EBioMedicine 60 (2020) 103033

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Genetic characterization of pancreatic cancer patients and prediction of carrier status of germline pathogenic variants in cancer-predisposing genes

Keijiro Mizukami[a], Yusuke Iwasaki[a], Eiryo Kawakami[b,c], Makoto Hirata[d,e], Yoichiro Kamatani[f], Koichi Matsuda[f], Mikiko Endo[a], Kokichi Sugano[e,g], Teruhiko Yoshida[e], Yoshinori Murakami[d], Hidewaki Nakagawa[h], Amanda B. Spurdle[i], Yukihide Momozawa[a,*]

[a] Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
[b] Medical Sciences Innovation Hub Program, RIKEN, Yokohama, Kanagawa, Japan
[c] Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan
[d] Institute of Medical Science, The University of Tokyo, Tokyo, Japan
[e] Department of Genetic Medicine and Services, National Cancer Center Hospital, Tokyo Japan
[f] Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
[g] Division of Cancer Prevention and Genetic Counseling, Genome Center, Tochigi Cancer Center, Japan
[h] Laboratory for Cancer Genomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
[i] Division of Genetics and Population Health, QIMR Berghofer Medical Research Institute, Brisbane, Australia

ARTICLE INFO

Article History:
Received 30 June 2020
Revised 11 September 2020
Accepted 11 September 2020
Available online xxx

Keywords:
Pancreatic cancer
Pathogenic variants
Universal screening for patients
Machine learning
BRCA
ATM

ABSTRACT

Background: National Comprehensive Cancer Network (NCCN) recently recommended germline genetic testing for all pancreatic cancer patients. However, the genes targeted by genetic testing and the feasibility of selecting patients likely to carry pathogenic variants have not been sufficiently verified. The purpose of this study was to genetically characterize Japanese patients and examine whether the current guideline is applicable in this population.

Methods: Using targeted sequencing, we analyzed the coding regions of 27 cancer-predisposing genes in 1,005 pancreatic cancer patients and 23,705 controls in Japan. We compared the pathogenic variant frequency between cases and controls and documented the demographic and clinical characteristics of carrier patients. We then examined if it was possible to use machine learning to predict carrier status based on those characteristics.

Findings: We identified 205 pathogenic variants across the 27 genes. Pathogenic variants in *BRCA2, ATM*, and *BRCA1* were significantly associated with pancreatic cancer. Characteristics associated with carrier status were inconsistent with previous investigations. Machine learning classifiers had a low performance in determining the carrier status of pancreatic cancer patients, while the same classifiers, when applied to breast cancer data as a positive control, had a higher performance that was comparable to that of the NCCN guideline.

Interpretation: Our findings support the clinical significance of multigene panel testing for pancreatic cancer and indicate that at least 3.4% of Japanese patients may respond to poly (ADP ribose) polymerase inhibitor treatments. The difficulty in predicting carrier status suggests that offering germline genetic testing for all pancreatic cancer patients is reasonable.

Funding: AMED under Grant Number JP19kk0305010 and Australian National Health and Medical Research funding (ID177524)

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Pancreatic cancer is devastating because of its high mortality rate, which has increased along with the incidence of the cancer in almost all countries and territories [1]. According to Japanese cancer statistics [2], the five-year survival rate of pancreatic cancer is the lowest among those of 19 cancers in the monitoring data of cancer incidence in Japan (cancers of the skin, larynx, bladder, uterine corpus, prostate, breast, liver, renal/urinary tract, stomach, uterine cervix, colon, rectal, thyroid, oral/pharynx, ovary, lung, esophagus, gallbladder/bile duct, and pancreas). The five-year survival rate for the localized stage of

* Corresponding author.
E-mail address: momozawa@riken.jp (Y. Momozawa).

https://doi.org/10.1016/j.ebiom.2020.103033
2352-3964/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Research in context

Evidence before this study

The current National Comprehensive Cancer Network (NCCN) guideline recommends that all patients with pancreatic cancer should undergo germline genetic testing for 11 cancer-predisposing genes including *BRCA1/2* and Lynch syndrome genes. Despite the importance of reviewing a sufficient number of investigations to improve the precision and detail of such guidelines, there have been a few large-scale case-control association studies using a multigene panel in pancreatic cancer. Additionally, most investigations that examined several hundred patients almost only targeted non-Hispanic European subjects, despite interethnic differences in susceptibility to pancreatic cancer. Thus, it is unclear whether the current NCCN guideline recommendation based on findings from the European population is reproducible and applicable to other populations.

