studied.

**Data analysis.** Cases were defined as patients with prostate cancer and controls were men with no evidence of cancer. Allele frequencies for each SNP were calculated for cases and controls and the distributions were compared. Genotype groupings were tested based on additive, dominant, and recessive genetic models for each SNP and the one with the highest likelihood was chosen as the best model. For the SNPs examined by Zheng et al., we used their genotype groupings (1).

We examined three models of SNP combinations: the first was the panel of five SNPs reported by Zheng et al., the second model was based on selecting other SNPs from those that we and others have shown to be significantly associated with prostate cancer, and the third model was based on selecting the best SNPs from models 1 and 2. This was based on the highest ORs and smallest  $P$  values from the first and second models. We tested the cumulative effects of selected SNPs for each model by counting the number of genotypes associated with prostate cancer; the ORs for prostate cancer for patients with one or more variant genotypes were estimated using univariate and multivariate analyses. In the multivariate analysis, we adjusted for age, ethnic group, family history of prostate cancer, presence of lower urinary tract voiding symptoms, total PSA level, free:total PSA ratio, and DRE. Unconditional logistic regression analysis was used to examine how each of these factors, alone and in combination, would predict the presence of prostate cancer.

**Receiver operating characteristic analysis and nomogram construction.** Receiver operating characteristics were constructed to estimate the area under the curve (AUC) of the various SNP models. The baseline model

**Table 1.** Comparison of baseline factors of patients undergoing prostate biopsy between cases and controls

Factor	Cancer	No cancer	P
<b>Total (n = 3,004)</b>	<b>n = 1,389 (46.2%), n (%)</b>	<b>n = 1,615 (53.8%), n (%)</b>	
Age group (y)			$1.8 \times 10^{-20}$
$\leq 50$	29 (2.1)	83 (5.1)	
51-60	334 (24.0)	491 (30.4)	
61-70	585 (42.1)	719 (44.5)	
$> 70$	441 (31.8)	322 (20.0)	
Family history of prostate cancer			0.0006
Absent	1,161 (83.6)	1,420 (87.9.0)	
Present	228 (16.4)	195 (12.1)	
Ethnicity			$2 \times 10^{-14}$
Asian	46 (3.3)	122 (7.6)	
Caucasian	1,160 (83.5)	1,314 (81.4)	
Black	155 (11.2)	117 (7.2)	
Other	28 (2.0)	62 (3.8)	
Lower urinary tract symptoms			$1.5 \times 10^{-5}$
$\leq 7$	797 (57.4)	812 (50.3)	
$> 7$	592 (42.6)	803 (49.7)	
DRE			$3 \times 10^{-16}$
No nodule	797 (57.4)	1,156 (71.6)	
Nodule	592 (42.6)	459 (28.4)	
PSA (ng/mL)			$1.4 \times 10^{-24}$
$\leq 4.0$	101 (7.3)	293 (18.1)	
4.1-10.0	846 (60.9)	930 (57.6)	
10.1-20.0	343 (24.7)	326 (20.2)	
$> 20.0$	99 (7.1)	66 (4.1)	
Free:total PSA ratio			$1.9 \times 10^{-43}$
$\leq 0.15$	906 (65.2)	649 (40.2)	
$> 0.15$	483 (34.8)	966 (59.8)	

