

## ARTICLE OPEN



# Predictive model of castration resistance in advanced prostate cancer by machine learning using genetic and clinical data: KYUCOG-1401-A study

Masaki Shiota<sup>1✉</sup>, Shota Nemoto<sup>2</sup>, Ryo Ikegami<sup>2</sup>, Shuichi Tatarano<sup>3</sup>, Toshiyuki Kamoto<sup>4</sup>, Keita Kobayashi<sup>5</sup>, Hideki Sakai<sup>6</sup>, Tsukasa Igawa<sup>7</sup>, Tomomi Kamba<sup>8</sup>, Naohiro Fujimoto<sup>9</sup>, Akira Yokomizo<sup>10</sup>, Seiji Naito<sup>10</sup> and Masatoshi Eto<sup>1</sup>

© The Author(s) 2024

**BACKGROUND:** The predictive power of the treatment efficacy and prognosis in primary androgen deprivation therapy (ADT) for advanced prostate cancer is not satisfactory. The objective of this study was to integrate genetic and clinical data to predict castration resistance in primary ADT for advanced prostate cancer by machine learning (ML).

**METHODS:** Clinical and single nucleotide polymorphisms (SNP) data obtained in the KYUCOG-1401-A study (UMIN000022852) that enrolled Japanese patients with advanced prostate cancer were used. All patients were treated with primary ADT. A point-wise linear (PWL) algorithm, logistic regression with elastic-net regularization, and eXtreme Gradient Boosting were the ML algorithms used in this study. Area under the curve for castration resistance and C-index for prognoses were calculated to evaluate the utility of the models.

**RESULTS:** Among the three ML algorithms, the area under the curve values to predict castration resistance at 2 years was highest for the PWL algorithm with all the datasets. Three predictive models (clinical model, small SNPs model, and large SNPs model) were created by the PWL algorithm using the clinical data alone, and 2 and 46 SNPs in addition to clinical data. C-indices for overall survival by the clinical, small SNPs, and large SNPs models were 0.636, 0.621, and 0.703, respectively.

**CONCLUSION:** The results demonstrated that the SNPs models created by ML produced excellent prediction of castration resistance and prognosis in primary ADT for advanced prostate cancer, and will be helpful in treatment choice.

*BJC Reports*; <https://doi.org/10.1038/s44276-024-00093-3>

## BACKGROUND

Androgen deprivation therapy (ADT) is widely used as the backbone therapy for advanced prostate cancer [1]. Current intensive therapies for metastatic prostate cancer include radiation, docetaxel, and novel androgen receptor signaling inhibitors, such as abiraterone, darolutamide, enzalutamide, and apalutamide, in addition to ADT [2]. Furthermore, triplet combination therapy, which adds such novel androgen receptor signaling inhibitors to ADT plus docetaxel, has recently been shown to prolong survival in metastatic prostate cancer [3, 4]. Therefore, prognostic estimation in ADT can help in choosing the best treatment. However, prognostic estimation in ADT using clinical parameters such as prostate-specific antigen (PSA), Gleason score, and TNM category is not satisfactory for producing C-indices of >0.7 [5–7]. Therefore, novel predictive models to more precisely estimate the response to ADT for advanced prostate cancer are needed.

Genetic background has been suggested to affect the efficacy and prognosis in ADT for prostate cancer, as indicated by different outcome among different ethnicities and consistent outcome within families [8–10]. Over the past few decades, genome wide association studies (GWAS) have discovered associations between single nucleotide polymorphisms (SNPs) and various features [11]. In a previous study, we investigated the association between the SNPs and prognosis in Japanese patients undergoing primary ADT for advanced prostate cancer by GWAS [12]. In that study, two SNPs, rs76237622 in *PRR27* and rs117573572 in *MTAP*, were validated to be associated with prognosis in ADT, but their predictive ability was not satisfactory [12].

Machine learning (ML) is a statistics-free approach that uses algorithms to identify patterns in rich and unwieldy data [13]. ML can resolve complex datasets of high dimensionality such as genomic data [14, 15]. In this study, we aimed to integrate genetic and clinical data that were obtained in the previous study [12], to

<sup>1</sup>Department of Urology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. <sup>2</sup>Industrial & Digital Business Unit, Hitachi, Ltd., Tokyo, Japan. <sup>3</sup>Department of Urology, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan. <sup>4</sup>Department of Urology, Faculty of Medicine, Miyazaki University, Miyazaki, Japan. <sup>5</sup>Department of Urology, Graduate School of Medicine, Yamaguchi University, Ube, Japan. <sup>6</sup>Department of Urology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan. <sup>7</sup>Department of Urology, School of Medicine, Kurume University, Kurume, Japan. <sup>8</sup>Department of Urology, Kumamoto University, Kumamoto, Japan. <sup>9</sup>Department of Urology, School of Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan. <sup>10</sup>Department of Urology, Harasanshin Hospital, Fukuoka, Japan. ✉email: [shiota.masaki.101@m.kyushu-u.ac.jp](mailto:shiota.masaki.101@m.kyushu-u.ac.jp)

Received: 26 June 2024 Revised: 13 August 2024 Accepted: 20 August 2024

Published online: 09 September 2024

predict castration resistance in primary ADT for advanced prostate cancer by ML.

## METHODS

### Study population

Japanese patients with *de novo* advanced prostate cancer (TanyN1M0 or TanyNanyM1) enrolled in the KYUCOG-1401-A study (UMIN000022852) that was conducted in conjunction with a prospective multi-institutional clinical trial (KYUCOG-1401; UMIN000014243, JRCTs071180035) were included in this study. Inclusion and exclusion criteria for the KYUCOG-1401-A study have been described previously [12]. Patients ( $n = 8$ ) censored before 2 years were excluded from this study. In the KYUCOG-1401 study, patients were randomized to receive gonadotropin-releasing hormone (GnRH) antagonist (degarelix) or agonist (leuporelin or goserelin) plus the antiandrogen bicalutamide [16]. This study was conducted in accordance with the Declaration of Helsinki and the Japanese Ethical Guidelines for Medical and Health Research Involving Human Subjects. Eligible patients provided written informed consent. This study was approved by the Kyushu University review board (23087-00).

