Original Article - Urological Oncology

Investig Clin Urol 2022;63:42-52.
https://doi.org/10.4111/icu.20210305
pISSN 2466-0493 • eISSN 2466-054X

INVESTIGATIVE AND CLINICAL UROLOGY
ICUROLOGY

Prediction of clinically significant prostate cancer using polygenic risk models in Asians

Sang Hun Song[1,2,*], Eunae Kim[3,*], Eunjin Woo[3], Eunkyung Kwon[1,3], Sungroh Yoon[4], Jung Kwon Kim[1], Hakmin Lee[1], Jong Jin Oh[1,2], Sangchul Lee[1], Sung Kyu Hong[1,2], Seok-Soo Byun[1,3,5]

[1]Department of Urology, Seoul National University Bundang Hospital, Seongnam, [2]Department of Urology, Seoul National University College of Medicine, Seoul, [3]Procagen, Seongnam, [4]Department of Electrical and Computer Engineering, Seoul National University, Seoul, [5]Department of Medical Device Development, Seoul National University College of Medicine, Seoul, Korea

Purpose: To develop and evaluate the performance of a polygenic risk score (PRS) constructed in a Korean male population to predict clinically significant prostate cancer (csPCa).

Materials and Methods: Total 2,702 PCa samples and 7,485 controls were used to discover csPCa susceptible single nucleotide polymorphisms (SNPs). Males with biopsy-proven or post-radical prostatectomy Gleason score 7 or higher were included for analysis. After genotype imputation for quality control, logistic regression models were applied to test association and calculate effect size. Extracted candidate SNPs were further tested to compare predictive performance according to number of SNPs included in the PRS. The best-fit model was validated in an independent cohort of 311 cases and 822 controls.

Results: Of the 83 candidate SNPs with significant PCa association reported in previous literature, rs72725879 located in *PRNCR1* showed the highest significance for PCa risk (odds ratio, 0.597; 95% confidence interval [CI], 0.555–0.641; $p=4.3\times10^{-45}$). Thirty-two SNPs within 26 distinct loci were further selected for PRS construction. Best performance was found with the top 29 SNPs, with AUC found to be 0.700 (95% CI, 0.667–0.734). Males with very-high PRS (above the 95th percentile) had a 4.92-fold increased risk for csPCa.

Conclusions: Ethnic-specific PRS was developed and validated in Korean males to predict csPCa susceptibility using the largest csPCa sample size in Asia. PRS can be a potential biomarker to predict individual risk. Future multi-ethnic trials are required to further validate our results.

Keywords: Genome-wide association study; Multifactorial inheritance; Polymorphism, single nucleotide; Prostatic neoplasms

INTRODUCTION

The incidence of prostate cancer (PCa) is constantly on the rise, currently second in cancer diagnosis and sixth in cancer mortality in men worldwide as of 2020 [1]. This trend is more apparent in Asia where PCa has traditionally had low prevalence, with literature pointing to the gradual transition to a westernized high-fat intake diet and obesity increase as well as broader implementation of nationwide PCa screening as notable factors [2]. As such, genetic origins of PCa pathogenesis have been of keen clinical interest, and with the approval of PARP (poly [adenosine diphosphate-

Received: 3 August, 2021 • **Revised:** 18 September, 2021 • **Accepted:** 12 October, 2021 • **Published online:** 29 November, 2021
Corresponding Author: Seok-Soo Byun https://orcid.org/0000-0001-9356-9500
Department of Urology, Seoul National University Bundang Hospital, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea
TEL: +82-31-787-7342, FAX: +82-31-787-4057, E-mail: ssbyun@snubh.org
*These authors contributed equally to this study and should be considered co-first authors.

ribose] polymerase) inhibitors as novel therapy for actionable genetic mutations including BRCA 1/2, the National Comprehensive Cancer Network (NCCN) now recommends germline testing for PCa not only in men with a positive family history but also in high- to very-high risk regional or metastatic PCa. Alterations in homologous DNA repair and mismatch repair genes are strongly linked to 3- to 6-fold increase in overall risk [3].

However, not all PCa can be explained with nominal genes of high penetrance, but rather can be attributed to the cumulative effect of small variant alleles that confer varying levels of individual risk to disease [3,4]. With the pivotal study that identified 5 singular single nucleotide polymorphisms (SNPs) having strong association with metastatic PCa [5], the search for PCa susceptible variants have culminated in over 100 SNPs found in large-scale genome-wide association studies (GWASs), attributing to 60% of heritability [6]. Associations from germline mutations provide valuable information in predicting lifetime trajectory, allowing individualized medical care at an early stage of screening and diagnosis. Polygenic risk score (PRS) takes advantage of common polymorphisms to assess a person's genetic risk of disease, and when used with other clinical variables, has added value in anticipating tumor aggressiveness and escaping unnecessary biopsy [7,8].

Current literature on PRS in PCa are scarce compared to the vast number of reported SNPs with variable levels of performance, limiting application in clinical practice. Most large scale models are also constructed in males of mainly Caucasian and European ancestry, resulting in lower predictive power when replicated in Asian cohorts likely due to non-shared genomic loci [4]. As such, ethnic-specific PRS is required to account for genetic variations. In this study, we developed and validated a PRS model predicting clinically significant PCa (csPCa) in a Korean population.

MATERIALS AND METHODS

1. Study population and genotyping

Participants of this study include PCa patients who received radical prostatectomy (RP) and/or core transrectal ultrasound prostate biopsy at one of the four tertiary hospitals in the Republic of Korea (Seoul National University of Bundang Hospital, Seoul National University Hospital, Chungbuk National University Hospital, and the Catholic University of Korea, St. Vincent's Hospital). All participants provided informed consent. This multi-center study was approved by the Institutional Review Board of Seoul National University Bundang Hospital (SNUBH) (approval number: B-1607/355-302). Gleason score (GS) of PCa was reviewed based on biopsy and/or RP specimens. For selection of csPCa cases, patients with GS 7 or higher were included for this study. Intact genomic DNA (200 ng) extracted from blood or saliva samples was genotyped using the Korea Biobank array (K-CHIP) following the manufacturer's instructions.

The control genotype and phenotype data were obtained from the Korean Genome and Epidemiology Study (KoGES) conducted by the Center for Genome Science, the Korea National Institute of Health (KNIH) [9]. The genotype data were produced using K-CHIP array. The control samples were further selected for analyses who are male, 60 years or older and had never been diagnosed with any cancer. More detailed information about the external validation cohort is available in a previously published article [9].