**Table 2.** Panel of SNPs analyzed by Zheng et al. and comparisons of allele frequencies in cases and controls

SNP information	Position	Alternative alleles	Associated allele	Frequency		Genotype		OR* (95% CI)	P
				Cases	Controls	Reference	Associated		
rs4430796 17q12	33,172,153	A, G	G	0.48	0.48	AA or AG	GG	1.04 (0.9-1.2)	0.70
rs7501939 17q12	33,175,269	C, T	T	0.38	0.38	CC or CT	TT	1.04 (0.8-1.3)	0.75
rs3760511 17q12	33,180,426	A, C	C	0.33	0.34	AA or AC	CC	1.02 (0.8-1.3)	0.85
rs1859962 17q24.3	66,620,348	G, T	G	0.51	0.46	GG or TG	GG	1.34 (1.1-1.6)	0.001
rs16901979 8q24	128,194,098	C, A	A	0.10	0.09	CC	CA or AA	1.07 (0.9-1.3)	0.50
rs6983267 8q24	128,482,487	G, T	G	0.59	0.55	GG	GT or GG	1.20 (1.0-1.4)	0.02
rs7000448 8q24	128,510,352	C, T	T	0.44	0.41	CC	CT or TT	1.16 (1.0-1.4)	0.08
rs1447295 8q24	128,554,220	C, A	A	0.15	0.10	CC	CA or AA	1.61 (1.3-1.9)	$7.3 \times 10^{-8}$
rs7017300 8q24	128,594,450	A, C	C	0.18	0.13	AA	AC or CC	1.50 (1.3-1.8)	$6.6 \times 10^{-7}$
rs7837688 8q24	128,608,542	G, T	T	0.12	0.08	GG	GT or TT	1.51 (1.2-1.8)	$8.4 \times 10^{-6}$

\*Allelic ORs are based on the multiplicative model.

included age, ethnic group, family history of prostate cancer, urinary voiding symptoms, PSA, free:total PSA ratio, and DRE. The AUC was then estimated after adding the information derived from SNPs from each of the three models. The revised model was then compared with that of the baseline model.

To develop a clinical instrument that incorporates the SNP findings, we added the four SNP genotype results to our previously established nomogram. This nomogram was designed to predict both prostate cancer and aggressive cancer defined as having intermediate to high-grade prostate cancer (Gleason score  $\geq 7$ ; ref. 9). Ordinal logistic regression was used to model the probability of having low- or high-grade cancer. Three outcome levels were defined: (a) no cancer, (b) low-grade cancer (Gleason score  $\leq 6$ ), and (c) intermediate to high-grade cancer (Gleason score  $\geq 7$ ). Continuous variables were modeled with restricted cubic splines to avoid linearity assumptions. The logistic regression model was the basis for constructing a nomogram. Two thirds of patients were used for model building, and the remaining third was used for model assessment. The nomogram was validated using two strategies. First, discrimination was quantified with the area under the receiver operating characteristic curve. Second, calibration was assessed. This was done by grouping patients into deciles (each of size 100) with respect to their nomogram-predicted probabilities and then comparing the mean of the group with the observed proportion of patients with any or high-grade cancer. All analyses were done using S-Plus 2000 Professional software (Statistical Sciences) with the Design and Hmisc libraries added (13).

## Results

Of the 3,004 men who underwent one or more prostate biopsies, 1,389 (46.2%) were found to have adenocarcinoma of the prostate. The mean age at the time of prostate biopsy was 64.5 years (range, 40-94 years). The characteristics of the cases and the cancer-free controls are presented in Table 1. Age, ethnicity, family history of prostate cancer, presence of urinary symptoms, PSA level, and free:total PSA ratio were each found to be significantly associated with the presence of prostate cancer (Table 1).

We evaluated the distribution of the genotypes of 19 SNPs in the cases and controls. Of the 10 SNPs reported by Zheng et al., 6 were significantly associated with the presence of prostate cancer at biopsy (Table 2). The strongest associations were found for 4 SNPs from the region of chromosome 8q24 (rs1447295, rs7017300, rs16901979, and rs7837688; Table 2). Of the 9 SNPs in the second panel (those not in the Zheng et al. study), all were significantly associated with prostate cancer (Table 3). We then constructed three models, which varied in choice of SNPs. In each model, all SNPs were from different chromosome loci; no 2 SNPs were in linkage disequilibrium. Although we selected SNPs with a call rate of  $>90\%$ , all SNPs