### Clinical data

Clinicopathological information and efficacy in treatment data were collected prospectively using an electronic data capture system, as described previously [12]. Progression was defined as PSA progression (defined as PSA level of 2.0 ng/mL or higher, a rise of 50% or more from the lowest value, and three consecutive increases in PSA measured at least one week apart) or radiographic progression, as described previously [12, 17, 18]. For the analysis of progression-free survival (PFS), cancer-specific survival (CSS), and overall survival (OS), progression or death from any cause, death from prostate cancer, and death from any cause were defined as events, respectively. Patients who did not experience any of these events were censored at the last follow-up visit. For the survival analysis, the number of days from enrollment to the earliest event or censoring date was calculated. Patients who progressed to castration resistance at 2 years were defined as non-responders and patients who did not were defined as responders. Risk stratification by J-CAPRA risk score was performed as described previously [19].

### Genetic data

Genetic data were obtained as described previously [12]. Genomic DNA was genotyped using a Japonica Array v2 according to the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA, USA) [20–22]. This Axiom Array was customized for the Japanese genome by the Tohoku Medical Megabank Organization. Genotype calling was conducted using Genotyping Console software v4.2 (Thermo Fisher Scientific). We used the PSA-PFS at 2years-associated 2 and 46 SNPs with  $p < 1.0 \times 10^{-5}$  and  $p < 1.0 \times 10^{-4}$  that were identified in a previous study [12].

### Construction of simple prediction scores

The variables were classified as binary (1 or −1) or quantitative. Quantitative variables were normalized by subtracting the mean value and dividing by the standard deviation. Missing values were set to 0.

The prediction models were constructed using three ML algorithms (point-wise linear algorithm, logistic regression with elastic-net regularization algorithm, and eXtreme Gradient Boosting) and three datasets (clinical, clinical and 2 SNPs, and clinical and 46 SNPs). Details of the clinical dataset are provided in Supplementary Table 1. The point-wise linear (PWL) algorithm is a deep learning-based algorithm that was implemented using PyTorch 1.5.1 and Python 3.7.4 [23]. The PWL algorithm uses a deep (multi-layered) neural network structure that generates a logistic regression model for each sample; i.e., a weight vector tailored to each sample. The importance of each feature is computed using its weight vector. Deep unified networks were used to construct the deep neural network in which the network layers and neurons are connected in a mesh-like structure that reduces the risk of overfitting [24]. The logistic regression with elastic-net regularization (LR) algorithm and eXtreme Gradient Boosting (XGBoost) were used to build the baseline models (implemented using scikit-learn v0.24.2, xgboost 1.0.2, and Python 3.7.4) [25]. The best hyper-parameter of each model was determined by 5-fold cross validation using the discovery cohort [23]. The prediction performance of each model was calculated by area under the curve (AUC) and evaluated using the validation cohort and the models fitted by the best hyper-parameter.

The important features to predict castration resistance were determined based on an importance score that was calculated using the weight vector of the PWL algorithm. The sample-wise importance score was calculated as described previously [26]. Importance score was defined by the rate at which a sample was ranked in the top 10% of features with sample-wise importance scores. Parameters with importance scores  $\geq 0.1$  were extracted as important features. An importance score of 0.1 indicated that at least 10% of the samples had parameters that were in the top 10% of important features. Simple prediction scores were constructed using important features and the sign of the median of sample-wise weights in those features. We used the original values for the variables in the simple prediction scores. The prediction performance of the simple prediction scores was evaluated by AUC using discovery and validation cohorts.

### Estimation of effect by genetic background among different ethnic populations

Allele frequency data were obtained from the 1000 Genomes Project (<https://www.internationalgenome.org/home>). Estimated effect was the sum of the value for each SNP calculated as: coefficient  $\times 2 \times$  (minor allele frequency)  $\times (1 - \text{minor allele frequency}) + 2 \times$  coefficient  $\times (\text{minor allele frequency})^2$ .

### Statistical analyses

Statistical analyses were performed using JMP16 software (SAS Institute, Cary, NC, USA). Continuous and categorical data are presented as median with interquartile range and number with percentage, respectively. The association among the categorical data was analyzed by the chi-square test. Survival analysis was performed using the Kaplan–Meier method and log-rank test. Harrell's C-index was calculated using Stata v18 (College Station, TX, USA) as described previously [7]. All P-values were two-sided, and P-values  $< 0.05$  were considered significant for all the analyses.

## RESULTS

### Patients assignment

A total of 119 patients were included in the study, and were divided randomly in a 7:3 ratio into discovery ( $n = 82$ ) and validation ( $n = 37$ ) cohorts. Clinical parameters of the patients in each cohort are provided in Supplementary Table 1. Several clinical parameters including Gleason score, extent of disease grade, PSA level, and hemoglobin value were different between non-responders and responders in the discovery cohort (Supplementary Table 1). In addition, history of cerebral infarction, total type I procollagen-N-propeptide (P1NP), white blood cell count, and neutrophil count were higher in non-responders compared with their levels in responders in the discovery cohort (Supplementary Table 1).