2. Quality control of genotype data and genotype imputation

All sample and marker quality control (QC) was performed using PLINK software (v1.90 beta) [10]. Samples were excluded based on the following criteria: i) low genotype call rate (<95%), ii) excessive heterogeneity, iii) genetic relatedness, and iv) sex inconsistencies. Markers i) that are not SNPs, ii) MAF <5%, iii) low call rate (<95%), or iv) significantly deviated from Hardy–Weinberg equilibrium ($p<1.0\times10^{-6}$) were excluded. Imputation of missing genotypes in the PCa cases were conducted using the Michigan Imputation Server which utilizes Eagle v2.4 for phasing and Minimac 4 with the 1000 Genomes Project Phase 3 as the reference genome panel for genotype imputation [11,12]. Genotype-imputed SNPs were filtered with genotype quality (R^2>0.8) for the further analyses. For control samples, the genotype data that had been imputed were provided by KNIH, after phasing using Eagle v2.3 and genotype imputation using IMPUTE 4 with 1000 Genomes project phase 3 and Korean reference genome (397 samples) as the reference panel [11,12]. The imputed SNPs were filtered based on INFO score >0.8 and MAF >0.1. The imputed genotype data for cases and controls were merged after SNPs commonly found in both cases and controls were kept for the further analyses. The post-imputation QC was performed based on the same criteria applied in the pre-imputation QC. After sample and marker QC, the resulting 11,320 samples (3,013 cases and 8,307 controls) and 4,724,872 SNPs were included for the downstream analyses.

3. Statistical analysis

A total of 10,187 samples (2,702 cases and 7,485 controls) were used for discovery of PCa-associated SNPs and their

Table 1. Characteristics of the study population

Variable	All (n=11,320)		Training set (n=10,187)		Test set (n=1,133)	
	Case (n=3,013)	Control (n=8,307)	Case (n=2,702)	Control (n=7,485)	Case (n=311)	Control (n=822)
Age (y)	67.8±7.5	64.7±3.6	67.6±7.5	64.7±3.6	68.8±7.7	64.9±3.6
BMI (kg/m^2)	24.6±2.7	24.1±2.7	24.6±2.7	24.1±2.7	NA	24.2±2.8
PSA (ng/mL)	35.1±209.4	NA	36.7±221.1	NA	21.3±31.0	NA
Gleason score						
7	2,105	-	1,928	-	177	-
8	411	-	338	-	73	-
≥9	497	-	436	-	61	-

Values are presented as mean±standard deviation or number only.
BMI, body mass index; PSA, prostate-specific antigen; NA, not available.

summary statistics. We performed association tests between csPCa and previously reported PCa-associated SNPs. Previously reported SNPs were extracted after literature review of GWAS studies on PCa. Logistic regression models were applied to test associations for each of those SNPs. Based on the association results, SNPs were filtered based on statistical significance of p<0.001. Linkage disequilibrium (LD)-clumping was performed to extract a set of lead SNPs within each LD block.

Based on the association results, we obtained odds ratio (OR) of PCa-associated SNPs on the risk of PCa development. PRS was calculated as log(OR)-weighted sum of the number of risk alleles as following:

$$S_{ij} = \frac{\sum_{j=1}^{M} X_{ij}\hat{\beta}_j}{n}$$

X_{ij} : number of risk alleles for the j^{th} variant (0,1 or 2) of i^{th} individual,

β_j : weighting (log(OR))for the j^{th} variant,

n : the total number of the variants calculated for PRS.

The individual PRS was computed upon the training set for internal validation and the test set for external validation.

To test performance of PRS models in predicting risk of PCa development, the test set (311 cases and 822 controls) was used. The predictive performance of PRS was evaluated using the area under the ROC (receiver operating characteristics) curve (AUC) [13]. The AUC was compared according to the number of SNPs included for PRS models. Improvement in AUC between ROC curves was tested using Delong's method [6]. The optimal PRS cutoff value was determined at the maximal Youden's Index (*J*, sensitivity+specificity-1) value.

For calculation of individual risk, we fitted the logistic regression model to obtain the regression coefficient of PRS on PCa risk and calculated the individual risk with $OR_i = \exp^{(\beta PRS*(PRS_i-\mu_{PRS})}$, where β_{PRS} is the regression coefficient of PRS on PCa risk, PRS_i is the i^{th} individual's PRS, μ_{PRS} is the mean PRS of the controls.

RESULTS

The mean age of PCa cases and controls were 67.8 and 64.7 years, respectively. The mean BMI of cases and controls was 24.6 kg/m^2 and 24.1 kg/m^2, respectively. The mean prostate-specific antigen (PSA) level of PCa cases were 35.1 ng/mL with standard deviation of 209.4 ng/mL. Most of PCa cases had GS 7 (n=2,105), followed by GS 9 or higher (n=497) and GS 8 (n=411) (Table 1).

We identified 83 SNPs that were previously reported to be PCa-associated in one or more studies (Table 2). rs72725879 located within *PRNCR1* was most significantly associated with PCa risk (OR, 0.597; 95% confidence interval [CI], 0.555–0.641; $p=4.3\times10^{-45}$). Other 38 SNPs also showed negative associations between minor allele and PCa risk. Of the 83 SNPs, 23 SNPs were located within *8q24.21* region, spanning *PRNCR1*, *CASC8*, *PCAT2*, *PCAT1*, *LOC105375751*, and *CCAT2*, 8 SNPs were located within *HNF1B* (*17q12*) and 5 SNPs were located within *11q13.3*. After LD clumping, 31 SNPs located within 26 distinct genomic loci had been extracted as candidate SNPs for PRS construction (Table 3).

We calculated individual PRS in the test set (311 cases and 822 controls) and tested the performance in predicting PCa risk (Fig. 1). PRS models composed of top 29 SNPs, top 26 SNPs and top 28 SNPs showed comparably superior performance compared to other models (Table 4). The AUC of the prediction model of PRS composed of 29 SNPs was estimated to be 0.700 (95% CI, 0.667–0.734), with a sensitivity and specificity of 0.672 and 0.662, respectively (Fig. 2A).