Added value of this study

Using amplicon target sequencing, germline pathogenic variants in the coding region of 27 cancer-predisposing genes were found in 6.7% of 1005 Japanese patients with pancreatic cancer. By comparing with 23,705 Japanese controls, *BRCA2* ($P = 2.723 \times 10^{-18}$ [two-sided Fisher's exact test], OR = 14.7, 95% CI = 8.5–24.9), *ATM* ($P = 3.168 \times 10^{-11}$ [two-sided Fisher's exact test], OR = 10.7, 95% CI = 5.7–19.5), and *BRCA1* ($P = 3.333 \times 10^{-7}$ [two-sided Fisher's exact test], OR = 13.4, 95% CI = 5.2–32.2) were identified as statistically significant causative genes. When analyzing the association between pathogenic variants in the three genes and demographic and clinical characteristics, carriers were more likely to have a family history of gastric ($P = 0.028$ [two-sided Fisher's exact test], OR = 2.0, 95% CI = 1.0 to 3.7) and ovarian ($P = 0.045$ [two-sided Fisher's exact test], OR = 7.7, 95% CI = 0.7–48.6) cancers and severe lymphovascular invasion ($P = 0.018$ [Cochran-Armitage test]). Machine learning using six classifiers did not exhibit practical predictive performances for identifying patients with pathogenic variants in pancreatic cancer (AUC: 0.514–0.561).

Implications of all the available evidence

BRCA1/2 and *ATM* are important genes for pancreatic cancer not only in the European population but also in the Japanese population. The proportion of patients with pathogenic variants in the Japanese population (5.1%) was almost the same as that in the European population (4.8%). However, predicting a carrier of a pathogenic variant based on clinical and demographic characteristics was found to be difficult, even when using machine learning classifiers. Thus, offering germline genetic testing for all pancreatic cancer patients would be a reasonable strategy.

pancreatic cancer (6.1% of patients are diagnosed at this stage) is 38.6%, that of the regional stage (32.3%) is 10.2%, and that of the distant stage (42.8%) is 1.3%. The prevalence of pancreatic cancer per 100,000 individuals in Japan is 6.4 [3], and the proportion of familial pancreatic cancer in Japan is 7.5% of all pancreatic cancer patients [4], both of which are comparable to those in Western counties. Germline pathogenic variants in cancer-predisposing genes (e.g. *BRCA1/2, ATM, CDKN2A, MSH6, PALB2, TP53, APC, BARD1, BRIP, CHEK2, NBN, PMS2, RAD51, and RAD51C*) have been identified in nearly 10% of patients with pancreatic cancer [5-7]. Identifying a germline variant in a susceptibility gene not only would assist with early detection of at-risk individuals in the general population who would benefit from longitudinal surveillance programs [8], but also would be beneficial for selecting poly (ADP ribose) polymerase (PARP) inhibitor treatments for *BRCA1/2* genes [9] and for developing cancer prevention strategies for relatives of cancer patients [10]. However, several studies pointed out the difficulty of selecting patients with a pathogenic variant because an effective predictor for identifying those patients has not been identified [5,7,11,12].

The current National Comprehensive Cancer Network (NCCN) guideline (Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, version 1.2020) recommends that all patients with pancreatic cancer undergo germline genetic testing for *BRCA1/2, ATM, CDKN2A*, Lynch syndrome genes (*MLH1, MSH2, MSH6, EPCAM*), *PALB2, STK11, TP53* [13]. Despite the importance of reviewing a sufficient number of investigations to improve the precision and detail of such guidelines, there has been only two large-scale case-control association studies, conducted by the same research group, using a multigene panel that examined over 1000 patients with pancreatic cancer [7,14]. Additionally, most investigations that examined several hundred patients almost only targeted non-Hispanic European subjects [5,7,11,12], despite interethnic differences in susceptibility to pancreatic cancer [15]. We have previously identified differences in genetic characteristics of Japanese breast and prostate cancer patients compared to European patients [16,17]. Thus, it is unclear whether the current NCCN guideline recommendation based on findings from the European population is reproducible and applicable to other populations.

To examine the applicability of the current guideline to pancreatic cancer germline genetic testing in Japan, we first performed a genetic characterization of Japanese patients with pancreatic cancer based on a large case-control sequencing association study, and then documented the demographic and clinical characteristics of patients with pathogenic variants. We also examined whether pathogenic variant carrier status is predictable based on demographic and clinical characteristics using machine learning techniques.

2. Materials and methods

2.1. Study design and participants

We obtained all samples from BioBank Japan [18,19], which is a multi-institutional, hospital-based registry that collected DNA and clinical information from patients with various common diseases, including pancreatic cancer [19], throughout Japan between 2003 and 2018. Clinical characteristics of cancer patients and control individuals were collected with interviews or medical record survey using a standard questionnaire at the point of entry to BioBank Japan. In this study, we analyzed the DNA of all 1009 pancreatic cancer patients available when requested (March 2018). We used data from 23,780 controls, as in our previous study investigating germline pathogenic variants in colorectal cancer patients [16,17,20]. All participants provided written informed consent. The study was approved by the ethical committees of the Institute of Medical Sciences, the University of Tokyo, and the RIKEN Center for Integrative Medical Sciences (Approval number: 17−17−16(8)).