### Predictive ability of castration resistance by three ML algorithms using genetic and clinical data

The ability of three ML algorithms (PWL, LR, XGBoost) to predict castration resistance using genetic and clinical data was evaluated. Using only the clinical data (Supplementary Table 1) to predict castration resistance at 2 years, the AUC, sensitivity, and specificity in the discovery cohort were 0.710–0.785, 0.568–0.704, and 0.644–0.689, respectively (Table 1). In the validation cohort, AUC, sensitivity, and specificity were 0.720–0.786, 0.688–0.875, and 0.579–0.684, respectively (Table 1). When the two SNPs associated with PSA-PFS at 2 years with  $p < 1.0 \times 10^{-5}$  were used together with the clinical parameters, AUC, sensitivity, and specificity in the discovery cohort slightly improved to 0.796–0.810, 0.700–0.754, and 0.689–0.778, respectively (Table 1). In the validation cohort, AUC, sensitivity, and specificity also slightly improved to 0.701–0.878, 0.625–0.750, and 0.684–0.789, respectively (Table 1). Finally, when the 46 SNPs associated with PSA-PFS at 2 years with  $p < 1.0 \times 10^{-4}$  were used together with the clinical parameters, AUC, sensitivity, and specificity in greatly improved in the discovery cohort to

**Table 1.** Predictive performance by machine learning methods using the indicated parameters.

Parameters	Mechine learning method	Discovery cohort <sup>a</sup>			Validation cohort		
		AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
Clinical data	PWL	0.773	0.639	0.911	0.786	0.812	0.684
	LR	0.710	0.568	0.644	0.727	0.688	0.684
	XGBoost	0.785	0.704	0.689	0.720	0.875	0.579
Clinical data and 2 SNPs ( $p < 1.0 \times 10^{-5}$ )	PWL	0.810	0.754	0.778	0.878	0.750	0.789
	LR	0.801	0.704	0.756	0.862	0.750	0.842
	XGBoost	0.796	0.700	0.689	0.701	0.625	0.684
Clinical data and 46 SNPs ( $p < 1.0 \times 10^{-4}$ )	PWL	0.988	0.864	0.978	1.000	0.938	1.000
	LR	0.963	0.725	0.978	1.000	0.938	1.000
	XGBoost	0.962	0.600	0.978	0.984	0.875	1.000

AUC area under the curve, PWL point-wise linear, LR logistic regression with elastic-net regularization, SNP single nucleotide polymorphism, XGBoost eXtreme Gradient Boosting.

<sup>a</sup>Mean values by 5-fold cross validation.

**Table 2.** Important features in clinical model.

#	Parameter	Score	Median weight
1	Hypertension	0.776	0.012
2	Total testosterone	0.524	0.015
3	Total cholesterol	0.495	0.015
4	Lymphocyte (%)	0.356	0.012
5	Gleason score	0.337	−0.011
6	ALB	0.178	0.008
7	BUN	0.176	−0.012
8	N-category	0.171	−0.007
9	Age	0.132	0.006
10	Dyslipidemia	0.117	−0.007
11	AST	0.112	0.008
12	PSA	0.107	−0.009

ALB albumin, AST aspartate aminotransferase, BUN blood urea nitrogen, PSA prostate-specific antigen.

0.962–0.988, 0.600–0.864, and 0.978–0.978, respectively (Table 1). In the validation cohort, AUC, sensitivity, and specificity also greatly improved to 0.984–1.000, 0.875–0.938, and 1.000–1.000, respectively (Table 1).

### Model creation to predict castration resistance using genetic and clinical data by ML

The PWL algorithm produced the highest AUC values in the three ML algorithms in the discovery and validation cohorts, with the exception of the clinical model in the discovery cohort (Table 1). Therefore, we created a prediction model for castration resistance using genetic and clinical data by the PWL algorithms.

When critical parameters associated with castration resistance at 2 years were used, 12 clinical parameters were identified (Table 2). Besides the known predictive factors, Gleason score, PSA level, N-category, albumin level, and total testosterone level, other factors including comorbidity with hypertension, total cholesterol level, lymphocyte ratio, blood urea nitrogen level, comorbidity with dyslipidemia, and aspartate aminotransferase level were also identified as critical parameters (Table 2). When the predictive model (clinical model) was created using the formula in Supplementary Table 2, the AUCs in the discovery and validation cohorts were 0.730 (95% CI, 0.610–0.849) and 0.585 (95% CI,

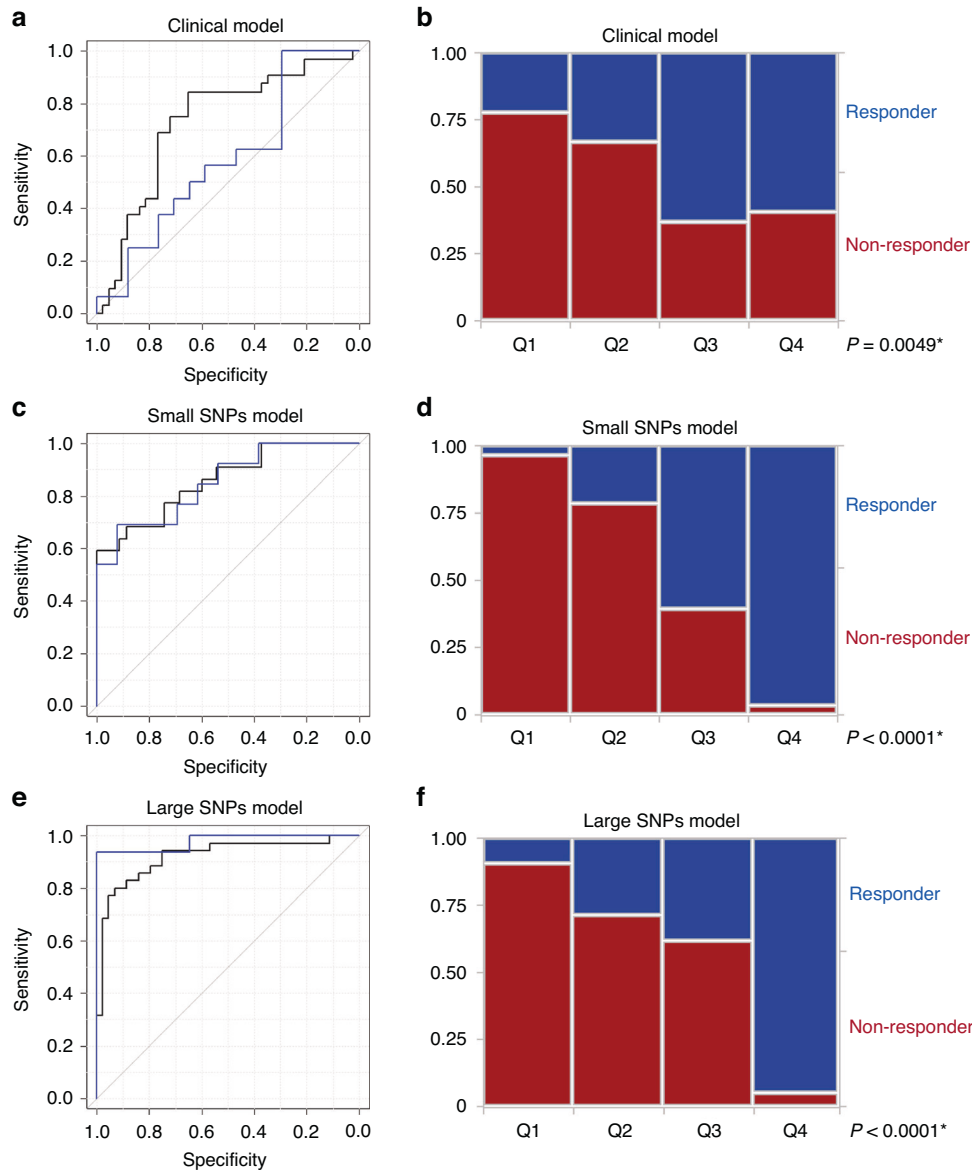
0.383–0.787), respectively (Fig. 1A). When prediction scores calculated by the clinical model were divided quarterly (Q1–Q4), the ranges were −1.42267 to 6.86566 in Q1, 6.901361–8.656931 in Q2, 8.7444–10.01411 in Q3, and 10.06535–13.23641 in Q4. The clinical model correctly predicted 21/27 (77.8%) in Q1 and 16/27 (59.3%) in Q4 to be non-responders and responders, respectively ( $p = 0.0049$ , Fig. 1B).