From the best performing PRS model composed of top 29 SNPs, we chose two different cutoff values, i) the optimal

https://doi.org/10.4111/icu.20210305

Table 2. SNPs associated with development of PCa

SNP	CHR	BP	Minor allele	OR (95% CI)	p-value	Locus
rs72725879	8	128103969	C	0.597 (0.555–0.641)	4.3E-45	*PRNCR1*
rs4242384	8	128518554	C	1.761 (1.628–1.906)	4.7E-45	*8q24.21*
rs7843031	8	128533473	T	1.790 (1.650–1.942)	1.1E-44	*8q24.21*
rs7837688	8	128539360	T	1.784 (1.645–1.935)	3.5E-44	*8q24.21*
rs1456315	8	128103937	C	0.584 (0.541–0.630)	8.1E-44	*PRNCR1*
rs4582524	8	128528435	G	1.748 (1.613–1.894)	2.3E-42	*8q24.21*
rs10090154	8	128532137	T	1.703 (1.570–1.846)	5.1E-38	*8q24.21*
rs11986220	8	128531689	A	1.690 (1.559–1.832)	4.6E-37	*8q24.21*
rs13254738	8	128104343	A	0.633 (0.590–0.680)	4.6E-36	*PRNCR1*
rs1447295	8	128485038	A	1.598 (1.476–1.730)	4.3E-31	*CASC8*
rs56005245	8	128113426	T	1.435 (1.342–1.533)	1.8E-26	*8q24.21*
rs1016343	8	128093297	T	1.417 (1.328–1.512)	7.0E-26	*PRNCR1/PCAT2*
rs12682344	8	128106784	G	1.441 (1.342–1.547)	6.6E-24	*8q24.21*
rs6983561	8	128106880	C	1.439 (1.341–1.545)	8.8E-24	*8q24.21*
rs16901979	8	128124916	A	1.413 (1.317–1.516)	6.3E-22	*8q24.21*
rs10505483	8	128125195	T	1.413 (1.317–1.516)	6.9E-22	*8q24.21*
rs11263763	17	36103565	G	0.700 (0.650–0.753)	1.0E-21	*HNF1B*
rs8064454	17	36101586	A	0.700 (0.651–0.753)	1.0E-21	*HNF1B*
rs11651052	17	36102381	A	0.702 (0.652–0.755)	1.6E-21	*HNF1B*
rs7501939	17	36101156	T	0.704 (0.654–0.757)	5.1E-21	*HNF1B*
rs13252298	8	128095156	G	0.712 (0.662–0.765)	2.8E-20	*PRNCR1/PCAT2*
rs4871009	8	128108416	T	1.367 (1.277–1.465)	4.4E-19	*8q24.21*
rs4430796	17	36098040	G	0.743 (0.692–0.797)	9.4E-17	*HNF1B*
rs2005705	17	36096300	A	0.746 (0.695–0.802)	1.2E-15	*HNF1B*
rs339331	6	117210052	C	0.775 (0.724–0.829)	1.4E-13	*RFX6*
rs339351	6	117200434	A	0.778 (0.727–0.832)	2.8E-13	*RFX6*
rs1512268	8	23526463	T	1.276 (1.195–1.363)	3.8E-13	*LOC107986930*
rs1160267	8	23529521	G	1.272 (1.191–1.358)	7.9E-13	*LOC107986930*
rs12549761	8	128540776	G	0.762 (0.706–0.823)	3.2E-12	*8q24.21*
rs11649743	17	36074979	A	0.781 (0.729–0.838)	4.0E-12	*HNF1B/LOC105371754*
rs10503733	8	23534018	T	1.246 (1.167–1.331)	6.0E-11	*LOC107986930*
rs11125927	2	62752975	G	1.247 (1.165–1.336)	2.3E-10	*2p15*
rs140783917	10	122834482	T	0.748 (0.683–0.818)	2.5E-10	*LOC105378521*
rs7821330	8	23522452	C	1.228 (1.150–1.310)	6.3E-10	*LOC107986930*
rs7489409	13	73716861	C	1.213 (1.139–1.292)	1.5E-09	*13q22.1*
rs10896449	11	68994667	G	1.404 (1.253–1.574)	5.3E-09	*11q13.3*
rs7463326	8	128027954	A	0.788 (0.723–0.858)	4.4E-08	*PCAT1/LOC105375751*
rs7929962	11	68985583	T	1.375 (1.226–1.543)	5.7E-08	*11q13.3*
rs376592364	11	69011693	T	1.383 (1.229–1.557)	7.6E-08	*11q13.3*
rs10086908	8	128011937	C	0.796 (0.732–0.867)	1.3E-07	*LOC105375751*
rs8023793	15	66942093	C	0.837 (0.783–0.895)	1.5E-07	*LINC01169*
rs12270641	11	69012244	A	1.374 (1.219–1.549)	2.1E-07	*11q13.3*
rs2659124	19	51354597	A	0.845 (0.792–0.902)	4.3E-07	*LOC105372441*
rs58235267	2	63277843	C	0.843 (0.788–0.903)	8.9E-07	*OTX1*
rs10993994	10	51549496	T	1.161 (1.091–1.235)	2.4E-06	*MSMB*
rs11228583	11	69009114	T	1.348 (1.190–1.527)	2.6E-06	*11q13.3*
rs77167534	2	173319930	T	0.833 (0.771–0.899)	3.3E-06	*ITGA6*
rs9600079	13	73728139	T	1.158 (1.087–1.233)	5.0E-06	*13q22.1*
rs7591218	2	43637998	G	0.850 (0.791–0.912)	6.6E-06	*THADA*
rs2735839	19	51364623	A	0.862 (0.808–0.920)	7.3E-06	*19q13.33*