**Table 3.** List of second panel of SNPs selected for analysis based on previous work and comparison of allele frequencies between cases and controls

SNP information	Position*	Alternative alleles	Associated allele	Frequency		Genotype		OR (95% CI)	P
				Cases	Controls	Reference	Associated		
ERG rs2836431	25,047,642	C, T	C	0.94	0.92	TC or TT	CC	1.36 (1.1-1.7)	0.006
ERG rs8131855	25,010,807	A, G	G	0.10	0.08	AA	AG or GG	1.34 (1.1-1.6)	0.003
HOGG1-326 rs1052133	9,733,475	C, G	C	0.79	0.75	GG	CG or CC	1.67 (1.2-2.3)	0.01
KLK2 rs198972	48,430,967	C, T	T	0.36	0.33	CC	CT or TT	1.16 (1.0-1.3)	0.05
KLK2 rs2664155	48,428,104	A, G	A	0.34	0.31	GG	AG or AA	1.24 (1.1-1.4)	0.001
TNF rs1800629	31,651,010	A, G	A	0.14	0.11	GG	AG or AA	1.27 (1.1-1.5)	0.001
rs1552895 9p22	14,773,799	C, G	C	0.41	0.37	GG	CG or CC	1.21 (1.0-1.4)	0.001
HPC1, 1q25 rs1930293	157,405,115	A, G	G	0.20	0.16	AA	AG or GG	1.27 (1.1-1.5)	0.003
ETV1, 7p21 rs2348763	7,864,240	A, C	A	0.70	0.67	CC or CA	AA	1.25 (1.1-1.4)	0.001

\*Chromosomal position based on National Cancer for Biotechnology Information database Build 36.2.

**Table 4.** Univariate and multivariate analysis of combinations of SNPs

Model	No. associated genotypes	Frequency (%)		Crude OR (95% CI)	P	P <sub>trend</sub>	Adjusted OR* (95% CI)	P	P <sub>trend</sub>
		Cases	Controls						
Model 1: five-SNP model (Zheng et al.): rs4430796, rs1859962, rs16901979, rs6983267, and rs1447295	0	24.0	32.0	1.00			1.00		
	1	39.5	37.7	1.40 (1.2-1.7)	0.0007		1.40 (1.1-1.7)	0.001	
	2	24.4	21.6	1.50 (1.2-1.9)	0.0002		1.47 (1.2-1.9)	0.002	
	3	9.0	7.0	1.73 (1.2-2.4)	0.0005		1.58 (1.1-2.2)	0.008	
	≥4	3.1	1.7	2.42 (1.4-4.1)	0.001	2 × 10 <sup>-5</sup>	1.55 (0.9-2.8)	0.14	0.0008
Model 2: four-SNP model from second panel of SNPs: KLK2, HPC1, TNF, and ETV1	0	5.5	8.4	1.00			1.00		
	1	26.6	31.5	1.29 (0.9-1.8)	0.12		1.32 (0.9-1.9)	0.12	
	2	39.0	38.6	1.54 (1.1-2.1)	0.007		1.44 (1.0-2.0)	0.04	
	3	23.6	18.3	1.97 (1.4-2.7)	<0.0001		1.69 (1.2-2.4)	0.004	
	4	5.3	3.2	2.53 (1.6-4.1)	0.0001	<0.0001	2.17 (1.3-3.6)	0.003	<0.0001
Model 3: combination of four-SNP model: rs1447295, rs1859962, TNF, and ETV1	0	20.2	27.8	1.00			1.00		
	1	41.0	43.4	1.30 (1.1-1.6)	0.01		1.23 (1.0-1.5)	0.05	
	2	28.8	24.0	1.65 (1.3-2.1)	1 × 10 <sup>-8</sup>		1.45 (1.1-1.8)	0.002	
	3	8.7	4.5	2.66 (1.9-3.8)	1 × 10 <sup>-10</sup>		2.22 (1.5-3.2)	1 × 10 <sup>-8</sup>	
	4	1.3	0.3	6.07 (2.0-18.5)	0.001	1 × 10 <sup>-15</sup>	5.09 (1.6-16.5)	0.006	<0.0001

\*Adjusted for age, family history of prostate cancer, ethnicity, urinary symptoms, PSA, free:total PSA ratio, and DRE within multivariate logistic regression model.

that were significantly associated with prostate cancer had call rates of >95%.