2.2. Sequencing and bioinformatics analysis

We analyzed 27 genes (*APC, ATM, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDK4, CDKN2A, CDH1, CHEK2, EPCAM, HOXB13, NBN, NF1, MLH1, MSH2, MSH6, MUTYH, PALB2, PMS2, PTEN, RAD51C, RAD51D, SMAD4, STK11*, and *TP53*), which are known cancer susceptibility genes and include the 11 genes recommended for germline testing for pancreatic cancer by the NCCN guideline but also other cancer susceptibility genes included in Myriad myRisk Hereditary Cancer Test [21]. We analyzed the complete coding regions and 2-bp

flanking intronic sequences of all 27 genes (84,822 bp), except exons 10 − 15 of *PMS2* due to the presence of *PMS2* pseudogene [22], using a multiplex polymerase chain reaction–based targeted sequence method [23]. We identified genetic variants, as described in our previous studies [16,17].

2.3. Annotation of variants

We determined the clinical significance (pathogenic, benign, or uncertain) of all variants as in our previous study [20], using the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines (Supplementary Table 1) [24,25]. We considered variants to be pathogenic if they were classified as such by the ACMG/AMP guidelines and/or by ClinVar interpretation irrespective of review status [26]. Variants not registered in ClinVar at May 20, 2019 were considered novel.

2.4. Predicting pathogenic variant carrier status using machine learning

To take advantage of 65 demographic and clinical characteristics collected from patients of BioBank Japan, we attempted to predict the pathogenic variant carrier status of patients using machine learning techniques. Machine learning performance was assessed using six types of supervised machine learning classifiers: random forest, gradient boosting machine, conditional random forest, naive bayes, neural network, and support vector machine. We also used a multiple logistic regression classifier as the baseline. All classifiers were implemented using the R package "caret" (ver. 6.0.84) [27]. Down-sampling was used to correct for imbalances between the number of cases and the number of control samples in training data. Classifiers were trained using repeated 10-fold cross-validation of the training dataset, and their predictive performances were evaluated based on the area under the ROC curve (AUC) using "pROC" (ver. 1.15.3) [28]. Hyperparameters tuned by grid-search (the number of levels for each tuning parameter, 3) are shown in Supplementary Table 2. Variables missing in over 50% of samples were excluded from analysis; 52 variables used in machine learning analysis are shown in Supplementary Table 3. Finally, 721 pancreatic cancer cases with complete data across 52 variables were applied to each of the six classifiers to predict the carrier status of pathogenic variants in genes associated with pancreatic cancer.

To confirm the predictive performance of the classifiers, we applied the same procedure to breast cancer data as a positive control, because breast cancer has practical criteria for predicting the carrier status of pathogenic variants [13]. For breast cancer patients, we analyzed germline pathogenic variant data and patient information across 65 characteristics. Sample collection, sequencing, and variant annotation of the data were conducted according to the same procedures as the present study [17]. The same machine learning classifiers and criteria for variable selection as those used to analyze the pancreatic cancer data were applied. Finally, 2496 breast cancer cases without missing data across 48 variables were used to predict the carrier status of pathogenic variants in four genes significantly associated with breast cancer (*BRCA1, BRCA2, TP53,* and *PALB2*) (Supplementary Table 3). To evaluate the effect of sample size on machine learning performance, we assessed the AUC of the random forest prediction using the same number of breast cancer samples as the number of pancreatic cancer samples. We randomly selected 721 breast cancer cases from the training and test cohorts. To reduce any potential bias due to random selection, we generated 100 sets of data for evaluation.

2.5. Statistical analysis

Case-control association analysis was performed using a two-sided Fisher's exact test with a recessive model for *MUTYH* and with dominant models for the other genes. Bonferroni correction was applied for the association analysis between each of 24 genes with pathogenic variants and pancreatic cancer ($P < 0.003 = 0.05/24$). Median survival time for carriers and non-carriers of pathogenic variants of a subset of genes (*BRCA1, BRCA2*, and *ATM*) was estimated using the Kaplan-Meier method, with a log-rank test for significance of the survival distribution using R package "survival" (ver. 2.44.1.1) [29]. A multivariable Cox regression model that adjusted for age at diagnosis and cancer stage was also used, along with a Wald test for the statistical significance of the relationship between carrier status and survival hazard. The patient population used for survival analysis was restricted to a subset of 351 pancreatic cancer cases that were enrolled in BioBank Japan within 3 months ($\leq$ 92 days) of diagnosis. In addition, to investigate the association between other demographic and clinical characteristics and pathogenic variants in any genes significantly associated with pancreatic cancer, we used t-tests for continuous variables and Fisher's exact tests or Cochran-Armitage tests for discrete variables. When a characteristic showed a significant association, we separately analyzed the association between the characteristic and pathogenic variants in each individual gene. These statistical tests were two-sided, and values of $P < 0.05$ were considered statistically significant. All analyses were performed using the R statistical package (ver. 3.5.1).