Table 2. Continued

SNP	CHR	BP	Minor allele	OR (95% CI)	p-value	Locus
rs1983891	6	41536427	T	1.155 (1.083–1.233)	1.3E-05	*FOXP4*
rs4714485	6	41536587	G	1.154 (1.081–1.231)	1.5E-05	*FOXP4*
rs2242652	5	1280028	A	0.835 (0.769–0.906)	1.7E-05	*TERT*
rs11817544	10	80236999	A	0.837 (0.772–0.909)	2.0E-05	*10q22.3*
rs12621278	2	173311553	G	0.849 (0.787–0.916)	2.3E-05	*ITGA6*
rs146618443	2	173309803	T	0.851 (0.789–0.918)	3.1E-05	*ITGA6*
rs1038822	2	43738173	C	0.863 (0.804–0.926)	4.4E-05	*THADA*
rs75204040	2	62780325	G	1.141 (1.070–1.215)	5.0E-05	*2p15*
rs10505477	8	128407443	A	1.139 (1.069–1.213)	5.6E-05	*CASC8*
rs6983267	8	128413305	G	1.138 (1.068–1.212)	6.0E-05	*CASC8/CCAT2*
rs817872	9	110144887	C	1.143 (1.070–1.221)	7.2E-05	*LOC107987111, RAD23B*
rs1465618	2	43553949	C	0.866 (0.806–0.930)	8.0E-05	*THADA*
rs56159348	11	76267331	G	0.836 (0.764–0.915)	1.0E-04	*11q13.5, EMSY*
rs74634457	15	66835704	G	1.202 (1.095–1.320)	1.1E-04	*ZWILCH*
rs4554825	10	80244623	T	0.853 (0.787–0.925)	1.2E-04	*10q22.3*
rs6660538	1	163295678	C	0.878 (0.821–0.940)	1.7E-04	*NUF2*
rs2252004	10	122844709	A	0.872 (0.812–0.937)	2.0E-04	*10q26.12*
rs4711748	6	43694598	T	1.125 (1.057–1.197)	2.1E-04	*POLR1C, RP1-261G23.5*
rs2659051	19	51345568	C	0.882 (0.825–0.943)	2.2E-04	*LOC105372441*
rs2238776	22	19757892	A	0.889 (0.835–0.947)	2.4E-04	*TBX1*
rs10807290	6	43710381	T	0.889 (0.835–0.947)	2.4E-04	*POLR1C*
rs6955627	7	92577760	T	0.879 (0.820–0.942)	2.6E-04	*7q21.2*
rs2660753	3	87110674	T	1.135 (1.060–1.216)	2.9E-04	*3p12.1*
rs17023900	3	87134800	G	1.163 (1.070–1.264)	3.8E-04	*3p12.1*
rs1283104	3	106962521	G	1.119 (1.051–1.192)	4.7E-04	*DUBR*
rs7153648	14	61122526	C	1.149 (1.062–1.244)	5.6E-04	*14q23.1*
rs9284813	3	87152169	G	1.130 (1.054–1.211)	6.1E-04	*LINC00506*
rs3110641	17	36047417	A	1.124 (1.051–1.202)	6.9E-04	*HNF1B/LOC105371755*
rs143745027	3	87144017	G	1.155 (1.062–1.256)	7.8E-04	*LINC00506*
rs8005621	14	61106699	G	1.160 (1.064–1.266)	7.9E-04	*SALRNA1*
rs12500426	4	95514609	A	1.113 (1.045–1.185)	8.0E-04	*PDLIM5*
rs6545977	2	63301164	A	0.881 (0.817–0.949)	8.5E-04	*2p15*
rs56103503	1	154980351	C	0.849 (0.771–0.936)	9.5E-04	*ZBTB7B*

SNP, single nucleotide polymorphism; PCa, prostate cancer; CHR, chromosome; BP, base pairs; OR, odds ratio; CI, confidence interval.

PRS value at the maximum Youden's index (-0.05) and ii) 95th percentile PRS cutoff (0.02) (Fig. 2B). We defined those with PRS below the optimal cutoff, those with PRS above 95th percentile cutoff and those with PRS between the two cutoff values as moderate, very-high and high-risk group, respectively. High PRS and very-high PRS group showed 3.80-fold and 5.93-fold increased risk, respectively, for PCa, compared to the moderate PRS group (the reference group) (Fig. 2C). The very-high PRS group had a 4.92-fold increased risk compared to the remaining population.

DISCUSSION

Based on the OR estimates of each SNP in PCa specific to a Korean population, we developed and evaluated PRS in Korean males. Of the 83 SNPs identified with significant association, 31 SNPs were selected for PRS construction. The best performing PRS composed of 29 SNPs reported an AUC of 0.700, with the top 5th percentile having more than a 4-fold increased risk of Grade Group (GG) ≥2 PCa. To note, minor alleles of many SNPs included in this study showed protective effects on PCa risk, which highlights that common risk alleles may have a cumulative impact in the development of PCa.

With trends of increasing prevalence worldwide, familial PCa and heritability is thoroughly researched for potential identification of actionable mutations. Hereditary PCa is caused by inherited genes of high penetrance, including