**Model 1.** We first examined the model defined by Zheng et al. (1), in which five SNPs were selected. Three of the five SNPs (rs1859962, rs6983267, and rs1447295) chosen by Zheng et al. were also associated with prostate cancer detection in our cohort. The crude OR for having prostate cancer for men with four or more of the five variant genotypes was 2.4 [95% confidence interval (95% CI), 1.4-4.1; *P* = 0.001; Table 4]. However, after adjusting for age, family history of prostate cancer, ethnicity, presence of urinary voiding symptoms, PSA level, free:total PSA ratio, and DRE, the adjusted OR for patients with four or more variant genotypes compared with patients with no variant genotype was 1.55 (95% CI, 0.9-2.8; *P* = 0.14; Table 4). The ORs for prostate cancer for those with one to

three variant genotypes did not change significantly after adjustment.

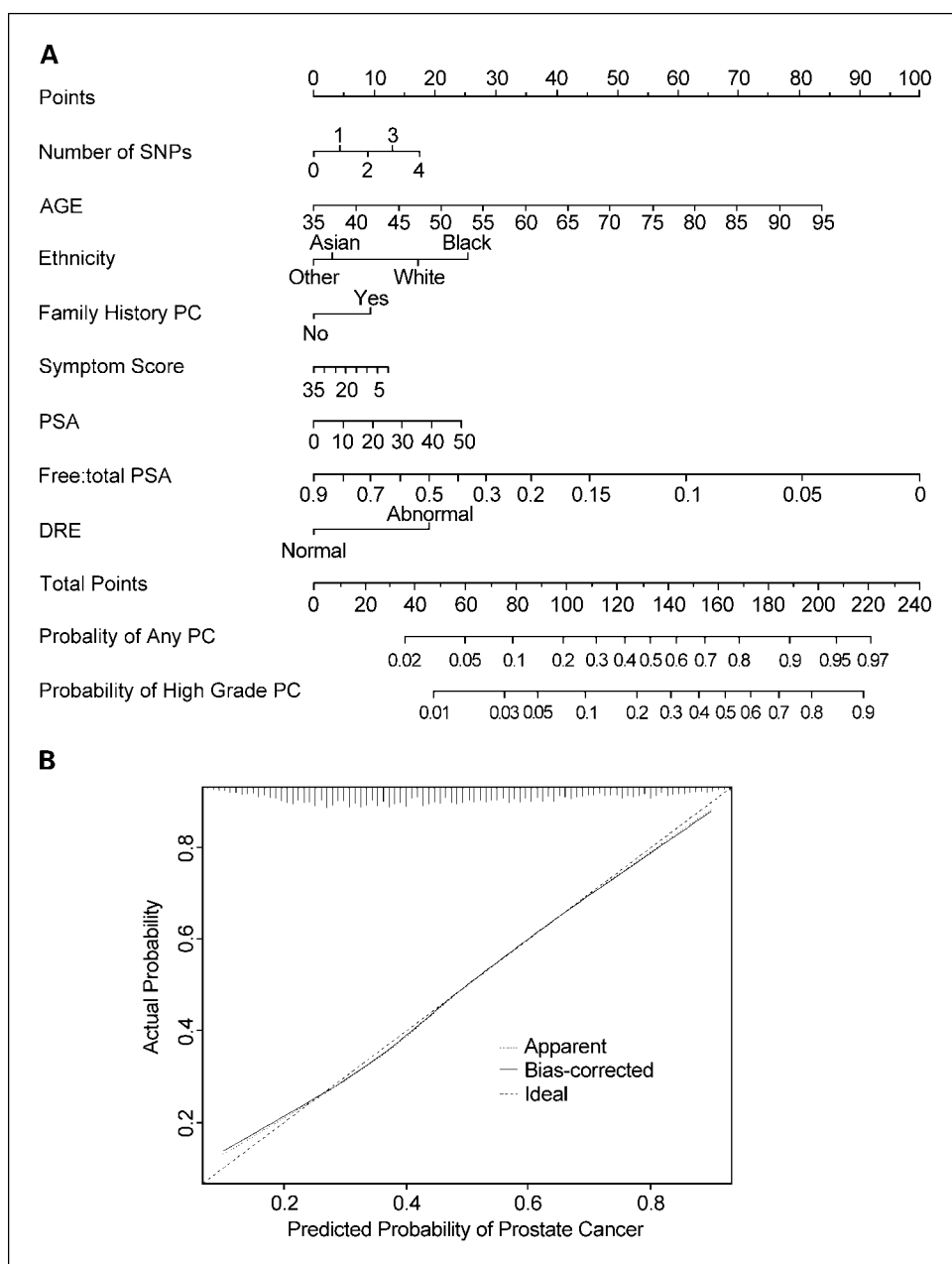
**Model 2.** Similar to the approach by Zheng et al., we chose four SNPs with the most significant *P* values selected from a panel based on our previous work. We used a combination of four SNP variants including, KLK2, HPC1, TNF, and ETV1-rs2348763. The crude OR for prostate cancer for patients with four variant genotypes compared with patients with no variant genotype was 2.5 (95% CI, 1.6-4.1; *P* = 0.0001; Table 4).