2.6. Role of the funding source

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the Article. KeM, YI, ME, and YuM had access to the raw data. The corresponding author had full access to all of the data and the final responsibility for the decision to submit for publication.

3. Results

3.1. Patient characteristics

We fully analyzed 1009 pancreatic cancer cases and 23,780 controls. The mean age at diagnosis of pancreatic cancer was 67.2 years (standard deviation = 9.9) (Table 1). The incidence of pancreatic cancer was higher in males (62.0% of cases) than in females (38.0% of cases, $P = 4.236 \times 10^{-9}$ [*t*-test]). There were significantly higher rates of positive smoking and alcohol history among cancer cases (59.8% and 55.0% of cases, respectively) than among controls (46.9% and 45.9% of cases, respectively; $P = 2.762 \times 10^{-16}$ and 5.045×10^{-9} [two-sided Fisher's exact test], respectively). Personal history of any other cancer was reported by 15.1% of patients. First- and second-degree relatives were included in the analysis for familial history. Family history of any cancer was reported by 57.0% of patients. Family history of pancreatic cancer was observed in 8.1% of cases. Other clinical characteristics of cases are shown in Supplementary Table 4.

3.2. Pathogenic germline variants

Sequencing analysis covered $\geq$ 20 sequence reads in 99.95% of the target region in 27 genes and identified 3610 genetic variants in 1005 patients and 23,705 controls. Among them, 65 variants had a minor allele frequency (MAF) $>$ 5%, 202 variants MAF $\leq$ 5% and $>$ 0.1%, and 3343 variants MAF $\leq$ 0.1% in controls. According to the ACMG/AMP guidelines, we categorized the variants as follows: 14 pathogenic, 147 likely pathogenic, 133 benign, 227 likely benign, and 3089 variants of uncertain significance (VUS) with insufficient evidence. Three variants designated as pathogenic using our application of the ACMG/AMP criteria had a MAF $>$ 0.05% in controls, $P > 0.05$ [two-sided Fisher's exact test], and OR $<$ 2, and were thus considered to represent at best low-risk variants and were removed from subsequent analyses (Supplementary Table 5): p.Arg223Cys in *CHEK2*

Table 1
Demographic and clinical characteristics of study participants.

Characteristic	PC patients No. (%)	Controls No. (%)
No. of participants	1009 (100)	23,780 (100)
Mean age at entry ± SD, (range), y	67.71 ± 9.77 (27 - 96)	71.11 ± 7.25 (60 - 105)*
Mean age at diagnosis ± SD, (range), y	67.17 ± 9.92 (24 - 96)	–
Sex†		
Male	626 (62.0)	12,520 (52.6)
Female	383 (38.0)	11,260 (47.4)
Smoking history†		
Yes	603 (59.8)	11,155 (46.9)
No	397 (39.3)	12,523 (52.7)
Missing	9 (0.9)	102 (0.4)
Alcohol history†		
Yes	555 (55.0)	10,907 (45.9)
No	443 (43.9)	12,731 (53.5)
Missing	11 (1.1)	142 (0.6)
BMI at diagnosis ± SD, (range), kg/m^2	20.85 ± 3.31 (12.2 - 37.9)	–
Personal history of cancer(s) other than PC		
Yes	152 (15.1)	0 (0)*
No	857 (84.9)	23,780 (100)*
Family history of cancer (s)		
Yes	575 (57.0)	0 (0)*
No	434 (43.0)	23,780 (100)*
Stage of cancer		
0	10 (1.0)	–
I	42 (4.2)	–
II	46 (4.6)	–
III	92 (9.1)	–
IVa	112 (11.1)	–
IVb	88 (8.7)	–
Missing	619 (61.3)	–

* Controls who were 60 years old or over and did not have personal nor family history of cancers were used.
† P-value: 4.236×10^{-9} in sex, 2.762×10^{-16} in smoking history, and 5.045×10^{-9} in alcohol history. PC, pancreatic cancer; BMI, body mass index; SD, standard deviation.

(present in 0.20% of cases and in 0.21% of controls, $P = 1.000$, OR = 1.0, 95% confidence interval [CI] = 0.1–3.7), p.Arg562* in *CHEK2* (present in 0% of cases and in 0.05% of controls, $P = 1.000$, OR = 0.0, 95% CI = 0.0–8.5), and p.Lys111fs in *RAD51D* (present in 0.10% of cases and in 0.14% of controls, $P = 1.000$, OR = 0.7, 95% CI = 0.0–4.4). Incorporating the ClinVar interpretation, we finally considered 205, 3, 787, and 2615 variants as (likely) pathogenic, low-risk, (likely) benign, and VUS, respectively. Single-variant association results are shown in Supplementary Table 5 for the 205 pathogenic and the 3 low-risk variants and in Supplementary Table 6 for 787 benign variants and the 2615 VUS.