https://doi.org/10.4111/icu.20210305

Table 3. Candidate SNPs for PRS construction

SNP	CHR	BP	Minor allele	β (95% CI)	p-value	SNPs in LD block	Locus
rs72725879	8	128103969	C	-0.517 (-0.589 to -0.445)	4.3E-45	rs1016343, rs13252298, rs1456315, rs13254738, rs12682344, rs6983561, rs4871009, rs56005245, rs16901979, rs10505483	*PRNCR1*
rs4242384	8	128518554	C	0.566 (0.487 to 0.645)	4.7E-45	rs1447295, rs4582524, rs11986220, rs10090154, rs7843031, rs7837688, rs12549761	*8q24.21*
rs11263763	17	36103565	G	-0.357 (-0.431 to -0.284)	1.0E-21	rs3110641, rs11649743, rs2005705, rs4430796, rs7501939, rs8064454, rs11651052	*HNF1B*
rs339331	6	117210052	C	-0.255 (-0.323 to -0.188)	1.4E-13	rs339351	*RFX6*
rs1512268	8	23526463	T	0.244 (0.178 to 0.310)	3.8E-13	rs7821330, rs1160267, rs10503733	*LOC107986930*
rs11125927	2	62752975	G	0.221 (0.153 to 0.290)	2.3E-10	rs75204040	*2p15*
rs140783917	10	122834482	T	-0.291 (-0.381 to -0.201)	2.5E-10	rs2252004	*LOC105378521*
rs7489409	13	73716861	C	0.193 (0.130 to 0.256)	1.5E-09	rs9600079	*13q22.1*
rs10896449	11	68994667	G	0.339 (0.226 to 0.454)	5.3E-09	rs7929962, rs11228583, rs376592364, rs12270641	*11q13.3*
rs7463326	8	128027954	A	-0.239 (-0.324 to -0.153)	4.4E-08	rs10086908	*PCAT1/LOC105375751*
rs8023793	15	66942093	C	-0.178 (-0.244 to -0.111)	1.5E-07	rs74634457	*LINC01169*
rs2659124	19	51354597	A	-0.168 (-0.233 to -0.103)	4.3E-07	rs2659051, rs2735839	*LOC105372441*
rs58235267	2	63277843	C	-0.171 (-0.239 to -0.103)	8.9E-07	rs6545977	*OTX1*
rs10993994	10	51549496	T	0.149 (0.087 to 0.211)	2.4E-06	-	*MSMB*
rs77167534	2	173319930	T	-0.183 (-0.260 to -0.106)	3.3E-06	rs146618443, rs12621278	*ITGA6*
rs7591218	2	43637998	G	-0.163 (-0.234 to -0.092)	6.6E-06	rs1465618, rs1038822	*THADA*
rs1983891	6	41536427	T	0.144 (0.080 to 0.209)	1.3E-05	rs4714485	*FOXP4*
rs2242652	5	1280028	A	-0.180 (-0.262 to -0.098)	1.7E-05	-	*TERT*
rs11817544	10	80236999	A	-0.177 (-0.259 to -0.096)	2.0E-05	rs4554825	*10q22.3*
rs10505477	8	128407443	A	0.130 (0.067 to 0.193)	5.6E-05	rs6983267	*CASC8*
rs817872	9	110144887	C	0.134 (0.068 to 0.200)	7.2E-05	-	*LOC107987111, RAD23B*
rs56159348	11	76267331	G	-0.179 (-0.269 to -0.089)	1.0E-04	-	*11q13.5, EMSY*
rs6660538	1	163295678	C	-0.130 (-0.197 to -0.062)	1.7E-04	-	*NUF2*
rs4711748	6	43694598	T	0.118 (0.055 to 0.180)	2.1E-04	rs10807290	*POLR1C, RP1-261G23.5*
rs2238776	22	19757892	A	-0.118 (-0.181 to -0.055)	2.4E-04	-	*TBX1*
rs6955627	7	92577760	T	-0.129 (-0.198 to -0.060)	2.6E-04	-	*7q21.2*
rs2660753	3	87110674	T	0.127 (0.058 to 0.196)	2.9E-04	rs17023900, rs143745027, rs9284813	*3p12.1*
rs1283104	3	106962521	G	0.112 (0.050 to 0.176)	4.7E-04	-	*DUBR*
rs7153648	14	61122526	C	0.139 (0.060 to 0.218)	5.6E-04	rs8005621	*14q23.1*
rs12500426	4	95514609	A	0.107 (0.044 to 0.170)	8.0E-04	-	*PDLIM5*
rs56103503	1	154980351	C	-0.164 (-0.261 to -0.067)	9.5E-04	-	*ZBTB7B*

SNP, single nucleotide polymorphism; PRS, polygenic risk score; CHR, chromosome; BP, base pairs; CI, confidence interval; LD, linkage disequilibrium.

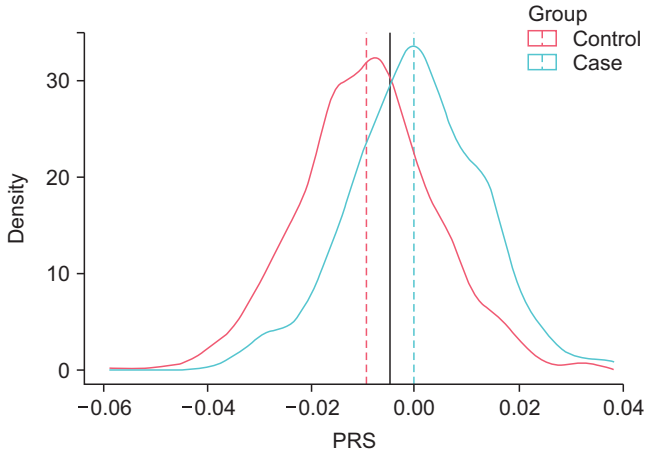

Fig. 1. Distribution of PRS in cases and controls. Mean PRS in controls: pink dashed line, Mean PRS in cases: blue dashed line, and optimal PRS cutoff: black solid line. PRS, polygenic risk score.

homologous DNA pair genes (e.g., BRCA1/2, ATM, CHEK2, PALB2) and mismatch repair genes (e.g., MLH1, MSH2, MSH6, PMS2), the latter related to Lynch syndrome. While such nominal mutations are easily identified and linked to PCa risk, not all familial PCa can be explained by singular mutations, but rather is accountable by the sum of multiple common variants with small effect size. PRS is a novel and only biomarker that uses the concept of polygenic risk as a single combinatory value estimate of the individual's genetic propensity of a complex disease to potentially stratify high- and low-risk patients in order to apply different methods of screening and prevention.

While the exact number required to achieve the full potential of PRS is unknown, summary statistics based on sufficient data is required to detect rare but significant variants shared between phenotypes and approximate the populational burden of disease. PCa is reported to have variants with relatively larger effect size and hence has been suggested to require a smaller sample to assess heritability compared to other, more polygenic cancers. While falling short of over 60,000 cases required to explain 80% of heritability as suggested by Zhang et al. (2020) [14], one of the largest cohorts of over 2,700 males in biopsy and RP-proven GS ≥7 PCa was utilized, compared to previous Korean [15] and Japanese [16] PRSs developed from 1,001 and 689 cases, respectively.

Selection of appropriate SNPs for inclusion critically influences PRS performance, due to obvious uncertainties inherent in estimating effect size from a potentially biased population. As such, statistical techniques must be applied prior to PRS development, a process necessary to weed out insignificant variants and to avoid overfitting or under-fitting of risk alleles. p-value thresholds are often implemented as a cut-off, and while $p<1\times10^{-5}$ is most commonly used, it is often an arbitrary value. A strict, parsimonious threshold risks the loss of otherwise powerful variants confounded by nearby SNPs resulting in false negatives, whereas an overly flexible threshold risks inclusion less predictive alleles as well as a resulting SNP combination too large to be feasible in actual practice. Hence, to optimize SNP selection, a relative lenient p-value threshold of $p<10^{-3}$ was applied to filter a relatively large set of potential candidates, followed by LD clumping to extract lead SNPs and quality control via imputation for missing genotypes as well as SNP combination comparisons, a similar approach described in our previous report [15].

To additionally bolster performance, our study selectively used previously reported, well-established PCa-susceptible SNPs, many found in the *8q24.21* region. *8q24* is commonly considered a "desert" region due to scarcity of genes, but has been the source of numerous PCa susceptible variants, contributing to 8% to 9% of two-fold increase in familial risk [17]. A proto-oncogene *Myc* is closest, and while polymorphisms in this region have been associated with regulatory enhancers and androgen receptor (AR) response, mechanisms of action are not yet well understood. Nonetheless, strong associations have been made in both European and East Asian populations [18,19], with rs7837688 and rs1016343 more common in Korean males [15], a finding comparable in this study as 4th and 12th in statistical significance. Eleven candidate SNPs were associated with *PRNCR1* in the *8q24* region, whose up-regulation is implicated in aggressive PCa via AR-mediated transcriptional programs [20]. Thirteen were located in *HNF1B* (*17q12*) and *11q13.3*, both well-documented as susceptible loci in males of European and Asian descent [21,22].