When the two SNPs ( $p < 1.0 \times 10^{-5}$ ) were added to the clinical parameters, six clinical parameters and two SNPs were identified to be critical to predict castration resistance at 2 years (Table 3). Besides the known predictive factors, Gleason score and extent of disease grade, other factors including comorbidity with diabetes mellitus, creatine kinase level, total P1NP level, and lymphocyte ratio were also identified as critical parameters (Table 3). When the predictive model (small SNPs model) was created using the formula in Supplementary Table 2, the AUCs in the discovery and validation cohorts improved to 0.857 (95% CI, 0.756–0.959) and 0.852 (95% CI, 0.706–0.998), respectively (Fig. 1C). When scores calculated by small SNPs model were divided quarterly (Q1–Q4), the ranges were −1.73 to 2.71 in Q1, 2.71–4.76 in Q2, 5.01–7.43 in Q3, and 7.6–13.24 in Q4. The small SNPs model discriminated responders and non-responders; 27/28 (96.4%) in Q1 and 27/28 (96.4%) in Q4 were correctly predicted to be non-responders and responders, respectively ( $p < 0.0001$ , Fig. 1D).

When the 46 SNPs ( $p < 1.0 \times 10^{-4}$ ) were added to the clinical parameters, 4 clinical parameters and 19 SNPs were identified to be critical to predict castration resistance at 2 years (Table 4). Besides the known predictive factors, M-category and Gleason score, other factors including total bilirubin level and glucose level were also identified as critical parameters (Table 4). When the predictive model (large SNPs model) was created using the formula in Supplementary Table 2, the AUCs in the discovery and validation cohorts were prominently improved to 0.920 (95% CI, 0.854–0.986) and 0.978 (95% CI, 0.932–1.000), respectively (Fig. 1E). When scores calculated by the large SNPs model were divided quarterly (Q1–Q4), the ranges were 0.8439–5.716 in Q1, 5.834–7.5162 in Q2, 7.7932–9.5307 in Q3, 9.6555–14.3819 in Q4. The large SNPs model discriminated responders and non-responders; 19/21 (90.5%) in Q1 and 19/20 (95.0%) in Q4 were correctly predicted to be non-responder and responder, respectively ( $p < 0.0001$ , Fig. 1F).

### Prognosis stratification by predictive models created by ML in advanced prostate cancer

We applied the three predictive models for prognosis stratification. PFS was significantly stratified by Q1–Q4 groups in all three



**Fig. 1 Predictive models of castration resistance at 2 years created by machine learning using clinical and genetic parameters.** Receiver operating characteristic curve for castration resistance at 2 years in clinical model (A), small SNPs model (C), and large SNPs model (E). Distribution of predictive scores of samples among discovery and validation cohorts according to quarterly-divided groups (Q1–Q4) based on scores by clinical model (B), small SNPs model (D), and large SNPs model (F).

**Table 3.** Important features in small SNPs model.

#	Parameter	Gene name	Score	Median weight
1	Diabetes mellitus		0.763	−0.026
2	Gleason score		0.659	−0.036
3	rs11231949	LOC101928443	0.561	−0.032
4	EOD grade		0.507	−0.025
5	rs2035081	PRIM1	0.432	−0.023
6	CK		0.224	0.032
7	Total P1NP		0.202	−0.029
8	Lymphocyte (%)		0.137	0.015

CK creatine kinase, EOD extent of disease, P1NP type I procollagen-N-propeptide, SNP single nucleotide polymorphism.

models (Fig. 2A). The PFS was more prominently stratified in the small SNPs and large SNPs models than it was in the clinical model (Fig. 2A). The C-indices for PFS in the clinical, small SNPs, and large SNPs models were 0.617 (95% CI, 0.556–0.678), 0.727 (95% CI, 0.681–0.774), and 0.730 (95% CI 0.667–0.793), respectively (Supplementary Table 3). The CSS was also significantly stratified by the Q1–Q4 groups in all three models, but was stratified more prominently in the large SNPs model than it was in the clinical and small SNPs models (Fig. 2B). The C-indices for CSS in the clinical, small SNPs, and large SNPs models were 0.678 (95% CI, 0.546–0.809), 0.670 (95% CI, 0.551–0.790), and 0.781 (95% CI 0.671–0.890), respectively (Supplementary Table 3). The OS was significantly stratified by the Q1–Q4 groups only in the large SNPs model (Fig. 2C). The C-indices for OS in the clinical, small SNPs, and large SNPs models were 0.636 (95% CI, 0.520–0.753), 0.621 (95% CI, 0.512–0.731), and 0.703 (95% CI 0.583–0.822), respectively (Supplementary Table 3).