We focused on the 205 pathogenic variants. We compared the frequency of pathogenic variant carriers in cancer cases with those in controls (Table 2). In total, 67 patients (6.7%) and 225 control participants (1.0%) carried a pathogenic variant in one of 27 genes ($P = 1.525 \times 10^{-31}$ [two-sided Fisher's exact test], OR = 7.5, 95% CI = 5.5–9.9). Six genes (*BRCA2, ATM, BRCA1, APC, PALB2*, and *CDKN2A*) had values of $P < 0.05$ [two-sided Fisher's exact test]. *BRCA2* ($P = 2.723 \times 10^{-18}$ [two-sided Fisher's exact test], OR = 14.7, 95% CI = 8.5–24.9), *ATM* ($P = 3.168 \times 10^{-11}$ [two-sided Fisher's exact test], OR = 10.7, 95% CI = 5.7–19.5), and *BRCA1* ($P = 3.333 \times 10^{-7}$ [two-sided Fisher's exact test], OR = 13.4, 95% CI = 5.2–32.2) were significantly associated with pancreatic cancer after Bonferroni correction. Although the direction of OR in *APC* (OR = 23.6, No. of carrier cases = 2), *PALB2* (OR = 7.9, No. of carrier case = 2), and *CDKN2A* (OR = ∞, No. of carrier case = 1) also indicated association, the number of patients with pathogenic variants in these genes was very limited. We checked the locations and frequencies of the pathogenic variants in the three most significantly associated genes (Fig. 1). Among pathogenic variants in *BRCA2, ATM*, and *BRCA1* in cases, 78.9% of them were singleton, and the most frequent pathogenic variant was c.4776+2T>*A* in *ATM* ($n = 4$, $P = 3.821 \times 10^{-5}$ [two-sided Fisher's exact test], OR = 47.3, 95% CI = 6.8–520.7), which were not found in 141,456 individuals in gnomAD 2.1.1 (https://gnomad.broadinstitute.org/about).

3.3. Associations between germline pathogenic variant carrier status and survival

The median overall survival time for patients with pathogenic variants in the three genes associated with pancreatic cancer (*BRCA2, ATM*, and *BRCA1*) was 0.97 years (95% CI, 0.48−1.60 years), whereas that for patients without pathogenic variants was 0.92 years (95% CI, 0.81−1.04 years). The association between the carrier status of pathogenic variants in those genes and median survival time was not statistically significant ($P = 0.609$ [log-rank test]) (Fig. 2). There was no difference by carrier status in survival hazard (adjusted by age at diagnosis and cancer stage) (hazard ratio: 0.85; 95% CI, 0.22−3.33; $P = 0.814$ [Wald test]).

3.4. Association of pathogenic variant carrier status with demographic and clinical characteristics

To investigate the association between pathogenic variants in the three genes of interest and demographic and clinical characteristics, we compared the characteristics between 51 carrier patients and 954 non-carrier patients. First, we analyzed the association of carrier status with diagnosis age because there were clear negative correlations between diagnosis age groups and the proportion of patients with pathogenic variants in breast [17], prostate [16], and colorectal [20] cancers. The mean age of carriers and non-carriers was 67.0 and 67.2 years old, respectively, and no significant association between carrier status and diagnosis age was observed ($p = 0.932$ [t-test]). There was no unidirectional trend of the carrier frequency in age groups (Supplementary Fig. 1). Next, we analyzed the association of carrier status with the other characteristics. The carriers more often had a family history of gastric or ovarian cancers (37.3% vs 23.2% in non-carriers, $P = 0.028$ [two-sided Fisher's exact test], and 3.9% vs 0.5% in non-carriers, $P = 0.045$ [two-sided Fisher's exact test], respectively) (Supplementary Table 7). Nine of 17 *ATM* pathogenic variant carrier patients (52.9%) had a family history of gastric cancer ($P = 0.008$ [two-sided Fisher's exact test], OR = 3.7, 95% CI = 1.3–11.2). Two of 9 *BRCA1* pathogenic variant carrier patients (22.2%) had a family history of ovarian cancer ($P = 0.002$ [two-sided Fisher's exact test], OR = 52.4, 95% CI = 4.3–396.7). Regarding the family history of cancer other than gastric and ovarian cancers and the personal history of cancers other than pancreatic cancer, there were no significant associations observed in this study. For example, the proportions of carriers and non-carriers with a family history of breast cancer were 11.8% and 6.1%, respectively ($P = 0.131$ [two-sided Fisher's exact test]). A family history of pancreatic cancer was identified in 9.8% of carriers and 8.1% of non-carriers ($P = 0.600$ [two-sided Fisher's exact test]). The proportions of carriers and non-carriers having at least one first degree relative with pancreatic cancer were 9.8% and 7.9%, respectively ($P = 0.593$ [two-sided Fisher's exact test]). Pathogenic variants in the three genes were identified in 5 of 80 patients with at least one first degree relative (6.3%). Carriers also exhibited severe lymphovascular invasion ($P = 0.018$ [Cochran-Armitage test]), and *ATM* and *BRCA2* were individually associated with severe lymphovascular invasion ($P = 0.009$ and $P = 0.048$ [Cochran-Armitage test], respectively) (Supplementary Table 8).