The ultimate goal of this research was to evaluate whether PRS can be utilized in actual clinical practice to avoid unnecessary intervention in low-risk PCa. Therefore, we selected patients with biopsy or pathology-proven csPCa with ≥GG2, excluding males under the age of 60. While generally preferred in the Korean population over active surveillance (AS) or watchful waiting, early intervention and surgery in low-grade PCa risks overtreatment and economic burden. As such, imaging modalities and biomarkers have been introduced to focus biopsy and treatment in csPCa, which generally includes any of histopathology ISUP grade ≥2 (≥GG2) and/or tumor volume ≥0.5 cc [23]. Our study, selecting only males of intermediate- to high-risk PCa category, showed performance comparable to previous literature. Seibert et al. (2018) [7], defining aggressive PCa as any cancer not eligible for surveillance based on the NCCN guidelines (i.e., any ≥GG2, stage T3–4, PSA ≥10 ng/mL, and any nodal or distant

https://doi.org/10.4111/icu.20210305

Table 4. Predictive performance of PRS models for development of PCa

Number of top SNPs included for PRS calculation	Mean PRS of controls	Mean PRS of cases	AUC (95% CI)	J	Sensitivity	Specificity	Improvement in AUC (p)
29	-0.009	0.000	0.700 (0.667–0.734)	0.334	0.672	0.662	-
26	-0.014	-0.004	0.700 (0.667–0.734)	0.321	0.659	0.662	0 (0.993)
28	-0.010	-0.001	0.700 (0.666–0.733)	0.320	0.637	0.684	-0.001 (0.726)
27	-0.012	-0.003	0.698 (0.664–0.732)	0.323	0.646	0.676	-0.002 (0.411)
31	-0.008	0.001	0.697 (0.664–0.731)	0.313	0.598	0.715	-0.003 (0.263)
30	-0.007	0.002	0.697 (0.664–0.731)	0.314	0.675	0.639	-0.003 (0.115)
25	-0.013	-0.003	0.696 (0.663–0.730)	0.337	0.640	0.697	-0.004 (0.350)
17	-0.017	-0.004	0.692 (0.658–0.726)	0.310	0.688	0.622	-0.008 (0.292)
24	-0.012	-0.001	0.692 (0.658–0.726)	0.315	0.624	0.691	-0.009 (0.065)
18	-0.018	-0.005	0.690 (0.656–0.724)	0.319	0.666	0.653	-0.010 (0.173)
20	-0.015	-0.003	0.690 (0.656–0.724)	0.331	0.656	0.675	-0.011 (0.106)
22	-0.013	-0.002	0.689 (0.655–0.723)	0.315	0.675	0.640	-0.012 (0.035)
21	-0.012	-0.001	0.688 (0.654–0.722)	0.311	0.630	0.681	-0.012 (0.420)
13	-0.024	-0.008	0.687 (0.654–0.721)	0.305	0.588	0.717	-0.013 (0.168)
16	-0.021	-0.007	0.687 (0.653–0.722)	0.298	0.569	0.729	-0.013 (0.110)
23	-0.014	-0.004	0.687 (0.653–0.721)	0.312	0.624	0.689	-0.013 (0.013)
19	-0.019	-0.007	0.687 (0.653–0.721)	0.316	0.675	0.641	-0.013 (0.059)
12	-0.022	-0.004	0.684 (0.650–0.718)	0.309	0.643	0.665	-0.016 (0.094)
15	-0.019	-0.005	0.684 (0.650–0.718)	0.290	0.540	0.749	-0.016 (0.060)
14	-0.018	-0.003	0.683 (0.649–0.718)	0.293	0.572	0.720	-0.017 (0.061)
11	-0.018	0.000	0.683 (0.649–0.717)	0.292	0.656	0.636	-0.018 (0.090)
10	-0.013	0.006	0.680 (0.646–0.714)	0.284	0.621	0.663	-0.020 (0.064)
8	-0.014	0.009	0.675 (0.641–0.709)	0.283	0.637	0.646	-0.026 (0.030)
5	-0.040	-0.006	0.675 (0.640–0.709)	0.260	0.653	0.607	-0.026 (0.060)
7	-0.028	-0.002	0.674 (0.639–0.708)	0.264	0.778	0.485	-0.027 (0.027)
6	-0.024	0.005	0.672 (0.638–0.707)	0.270	0.595	0.675	-0.028 (0.029)
4	-0.070	-0.031	0.670 (0.636–0.705)	0.285	0.675	0.609	-0.030 (0.042)
9	-0.010	0.011	0.669 (0.635–0.704)	0.273	0.666	0.607	-0.031 (0.006)

PRS, polygenic risk score; PCa, prostate cancer; SNP, single nucleotide polymorphism; AUC, area under the receiver operating characteristics curve.