**Model 3.** We incorporated two SNPs from the model of Zheng et al. (rs1447295-8q24 and rs1859962-17q24) and two SNPs from our second panel of SNPs (TNF and ETV1-rs2348763). The crude OR for prostate cancer for patients with four variant genotypes compared with patients with no variant genotype was 6.1 (95% CI, 2.0-19; *P* = 0.001; Table 4). After

**Table 5.** OR for prostate cancer restricted to subjects of Caucasian descent only for each SNP model

Model	No. associated genotypes	Frequency (%)		Crude OR (95% CI)	P	P <sub>trend</sub>
		Cases	Controls			
Model 1: five-SNP model (Zheng et al.): rs4430796, rs1859962, rs16901979, rs6983267, and rs14472951	0	27.3	35.3	1.00		
	1	42.0	38.7	1.41 (1.2-1.7)	0.0009	
	2	24.5	20.8	1.53 (1.2-1.9)	0.0004	
	3	5.0	4.9	1.33 (0.9-2.0)	0.17	
	≥4	1.2	0.3	4.46 (1.4-13.9)	0.0001	1 × 10 <sup>-5</sup>
Model 2: four-SNP model from second panel of SNPs: KLK2, HPC1, TNF, and ETV1	0	5.8	8.4	1.00		
	1	27.2	32.2	1.22 (0.9-1.7)	0.26	
	2	38.8	37.5	1.49 (1.1-2.1)	0.02	
	3	23.1	18.9	1.76 (1.2-2.5)	0.002	
	4	5.1	3.1	2.38 (1.4-4.0)	0.0009	0.0001
Model 3: combination of four-SNP model: rs1447295, rs1859962, TNF, and ETV1	0	22.9	30.5	1.00		
	1	41.7	43.9	1.26 (1.0-1.6)	0.03	
	2	26.3	21.7	1.61 (1.3-2.1)	0.0001	
	3	8.1	3.5	3.05 (2.0-4.6)	1 × 10 <sup>-8</sup>	
	4	1.0	0.4	3.81 (1.2-12.3)	0.02	1 × 10 <sup>-11</sup>