Table 2
Results of the gene-based association test using pathogenic variants.

Gene	No. of pathogenic variants	Case (n = 1005) No. of carriers (%)	Control (n = 23,705) No. of carriers (%)	P-value*	OR (95% CI)
BRCA2†	34	25 (2.49)	41 (0.17)	2.723×10^{-18}	14.7 (8.5 - 24.9)
ATM†	35	17 (1.69)	38 (0.16)	3.168×10^{-11}	10.7 (5.7 - 19.5)
BRCA1†	18	9 (0.90)	16 (0.07)	3.333×10^{-7}	13.4 (5.2 - 32.2)
APC	2	2 (0.20)	2 (0.01)	0.009	23.6 (1.7 - 324.7)
PALB2†	8	2 (0.20)	6 (0.03)	0.039	7.9 (0.8 - 44.1)
CDKN2A†	1	1 (0.10)	0 (0)	0.041	Inf (0.6 - Inf)
MSH6†	13	3 (0.30)	21 (0.09)	0.072	3.4 (0.6 - 11.3)
TP53†	9	2 (0.20)	12 (0.05)	0.109	3.9 (0.4 - 17.7)
PTEN	4	1 (0.10)	3 (0.01)	0.153	7.9 (0.1 - 97.8)
EPCAM†	5	1 (0.10)	4 (0.02)	0.187	5.9 (0.1 - 59.7)
BARD1	8	1 (0.10)	8 (0.03)	0.312	3.0 (0.1 - 22)
BRIP1	8	1 (0.10)	8 (0.03)	0.312	3.0 (0.1 - 22)
MLH1†	5	1 (0.10)	8 (0.03)	0.312	3.0 (0.1 - 22)
CHEK2	8	1 (0.10)	9 (0.04)	0.340	2.6 (0.1 - 19)
MUTYH	9	0 (0)	0 (0)	1	0 (0 - Inf)
NBN	7	0 (0)	8 (0.03)	1	0 (0 - 13.8)
NF1	7	0 (0)	7 (0.03)	1	0 (0 - 16.4)
RAD51C	6	0 (0)	7 (0.03)	1	0 (0 - 16.4)
PMS2	5	0 (0)	10 (0.04)	1	0 (0 - 10.5)
MSH2†	4	0 (0)	4 (0.02)	1	0 (0 - 35.8)
RAD51D	3	0 (0)	6 (0.03)	1	0 (0 - 20)
HOXB13	2	0 (0)	2 (0.01)	1	0 (0 - 125.7)
STK11†	2	0 (0)	2 (0.01)	1	0 (0 - 125.7)
CDH1	1	0 (0)	1 (<0.01)	1	0 (0 - 906.9)
SMAD4	1	0 (0)	2 (0.01)	1	0 (0 - 125.7)
BMPR1A	0	0 (0)	0 (0)	1	0 (0 - Inf)
CDK4	0	0 (0)	0 (0)	1	0 (0 - Inf)
Sum	205	67 (6.67)	225 (0.95)	1.525×10^{-31}	7.5 (5.5 - 9.9)

* Fisher's exact test under a recessive model for *MUTYH* and dominant model for the other genes.
† Genes described in the NCCN guideline. OR, odds ratio; CI, confidence interval.