**Table 4.** Important features in large SNPs model.

#	Parameter	Gene name	Score	Median weight
1	rs12979986	ZNF702P	0.988	−0.106
2	rs9625031	SRRD	0.968	−0.098
3	M-category		0.629	−0.038
4	rs1931229	VGLL2	0.512	−0.103
5	rs1660281	—	0.488	0.064
6	rs10064620	—	0.459	0.060
7	Gleason score		0.432	−0.066
8	rs10860210	RMST	0.395	−0.048
9	rs941207	BAZ2A	0.385	0.057
10	rs9868579	RPN1	0.317	−0.031
11	rs62174680	—	0.300	0.040
12	rs11232056	LOC101928443	0.259	0.040
13	rs8124833	C20orf78	0.251	0.037
14	rs74522810	—	0.234	−0.049
15	rs74369678	LRRN1	0.205	−0.036
16	NA (chr3:74213366)	—	0.190	0.104
17	rs2035081	PRIM1	0.188	−0.045
18	T-Bil		0.176	0.033
19	Glucose		0.149	−0.041
20	rs79404120	FAM19A5	0.146	0.058
21	rs11672661	COL5A3	0.129	−0.030
22	rs28625772	—	0.112	0.028
23	rs9298681	—	0.105	−0.036

NA not available, SNP single nucleotide polymorphism, T-Bil total bilirubin.

The preexisting well-known risk model, J-CAPRA risk group using TNM category, Gleason score, and PSA level stratified PFS, but not CSS and OS (Supplementary Fig. 1). The C-indices for PFS, CSS, and OS by J-CAPRA risk group were 0.588 (95% CI, 0.536–0.639), 0.602 (95% CI, 0.512–0.692), and 0.528 (95% CI 0.429–0.627), respectively (Supplementary Table 3).

#### Allele frequency by ethnicity and estimated effect of important SNPs in the large SNPs model on the response to ADT

The allele frequency of SNPs is known to differ among ethnic populations, which may affect the impact of genotype on outcomes. We investigated the allele frequencies of important SNPs in the large SNPs model. Minor allele frequency data for 16 SNPs were available in the 1000 Genomes Project, and differed among different ethnic groups as shown in Table 5. We estimated the effect of critical SNPs in the large SNPs model on the response to ADT. The estimated effect of the 16 SNPs was 0.94 in East Asians and −1.20 in Europeans, where a high value indicates higher probability of responder (Table 5).

#### DISCUSSION

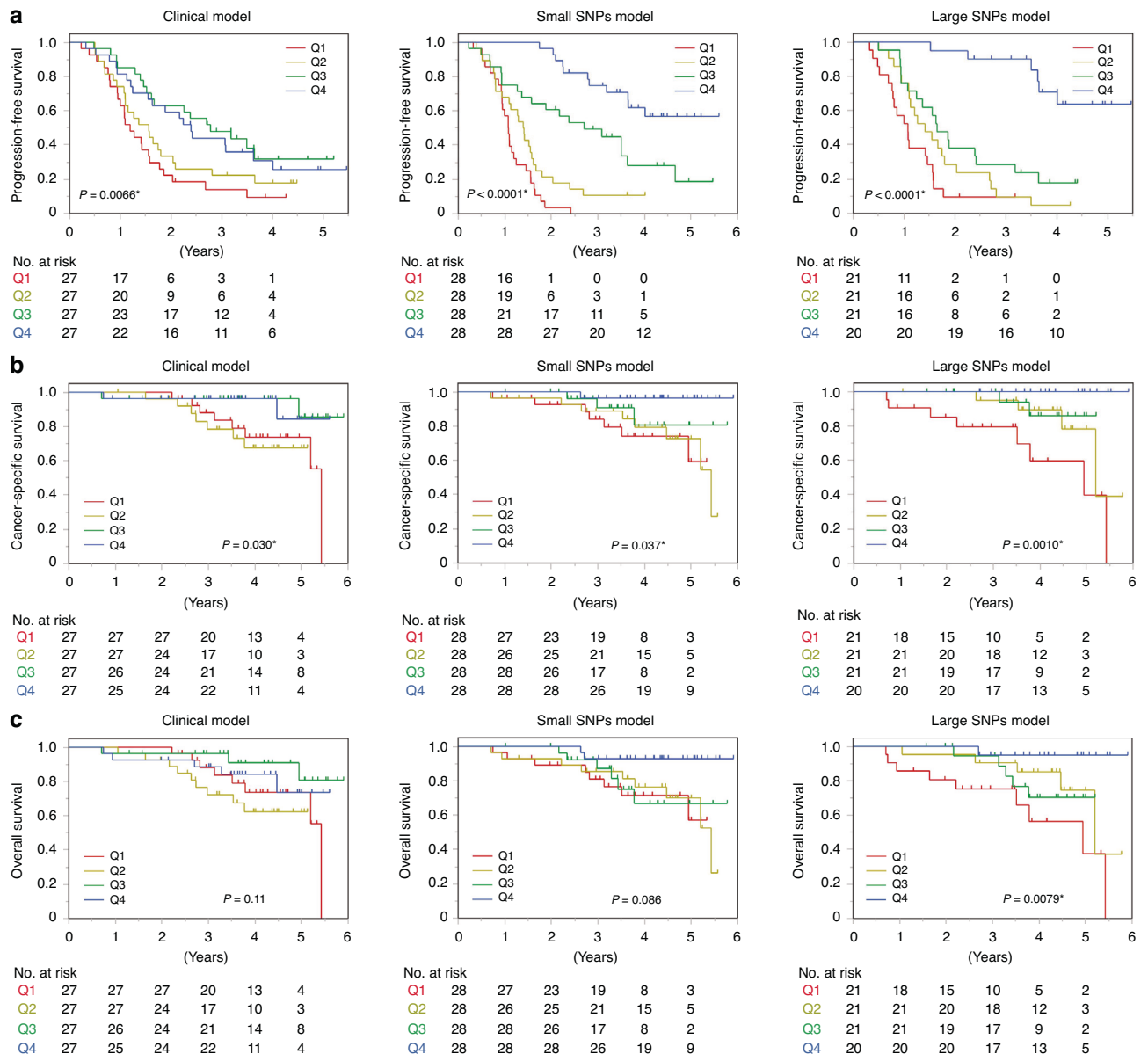
The results also showed higher predictive ability when SNPs were used in addition to clinical parameters. Several risk models using clinical parameters have been developed to predict the response and prognosis of ADT. However, their predictive power is modest, as indicated by AUCs of <0.7 [5–7]. We also found that the clinical model had limited predictive power even when created by ML, although the C-index was modestly higher than that of previous risk models. The predictive power of the