**Fig. 1.** *A*, nomogram prediction model for predicting prostate cancer at the time of biopsy. The nomogram is used by first locating a patient's position for each variable on its horizontal scale and then a point value is assigned according to the points scale (*top axis*) and summed for all variables. Total points correspond to a probability value for having prostate cancer or aggressive prostate cancer. Number of SNPs refers to the number of variant genotypes for the SNP combination of model 3. *PC*, prostate cancer. High-grade prostate cancer is defined as having Gleason score  $\geq 7$  prostate cancer. *B*, calibration of the nomogram model when predicting any cancer. *X axis*, histogram of the calculated probabilities for the testing data set. *Y axis*, actual, observed incidence (actual probability); *X axis*, probability calculated by the nomogram (predicted probability). If the model were perfect, *triangles* would lie on the *dotted line* with a slope of 1.



adjusting for age, family history of prostate cancer, ethnicity, presence of urinary symptoms, PSA, free:total PSA ratio, and DRE, the OR was 5.1 (95% CI, 1.6-16.5;  $P = 0.006$ ).

We also restricted our analysis to subjects from Caucasian descent only for models 1 to 3. We could not examine other ethnic groups due to sample size limitations for a four-SNP combination setting. Among White subjects, the OR for prostate cancer did change in magnitude, but their statistical significance remained (Table 5).

**Clinical application of genotypes in assessing prostate cancer risk.** From multivariate receiver operating characteristic analysis, the AUC for the baseline model that included age, family history of prostate cancer, ethnicity, presence of urinary voiding symptoms, PSA level, free:total PSA ratio, and DRE was 0.72 (95% CI, 0.70-0.74). When adding the SNPs from model 1 (five-SNP combination by Zheng et al.) to the multivariate

model, the AUC was 0.73 (95% CI, 0.71-0.75). The AUC from model 2 (four-SNP combination from second panel) was also 0.73 (95% CI, 0.71-0.74). For the SNPs in model 3, the multivariate AUC was 0.74 (95% CI, 0.72-0.76), which was significantly higher than the AUC of the baseline model (contrast test for difference,  $P = 0.001$ ).

To develop an instrument for these genotypes to be used in a clinical setting, we developed a nomogram that incorporates the four-SNP combination (model 3) with age, family history of prostate cancer, ethnicity, urinary voiding symptoms, PSA level, free:total PSA ratio, and DRE in predicting all prostate cancer and intermediate to high-grade cancer (Gleason score  $\geq 7$ ; Fig. 1). We evaluated the accuracy of a nomogram in predicting the presence of prostate cancer and aggressive prostate cancer by comparing the predicted and actual probabilities for prostate cancer and aggressive prostate cancer

in the validation set. We examined the incremental drop in AUC from the each of the variables based on the nomogram results, which we showed previously to be a method to quantify how much each factor contributes to the model (9). We examined how the AUC of the predictive model would drop by removing the SNP genotype combination and compared it with the incremental drops of established variables including age, family history of prostate cancer, ethnicity, urinary symptoms, PSA, free:total PSA ratio, and DRE. We found that the incremental drop in AUC for the SNP genotype combination was greater (and therefore contributed more predicative information) than PSA, family history of prostate cancer, ethnicity, urinary symptoms, and DRE (Table 6).

Finally, we also examined how the inclusion of the four-SNP combination could affect the positive predictive values (PPV) of the PSA level. Using the standard cutoff of 4.0 ng/mL, the PPV of PSA among all patients was 49%. The PPV changed from 42% to 94% based on the number of variant genotypes present ( $P < 1 \times 10^{-15}$ , Table 7).