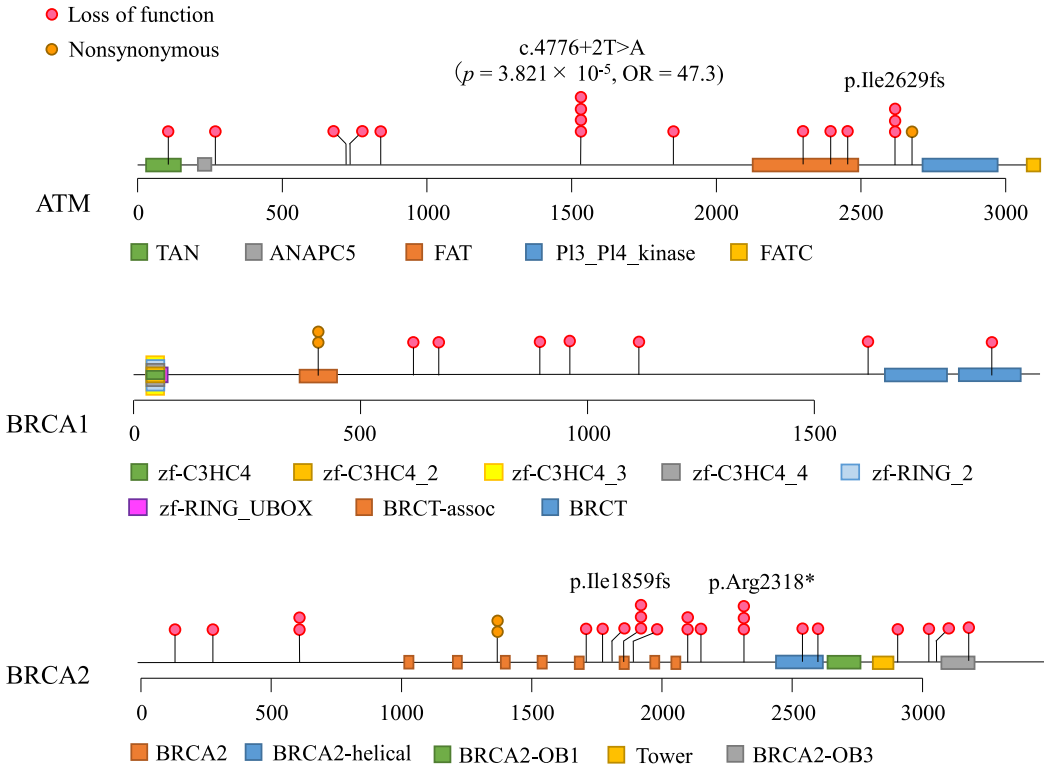

Fig 1. Location and number of frequent pathogenic variants in three genes associated with pancreatic cancer. Locations of frequent pathogenic variants found in patients and of Pfam domains downloaded from the UCSC genome browser table function in proteins are shown with lollipop structures, with the variant type indicated by color. Pink and yellow circles indicate variants of loss of function and nonsynonymous, respectively. The x-axis reflects the number of amino acid residues. ANAPC5 = anaphase-promoting complex subunit 5; BRCT = BRCA1 C terminus; FAT = FRAP, ATM and TRRAP; FATC = FRAP, ATM, TRRAP C-terminal; PI3_PI4_kinase = phosphatidylinositol 3- and 4-kinase; TAN = telomere-length maintenance and DNA damage repair; zf-C3HC4 = zinc finger, C3HC4 type; zf-RING=zing finger, RING type.