models created by ML was improved by adding small and large numbers of SNPs to the clinical parameters. In particular, the large SNPs model achieved C-indices >0.70 for PFS, CSS, and OS. Considering various previous predictive models failed to achieve C-indices >0.70, achieving higher prediction power of castration resistance at 2 years and prognosis by measuring 19 SNPs in addition to four clinical parameters would be valuable. Currently, intensive treatments have emerged as novel standard treatments for advanced prostate cancer [2]. Therefore, the large SNPs model will be helpful in choosing the best treatment for individual patients. Intensive treatment may be preferable if patient is a non-responder while de-escalated treatment may be preferable if patient is predicted to be a responder.

In addition, genetic parameters in the large SNPs model supported the ethnic differences of response to ADT. Several studies reported that Asians have a higher survival rate after primary ADT compared with that of Caucasians and African Americans [8, 9, 27]. Consistently, the score from allele frequencies of 16 SNPs in the large SNPs model indicated a higher possibility of responders in East Asians compared with that in people with European ancestry.

We identified various clinical and genetic parameters that were associated with response to ADT. Among 19 SNPs in the large SNPs model, rs1931229 was associated with the expression of *TSPYL1*, which is known to be a CYP17A1 and CYP3A4 regulator, by expression quantitative trait loci (eQTL) analysis, whereas rs941207 and rs2035081 were associated with *HSD17B6* expression (data not shown). Because *TSPYL1* and *HSD17B6* are both involved in androgen synthesis, they are associated with the response to ADT through their role in this pathway [28, 29].





**Fig. 2 Prognosis stratification by predictive models of castration resistance at 2 years created by machine learning using clinical and genetic parameters.** Progression-free survival (A), cancer-specific survival (B), and overall survival (C) among discovery and validation cohorts according to quarterly-divided groups (Q1–Q4) based on scores by clinical model (left), small SNPs model (middle), and large SNPs model (right).

Preexistence of hypertension was associated with favorable response to ADT, whereas preexistence of diabetes mellitus and high glucose level were associated with unfavorable response to ADT. This is consistent with our previous finding that comorbidity with hypertension and diabetes mellitus were associated with longer and shorter survival in primary ADT, respectively [30, 31]. Higher total cholesterol level was also associated with better response to ADT, whereas preexistence of dyslipidemia was associated with poor response to ADT, implying a close relationship between ADT and lipid metabolism.

This study had several limitations. The sample size was relatively small. Although the SNPs models had excellent predictive performance in the Japanese population, future work is needed to explore the generalizability of the predictive performance of these SNPs models in other populations. Primary ADT alone is no

longer a standard therapy, and utilized in a combination with other treatments such as androgen receptor signaling inhibitors. Conversely, a strong point of this study is that the clinical data were obtained from patients enrolled in a prospective trial, in which the treatment and testing schedules were subject to strict protocols.

## CONCLUSION

Our results demonstrate that the SNPs models using clinical and genetic parameters created by a PWL algorithm produced excellent prediction of castration resistance and prognosis in primary ADT for advanced prostate cancer. These models are expected to be helpful in treatment choice for advanced prostate cancer.

**Table 5.** Minor allele frequency by ethnics and estimated effect on the response to androgen deprivation therapy.

SNP identification	Coefficient	1000 Genome				
		SAS	AFR	EUR	AMR	EAS
rs12979986	−1	0.26	0.08	0.34	0.34	0.21
rs9625031	−1	0.14	0.02	0.29	0.19	0.23
rs1931229	−1	0.35	0.10	0.28	0.33	0.50
rs1660281	1	0.26	0.49	0.11	0.21	0.38
rs10064620	1	0.41	0.26	0.37	0.32	0.26
rs10860210	−1	0.45	0.50	0.38	0.46	0.32
rs941207	1	0.16	0.08	0.26	0.25	0.26
rs9868579	−1	NA	NA	NA	NA	NA
rs62174680	1	0.29	0.03	0.32	0.28	0.34
rs11232056	1	0.26	0.28	0.17	0.41	0.47
rs8124833	1	0.10	0.20	0.13	0.25	0.39
rs74522810	−1	0.20	0.18	0.21	0.18	0.18
rs74369678	−1	NA	NA	NA	NA	NA
NA (chr3:74213366)	1	NA	NA	NA	NA	NA
rs2035081	−1	0.22	0.33	0.38	0.33	0.26
rs79404120	1	0.17	0.01	0.14	0.11	0.16
rs11672661	−1	0.40	0.06	0.48	0.48	0.43
rs28625772	1	0.05	0.30	0.06	0.07	0.08
rs9298681	1	0.24	0.47	0.20	0.45	0.26
Estimated effect		−0.16	1.70	−1.20	0.08	0.94

AFR Africa, AMR native Americans, EAS East Asia, EUR Europe, NA not available, SAS South Asia, SNP single nucleotide polymorphism.