### Discussion

This study shows that SNP genotyping can be applied in a clinical setting to predict the presence of prostate cancer among men screened with PSA and DRE. We confirm that several of the SNPs of Zheng et al. are associated with the presence of prostate cancer at biopsy. We were able to replicate, in an ethnically mixed population, positive associations for 5 of the 10 SNPs presented by Zheng et al. and have extended the study to include 9 other SNPs. Our best predictive model included 2 of the SNPs from Zheng et al. and 2 SNPs not included in the Zheng et al. model. Men who carried all four genetic variants experienced an OR of 6.1. Using the 4 SNPs in combination with PSA, the PPV of the PSA test ranged from 41.5% to 93.8% based on the number of variants carried. Incorporating additional factors, such as family history, age, ethnic group, PSA, free:total PSA ratio, and DRE improved the model further as shown by the nomogram.

In our second panel of SNPs that we have established previously, it would be important to have other independent studies validate our findings. Future studies will be necessary to examine this. Until this is accomplished, it would be desirable

**Table 6.** Comparison of incremental drop in the AUC for each predictor variable when removed from the nomogram model

Factor	Incremental drop in AUC
SNP combination from model 3	0.014
Age	0.022
Family history of prostate cancer	0.003
Symptom score	0.001
PSA	0.001
Free:total PSA ratio	0.066
DRE	0.010

NOTE: The nomogram model includes the SNP combination from model 3.

**Table 7.** Changes in PPV of the PSA test based on the established cutoff level of 4.0 ng/mL using the genotype combination from model 3

Genotype combination	PSA cutoff (ng/mL)	No. cancer	No. without cancer	PPV* (%)
0	<4.0	17	74	41.5
	≥4.0	219	308	
1	<4.0	39	108	47.4
	≥4.0	441	490	
2	<4.0	27	59	53.4
	≥4.0	310	271	
3	<4.0	11	9	63.3
	≥4.0	91	53	
4	<4.0	0	3	93.8
	≥4.0	15	1	

\* $\chi^2$  for Mantel-Haenszel test of heterogeneity = 56.8 ( $P < 0.0001$ ).

to include all the SNPs from our second panel within the analysis. However, because our goal was to evaluate and validate the SNP combination model established by Zheng et al., we could only restrict our number of SNPs to four per model. Using more than four SNPs would result in very rare allelic combinations that the sample size could not support.

In the Zheng et al. study, the ORs associated with the variant genotypes ranged from 1.1 to 1.5. In our study, 5 of the 10 SNPs studied by Zheng et al. achieved statistical significance and the ORs, in general, were closer to unity. This is not surprising, given that Zheng et al. selected those SNPs with the most extreme ORs and *P* values; it is expected that the corresponding values in follow-up studies should be closer to unity.

We studied an unselected and heterogeneous group of men who were part of a PSA screening program. To be clinically useful, SNPs should predict cancer in a range of ethnic groups. Although the majority of men were of Caucasian descent, 18% of men were of African, Asian, or mixed ancestry. Our nomogram model includes both ethnic background and genotype status. Given the wide range of ethnic groups represented in Toronto, it is expected that our results will be similar in other large population centers in North America. However, in countries with different ethnicities, it is important that the choice of SNPs be validated.

Despite the confirmatory nature of our findings, the clinical utility of the multigenic SNP approach is inherently limited. The predictive ability of our basic model improved to only a small extent when the four SNPs were introduced (AUC, 0.72-0.74). Among men with no variant allele (in model 3), the PPV of a PSA test above 4.0 was 42%; at this level, a biopsy could still be appropriate. It is difficult to compare models that incorporate different combinations of SNPs because the baseline category (the referent group of unexposed men) changes. For example, the reference group contained 28% of all men in model 1, 7% in model 2, and 24% in model 3.

Ideally, a genetic classifier would predict that a high proportion of cancer cases would fall into a relatively small group of men, which is enriched for the genotypes of risk. Clearly, it would be valuable if we could identify a subgroup

of 50% of the men who had 90% of the cancers, but this would be equivalent to an OR of 10. We did not approach this level of discrimination in our study; for example, in model 3, 39% of the patients fell into the top 34% of the population, classified by genetic risk. The number of SNPs identified through genome-wide association studies continues to increase. We expect that other models will incorporate newly identified SNPs; however, the basic results presented here will not change unless the corresponding ORs are dramatically higher.