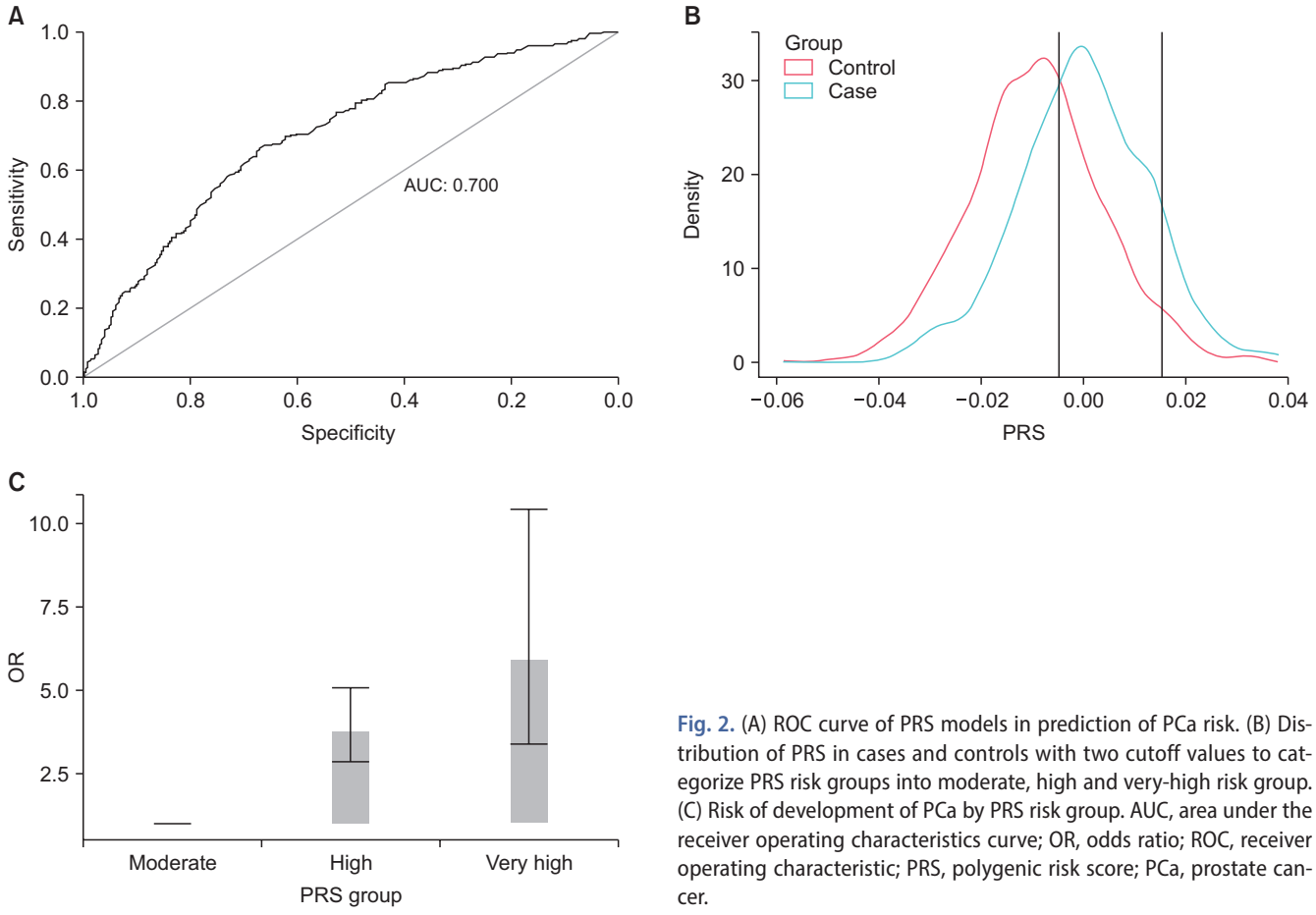

Fig. 2. (A) ROC curve of PRS models in prediction of PCa risk. (B) Distribution of PRS in cases and controls with two cutoff values to categorize PRS risk groups into moderate, high and very-high risk group. (C) Risk of development of PCa by PRS risk group. AUC, area under the receiver operating characteristics curve; OR, odds ratio; ROC, receiver operating characteristic; PRS, polygenic risk score; PCa, prostate cancer.

metastasis), developed a PRS from 54 SNPs to predict early age at onset of aggressive PCa (z=11.2, $p<10^{-16}$), with males in the 98th percentile having a 2.9-fold increased risk. Stockholm3 combined a PRS generated from 232 SNPs with 6 plasma proteins and other clinical variables to predict ≥GG2 PCa, outperforming PSA alone with a resulting AUC of 0.74 (95% CI, 0.72–0.75) [8]. Takata et al. (2019) [24] published results from a large Japanese cohort with a development set of 5,088 PCa cases from Biobank Japan and 10,682 controls from multiple institutions, but 1,806 (35.5%) patients had missing GS, and 1,293 (25.4%) had GS ≤6. Our effect size estimates calculated from csPCa better represent genetic polymorphisms that require intervention in actual practice.

Generalizability is key to the performance of a PRS, and population bias from a limited cohort poses a unique challenge in genomic prediction. Previously reported models are generated from cohorts of largely European ancestry and hence often fail to have the same predictive power in Asian males, primarily because unshared genomic loci, as well as variable population-specific effect size. A meta-analysis of multiancestry GWAS data found East Asians to have a 0.73-fold lower risk score compared to their European counterparts, highlighting the probability of potent germline variations among ethnic populations [25]. As such, accurate prediction of PCa risk requires tailoring to each race and ethnicity. Also, with increasing prevalence, even a modest performance is able to substantially stratify absolute populational risk, allowing application into clinical practice of early detection and prevention. AUC of 0.700 in our study is one of the highest discriminatory performances of PRS developed in an Asian cohort to date [4], compared to a previous report by our team conferring an AUC of 0.605 for any PCa utilizing 5 SNPs [15] and those developed in Japanese and Chinese populations with AUCs of 0.60 and 0.659, respectively [16,26]. The 4.92-fold increased risk of csPCa in the 95th percentile of our study is comparable to results validated in western counterparts with a 4 to 5-fold increased risk in the 99th percentile [18,27].

Despite successful replication of significant PCa susceptible polymorphisms and outperforming some previous models [15,16,28,29], our study is not without limitations. First, although we utilized data from 2,702 males with csPCa, it is still a relatively small sample size to accurately represent true populational variation. Second, because both

https://doi.org/10.4111/icu.20210305

development and validation cohorts consisted of Koreans, future evaluation in other race and ethnicities are required to fully assess predictive performance and generalizability. Due to the inherent selection bias in design, not all significant SNPs may be adequately represented in our data. In addition, while we limited the study population to csPCa, applying our model to males with any PCa at any age may further PRS utility, as both patients and clinicians can have additive information to guide screening schedules and opt for early intervention in AS-eligible males with high genetic risk.

Nonetheless, this study provides valuable evidence for prediction of csPCa based on the cumulative effects of small genetic variants specific to Asian populations. PRS calculated from individual GWAS data is the hallmark of modern precision medicine with the possibility of predicting a patient's lifetime trajectory for disease, offering the chance for early screening and personalized treatment, as in the case of coronary heart disease [30] where genetics is making great strides in real-world practice. Future acquisition of more sample summary statistics will further potentiate PRS performance in PCa and produce more accurate stratification of risk groups.

CONCLUSIONS

We successfully developed and validated PRS models for csPCa risk developed in a large sample of 2,702 ≥GS7 PCa cases and 7,485 healthy controls. A model utilizing the top 29 SNPs conferred the best predictive performance with an AUC 0.700, with over 4-fold risk predicted in males in above the 95th percentile of PRS. Future prospective, large-scale studies are required to further validate our results.