## DATA AVAILABILITY

The data sets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## REFERENCES

- Shiota M, Eto M. Current status of primary pharmacotherapy and future perspectives toward upfront therapy for metastatic hormone-sensitive prostate cancer. *Int J Urol*. 2016;23:360–9.
- Blas L, Shiota M, Eto M. Current status and future perspective on the management of metastatic castration-sensitive prostate cancer. *Cancer Treat Res Commun*. 2022;32:100606.
- Fizazi K, Foulon S, Carles J, Roubaud G, McDermott R, Fléchon A, et al. Abiraterone plus prednisone added to androgen deprivation therapy and docetaxel in de novo metastatic castration-sensitive prostate cancer (PEACE-1): a multicentre, open-label, randomised, phase 3 study with a 2×2 factorial design. *Lancet*. 2022;399:1695–707.
- Smith MR, Hussain M, Saad F, Fizazi K, Sternberg CN, Crawford ED, et al. Darolutamide and survival in metastatic, hormone-sensitive prostate cancer. *N Engl J Med*. 2022;386:1132–42.
- Akamatsu S, Kubota M, Uozumi R, Narita S, Takahashi M, Mitsuzuka K, et al. Development and validation of a novel prognostic model for predicting overall survival in treatment-naïve castration-sensitive metastatic prostate cancer. *Eur Urol Oncol*. 2019;2:320–8.
- Zelic R, Garmo H, Zugna D, Stattin P, Richiardi L, Akre O, et al. Predicting prostate cancer death with different pretreatment risk stratification tools: a head-to-head comparison in a nationwide cohort study. *Eur Urol*. 2020;77:180–8.
- Shiota M, Terada N, Kitamura H, Kojima T, Saito T, Yokomizo A, et al. Novel metastatic burden-stratified risk model in de novo metastatic hormone-sensitive prostate cancer. *Cancer Sci*. 2021;112:3616–26.
- Fukagai T, Namiki TS, Carlisle RG, Yoshida H, Namiki M. Comparison of the clinical outcome after hormonal therapy for prostate cancer between Japanese and Caucasian men. *BJU Int*. 2006;97:1190–3.
- Cooperberg MR, Hinotsu S, Namiki M, Carroll PR, Akaza H. Trans-Pacific variation in outcomes for men treated with primary androgen-deprivation therapy (ADT) for prostate cancer. *BJU Int*. 2016;117:102–9.
- Hemminki K, Ji J, Försti A, Sundquist J, Lennner P. Concordance of survival in family members with prostate cancer. *J Clin Oncol*. 2008;26:1705–9.
- Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20:467–84.
- Shiota M, Tatarano S, Kamoto T, Matsuyama H, Sakai H, Igawa T, et al. Genome-wide association studies in advanced prostate cancer: KYUCOG-1401-A study. *Endocr Relat Cancer*. 2023;30:e230044.
- Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15:233–4.
- Sigala RE, Lagou V, Shmeliov A, Atito S, Kouchaki S, Awais M, et al. Machine learning to advance human genome-wide association studies. *Genes*. 2023;15:34.
- Koido M. Polygenic modelling and machine learning approaches in pharmacogenomics: Importance in downstream analysis of genome-wide association study data. *Br J Clin Pharmacol*. 2023. <https://doi.org/10.1111/bcp.15913>.
- Yokomizo A, Shiota M, Morokuma F, Eto M, Matsuyama H, Matsumoto H, et al. GnRH antagonist monotherapy versus a GnRH agonist plus bicalutamide for advanced hormone-sensitive prostate cancer; KYUCOG-1401. *Int J Urol*. 2024;31:362–9.
- Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, et al. EAU guidelines on prostate cancer. Part II: Treatment of advanced, relapsing, and castration-resistant prostate cancer. *Eur Urol*. 2014;65:467–79.
- Scher HI, Halabi S, Tannock I, Morris M, Sternberg CN, Carducci MA, et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. *J Clin Oncol*. 2008;26:1148–59.
- Cooperberg MR, Hinotsu S, Namiki M, Ito K, Broering J, Carroll PR, et al. Risk assessment among prostate cancer patients receiving primary androgen deprivation therapy. *J Clin Oncol*. 2009;27:4306–13.
- Kawai Y, Mimori T, Kojima K, Nariiai N, Danjoh I, Saito R, et al. Japonica array: improved genotype imputation by designing a population-specific SNP array with 1070 Japanese individuals. *J Hum Genet*. 2015;60:581–7.
- Shiota M, Fujimoto N, Yamamoto Y, Takeuchi A, Tatsugami K, Uchiumi T, et al. Genome-wide association study of genetic variations associated with treatment failure after intravesical bacillus Calmette-Guérin therapy for non-muscle invasive bladder cancer. *Cancer Immunol Immunother*. 2020;69:1155–63.
- Shiota M, Miyake H, Takahashi M, Oya M, Tsuchiya N, Masumori N, et al. Effect of genetic polymorphisms on outcomes following nivolumab for advanced renal cell carcinoma in the SNIp-RCC trial. *Cancer Immunol Immunother*. 2023;72:1903–15.