Multigene SNP studies tend to emphasize ORs for cancer for individuals in the highest risk class compared with individuals in the class with no variant genotype. Our data show why the utility of the multigenic approach to risk estimation is limited. Only a small proportion of men will carry all variant genotypes, where the OR for prostate cancer is the highest. If additional SNPs were to be added to the model, the OR for men in the highest risk group would be increased, but the number of men included in this group would be proportionately smaller. Therefore, it is not possible to achieve the desired levels of discrimination by adding more SNPs. This is shown by the overall contribution in AUC of SNPs where it is relatively small despite the high ORs for the highest risk class.

Nevertheless, despite its limitations, we have shown the clinical utility of SNP genotyping combinations through the use

of a nomogram platform. Prostate cancer risk calculators are being used clinically online based on data from the Prostate Cancer Prevention Trial (14) from the National Cancer Institute in the United States<sup>10</sup> and from the Sunnybrook Prostate Cancer Risk Calculator in Canada.<sup>11</sup> The characteristics of our study population are very similar to other screening populations used by the National Cancer Institute prostate cancer risk calculator in regards to ethnicity, family history of prostate cancer, and urinary symptoms (9, 14). The addition of SNP profiling could further enhance the ability to better assess prostate cancer risk. Indeed, although the addition of a the SNP genotype combination did not make a large improvement in the AUC from the baseline model, the predictive ability of the SNP predictor variable was important within the overall multivariate model as shown by examining the incremental drop in AUC (Table 6).

Finally, notwithstanding a nomogram approach, the clinical use of SNP combinations can be illustrated by examining the change in positive predictive of PSA. SNP combination may be able to effectively complement and enhance the predictive value of the PSA test as shown in Table 7.

In summary, it is possible to incorporate a small number of SNPs in a clinical setting to improve the PPV of the PSA test. Further clinical utility can be used through a nomogram approach for prostate cancer risk calculators.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

<sup>10</sup> <http://www.compass.fhcr.org/edrnnci/bin/calculator/main.asp?t=prostate&sub=disclaimer&v=prostate&m=&x=Prostate%20Cancer>

<sup>11</sup> <http://prostatecancer.ca>

#### References

- 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.
- Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007;39:631–7.
- 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;39:645–9.
- Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007;39:638–44.
- Nam RK, Zhang WW, Loblaw DA, et al. A genome-wide association screen identifies regions on chromosomes 1q25 and 7p21 as risk loci for sporadic prostate cancer. *Prostate Cancer Prostatic Dis* 2007;11:241–6.
- Nam RK, Zhang WW, Jewett MA, et al. The use of genetic markers to determine risk for prostate cancer at prostate biopsy. *Clin Cancer Res* 2005;11:8391–7.
- Nam RK, Zhang WW, Klotz LH, et al. Variants of the hK2 protein gene (KLK2) are associated with serum hK2 levels and predict the presence of prostate cancer at biopsy. *Clin Cancer Res* 2006;12:6452–8.
- Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. *N Engl J Med* 2003;349:215–24.
- Nam RK, Toi A, Klotz LH, et al. Assessing individual risk for prostate cancer. *J Clin Oncol* 2007;25:3582–8.
- Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* 1974;111:58–64.
- Barry MJ, Fowler FJJ, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol* 1992;148:1549–57.
- Nam RK, Sugar L, Yang W, et al. Expression of the TMPRSS2:ERG fusion gene predicts cancer recurrence after surgery for localized prostate cancer. *Br J Cancer* 2007;97:1690–5.
- Harrell FE. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag; 2001.
- Thompson IM, Ankerst DP, Chi C, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *J Natl Cancer Inst* 2006;98:529–34.