CONFLICTS OF INTEREST

Eunae Kim and Eunjin Woo are employees, Seok-Soo Byun is the chief executive officer, Eunkyung Kown is the chief operating officer and Jong Jin Oh is the inside director of Procagen. Sang Hun Song, Eunae Kim, Eunjin Woo, Eunkyung Kwon, Jung Kwon Kim, Hakmin Lee, Jong Jin Oh, Sangchul Lee, and Seok-Soo Byun have an equity interest in Procagen. All other authors report no conflicts or other disclosure.

ACKNOWLEDGMENTS

This work was supported by Grant No. 13-2016-007 and 14-2019-024 from the SNUBH Research Fund and the Technology Development Program (S2864304 and S3074168) from the Ministry of SMEs and Startups (MSS, Korea). This study was conducted with bioresources from National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2020-021). This paper was selected for the Best Paper Award at the 73rd Annual Meeting of the Korean Urological Association in 2021.

AUTHORS' CONTRIBUTIONS

Research conception and design: Eunae Kim and Seok-Soo Byun. Data acquisition: Jung Kwon Kim, Hakmin Lee, Jong Jin Oh, Sangchul Lee, Sung Kyu Hong, and Seok-Soo Byun. Statistical analysis: Eunae Kim and Eunjin Woo. Data analysis and interpretation: Sang Hun Song, Eunae Kim, and Eunkyung Kwon. Drafting of the manuscript: Sang Hun Song and Eunae Kim. Critical revision of the manuscript: Sang Hun Song, Eunae Kim, and Seok-Soo Byun. Obtaining funding: Seok-Soo Byun. Administrative, technical, or material support: Sungroh Yoon and Seok-Soo Byun. Supervision: Seok-Soo Byun. Approval of the final manuscript: all authors.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209-49.
2. Zhu Y, Mo M, Wei Y, Wu J, Pan J, Freedland SJ, et al. Epidemiology and genomics of prostate cancer in Asian men. Nat Rev Urol 2021;18:282-301.
3. Das S, Salami SS, Spratt DE, Kaffenberger SD, Jacobs MF, Morgan TM. Bringing prostate cancer germline genetics into clinical practice. J Urol 2019;202:223-30.
4. Song SH, Byun SS. Polygenic risk score for genetic evaluation of prostate cancer risk in Asian populations: a narrative review. Investig Clin Urol 2021;62:256-66.
5. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. N Engl J Med 2016;375:443-53.
6. Allemailem KS, Almatroudi A, Alrumaihi F, Makki Almansour N, Aldakheel FM, Rather RA, et al. Single nucleotide polymorphisms (SNPs) in prostate cancer: its implications in diagnostics and therapeutics. Am J Transl Res 2021;13:3868-89.
7. Seibert TM, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large

scale cohorts. BMJ 2018;360:j5757.
8. Möller A, Olsson H, Grönberg H, Eklund M, Aly M, Nordström T. The Stockholm3 blood-test predicts clinically-significant cancer on biopsy: independent validation in a multi-center community cohort. Prostate Cancer Prostatic Dis 2019;22:137-42.
9. Kim Y, Han BG; KoGES group. Cohort profile: the Korean genome and epidemiology study (KoGES) consortium. Int J Epidemiol 2017;46:e20.
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
11. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-7.
12. Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 2016;48:1443-8.
13. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
14. Zhang YD, Hurson AN, Zhang H, Choudhury PP, Easton DF, Milne RL, et al. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. Nat Commun 2020;11:3353.
15. Oh JJ, Lee SJ, Hwang JY, Kim D, Lee SE, Hong SK, et al. Exome-based genome-wide association study and risk assessment using genetic risk score to prostate cancer in the Korean population. Oncotarget 2017;8:43934-43.
16. Akamatsu S, Takahashi A, Takata R, Kubo M, Inoue T, Morizono T, et al. Reproducibility, performance, and clinical utility of a genetic risk prediction model for prostate cancer in Japanese. PLoS One 2012;7:e46454
17. Dupont WD, Breyer JP, Plummer WD, Chang SS, Cookson MS, Smith JA, et al. 8q24 genetic variation and comprehensive haplotypes altering familial risk of prostate cancer. Nat Commun 2020;11:1523.
18. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, et al. Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet 2009;41:1058-60.
19. Zhao CX, Liu M, Xu Y, Yang K, Wei D, Shi XH, et al. 8q24 rs4242382 polymorphism is a risk factor for prostate cancer among multi-ethnic populations: evidence from clinical detection in China and a meta-analysis. Asian Pac J Cancer Prev 2014;15:8311-7.
20. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature 2013;500:598-602.
21. Zhang YR, Xu Y, Yang K, Liu M, Wei D, Zhang YG, et al. Association of six susceptibility Loci with prostate cancer in Northern Chinese men. Asian Pac J Cancer Prev 2012;13:6273-6.
22. Zheng SL, Stevens VL, Wiklund F, Isaacs SD, Sun J, Smith S, et al. Two independent prostate cancer risk-associated Loci at 11q13. Cancer Epidemiol Biomarkers Prev 2009;18:1815-20.
23. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, version 2. Eur Urol 2016;69:16-40.
24. Takata R, Takahashi A, Fujita M, Momozawa Y, Saunders EJ, Yamada H, et al. 12 new susceptibility loci for prostate cancer identified by genome-wide association study in Japanese population. Nat Commun 2019;10:4422.
25. Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. Nat Genet 2021;53:65-75.
26. Ren S, Xu J, Zhou T, Jiang H, Chen H, Liu F, et al. Plateau effect of prostate cancer risk-associated SNPs in discriminating prostate biopsy outcomes. Prostate 2013;73:1824-35.
27. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet 2018;50:928-36.
28. Zhu Y, Han CT, Chen HT, Liu F, Zhang GM, Yang WY, et al. Influence of age on predictiveness of genetic risk score for prostate cancer in a Chinese hospital-based biopsy cohort. Oncotarget 2015;6:22978-84.
29. Oh JJ, Kim E, Woo E, Song SH, Kim JK, Lee H, et al. Evaluation of polygenic risk scores for prediction of prostate cancer in Korean men. Front Oncol 2020;10:583625.
30. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N Engl J Med 2016;375:2349-58.

https://doi.org/10.4111/icu.20210305