23. Kumagai S, Togashi Y, Kamada T, Sugiyama E, Nishinakamura H, Takeuchi Y, et al. The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nat Immunol*. 2020;21:1346–58.
24. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak*. 2018;18:44.
25. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016. <https://doi.org/10.48550/arXiv.1603.02754>
26. Shibahara T, Wada C, Yamashita Y, Fujita K, Sato M, Kuwata J, et al. Deep learning generates custom-made logistic regression models for explaining how breast cancer subtypes are classified. *PLoS One*. 2023;18:e0286072.
27. Fujimoto N, Shiota M, Tomisaki I, Minato A. Gene polymorphism-related individual and interracial differences in the outcomes of androgen deprivation therapy for prostate cancer. *Clin Genitourin Cancer*. 2017;15:337–42.
28. Qin S, Liu D, Kohli M, Wang L, Vedell PT, Hillman DW, et al. TSPYL family regulates CYP17A1 and CYP3A4 expression: potential mechanism contributing to abiraterone response in metastatic castration-resistant prostate cancer. *Clin Pharmacol Ther*. 2018;104:201–10.
29. Ishizaki F, Nishiyama T, Kawasaki T, Miyashiro Y, Hara N, Takizawa I, et al. Androgen deprivation promotes intratumoral synthesis of dihydrotestosterone from androgen metabolites in prostate cancer. *Sci Rep*. 2013;3:1528.
30. Shiota M, Fujimoto N, Imada K, Kashiwagi E, Takeuchi A, Inokuchi J, et al. Prognostic impact of genetic polymorphism in mineralocorticoid receptor and comorbidity with hypertension in androgen-deprivation therapy. *Front Oncol*. 2018;8:635.
31. Hirata Y, Shiota M, Kobayashi T, Kashiwagi E, Takeuchi A, Inokuchi J, et al. Prognostic significance of diabetes mellitus and dyslipidemia in men receiving androgen-deprivation therapy for metastatic prostate cancer. *Prostate Int*. 2019;7:166–70.

## ACKNOWLEDGEMENTS

We thank Dr. Kaori Yasuda and Dr. Atsushi Doi (Cell Innovator, Fukuoka, Japan) for excellent guidance regarding the GWAS analysis, the Clinical Research Support Center Kyushu (CRS Kyushu, Fukuoka, Japan) for their secretarial and supportive assistance, and Ms. Noriko Hakoda and Ms. Eriko Gunshima for technical assistance (Department of Urology, Kyushu University, Fukuoka, Japan). We thank the patients who volunteered to participate in this trial, as well as the following KYUCOG-1401-A investigators and trial staff: Noriaki Tokuda in the Saga-ken Medical Centre Koseikan (Saga), Kentaro Kuroiwa in the Miyazaki Prefectural Miyazaki Hospital (Miyazaki), Kazuya Kawahara in the Kawahara Nephro-Urology Clinic (Aira), Motonobu Nakamura in the National Hospital Organization Kyushu Cancer Center (Fukuoka), Ken Goto in the Japanese Red Cross Fukuoka Hospital (Fukuoka), Sumitaka Mitsu in the Kagoshima Prefectural Oshima (Amami), Masafumi Nagano in the Fujimoto General Hospital (Miyakonojo), Naotaka Sakamoto in the National Hospital Organization Kyushu Medical Center (Fukuoka), Youji Doman in the Saiseikai Sendai Hospital (Satsumasendai), Hironobu Wakeda in the Chiyoda Hospital (Hyuga), Yasufumi Nabekura in the Kumamoto Urological Hospital (Kumamoto), Kohei Mizuma in the National Hospital Organization Ibusuki Medical Center (Ibusuki), and Hisato Inatomi in the Munakata Suikokai General Hospital (Fukutsu). We thank Margaret Biswas, PhD, from Edanz (<https://www.jp.edanz.com/ac>) for editing a draft of this manuscript.

## AUTHOR CONTRIBUTIONS

M Shiota: Conceptualization, funding acquisition, resources, formal analysis, visualization, writing—original draft. S Nemoto: formal analysis, visualization, writing—original draft. R Ikegami: formal analysis, visualization, writing—original draft. S Tatarano: Resources, writing—review and editing. T Kamoto: Resources, writing—review and editing. K Kobayashi: Resources, writing—review and editing. H Sakai: Resources, writing—review and editing. T Igawa: Resources, writing—review and editing. T Kamba: Resources, writing—review and editing. N Fujimoto: Resources, writing—review and editing. A Yokomizo: Resources, writing—review and editing. S

Naito: Resources, writing—review and editing. M Eto: Resources, supervision, writing—review and editing.

## FUNDING

This work was supported by Astellas Investigator Sponsored Research to MS, which had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, or decision to submit results. KYUCOG-1401 (UMIN000014243, jRCTs071180035) was supported by Astellas Pharma.

## COMPETING INTERESTS

MS received honoraria from Janssen Pharmaceutical, AstraZeneca, Astellas Pharma, Sanofi, and Bayer and research funding support from Daiichi Sankyo. SN and RI are employees of Hitachi Co. Ltd. Toshiyuki Kamoto received research funding support from Janssen Pharmaceutical, Astellas Pharma, Shin Nippon Biomedical Laboratories, Ono Pharmaceutical, Bayer, Sanofi, and Takeda Pharmaceutical. HS received honoraria from Takeda Pharmaceutical and Astellas Pharma. TI received honoraria from Janssen Pharmaceutical and Astellas Pharma. TK received honoraria from AstraZeneca and Merck. AY received honoraria from Janssen Pharmaceutical and Astellas Pharma. ME received honoraria from Janssen Pharmaceutical, Janssen Pharmaceutical, AstraZeneca, and Astellas Pharma and research funding support from Astellas Pharma, Sanofi, and Takeda Pharmaceutical. MS is an Associate Editor at BJC Reports. He was not involved in any aspect of handling of this manuscript or any editorial decisions.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was conducted in accordance with the Declaration of Helsinki and the Japanese Ethical Guidelines for Medical and Health Research Involving Human Subjects. Eligible patients provided written informed consent. This study was approved by the Kyushu University review board (23087-00).

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44276-024-00093-3>.

**Correspondence** and requests for materials should be addressed to Masaki Shiota.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024