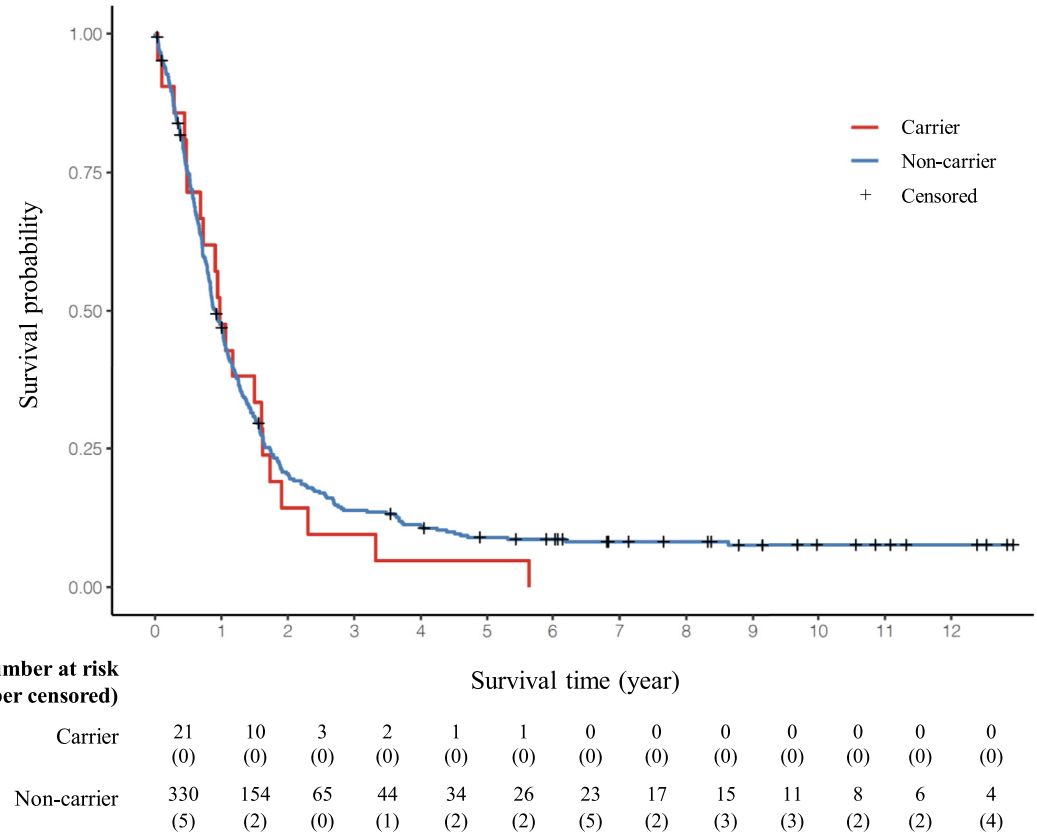

Fig 2. Kaplan-Meier analysis of overall survival for pancreatic cancer patients with and without pathogenic variants in three genes (*BRCA1*, *ATM*, and *BRCA2*).

3.5. Predicting germline pathogenic variant carrier status using machine learning

First, we assessed the predictive performance of machine learning classifiers using breast cancer data with 48 variables as a positive control. The AUCs of six classifiers using data from 2496 breast cancer cases ranged from 0.629 to 0.687; random forest, 0.687 (95% CI = 0.604–0.769); conditional random forest, 0.686 (95% CI = 0.607–0.765); gradient boosting machine, 0.666 (95% CI = 0.574–0.758); naive bayes, 0.673 (95% CI = 0.593–0.754); neural network, 0.629 (95% CI = 0.547–0.711); support vector machine, 0.667 (95% CI = 0.580–0.753) (Fig. 3a). The AUC of a multiple logistic regression classifier was 0.627 (95% CI = 0.536–0.718). Next, we assessed if breast cancer patients met the current NCCN guideline for genetic breast cancer testing (Supplementary Table 9) [13], and then calculated the sensitivity and specificity of the criteria (Supplementary Table 10). The results calculated based on the criteria of "Testing is clinically indicated" indicated 65.9% sensitivity and 61.3% specificity (Supplementary Table 10). The sensitivity and specificity calculated based on the criteria of "Testing may be considered" in addition to those of "Testing is clinically indicated" were 68.3% and 58.5%, respectively. Thus, the predictive performance of machine learning classifiers was equivalent to that of the NCCN guideline criteria for breast cancer testing (Fig. 3a).

We then applied the same classifiers to the pancreatic cancer data. Machine learning exhibited poorer predictive performance across all methods (AUC: random forest, 0.544 [95% CI = 0.439–0.650]; conditional random forest, 0.555 [95% CI = 0.414–0.697]; gradient boosting machine, 0.561 [95% CI = 0.460–0.663]; naive bayes, 0.544 [95% CI = 0.415–0.673]; neural network, 0.550 [95% CI = 0.418–0.682]; support vector machine, 0.514 [95% CI = 0.324– 0.704]) (Fig. 3b). The AUC of a multiple logistic regression classifier was 0.562 (95% CI = 0.425–0.698).

One hundred sets of reduced breast cancer data were examined to evaluate the effect of sample size on machine learning performance. The average performance of the random forest prediction that was applied to the reduced breast cancer datasets was higher than that of the prediction applied to pancreatic cancer dataset, although lower than that of the prediction applied to a complete breast cancer dataset with 2496 cases (AUC: mean ± sd, 0.629 ± 0.090) (Fig. 3c). Eighteen of the 100 reduced breast cancer datasets exhibited a lower AUC than did the pancreatic cancer dataset (Supplementary Fig. 2).

4. Discussion

We identified 205 pathogenic variants in 27 genes of 1005 Japanese patients with pancreatic cancer and 23,705 control individuals from the same racial and regional background. Pathogenic variants were found in 6.7% of unselected Japanese pancreatic cancer patients; *BRCA2, ATM*, and *BRCA1* were identified as statistically significant causative genes. Only a few demographic and clinical characteristics were associated with pathogenic variants in patients with pancreatic cancer. Machine learning classifiers exhibited practical predictive performances for identifying patients with pathogenic variants in breast cancer but not for identifying patients with pathogenic variants in pancreatic cancer.

A significant association between pancreatic cancer diagnosis and pathogenic variant carrier status was observed in *BRCA2, ATM,* and *BRCA1*. As a sensitivity analysis, we conducted a case-control association analysis using a subgroup of cases matched to controls (i.e. > 60 years old and no personal and/or family history of cancer). Although the number of cases decreased from 1005 to 296, *BRCA2, ATM*, and *BRCA1* showed $P < 0.05$ [two-sided Fisher's exact test] and the OR was relatively comparable to the original result (Supplementary Table 11). These genes have also been shown to contribute to pancreatic cancer in the European population, and are included in

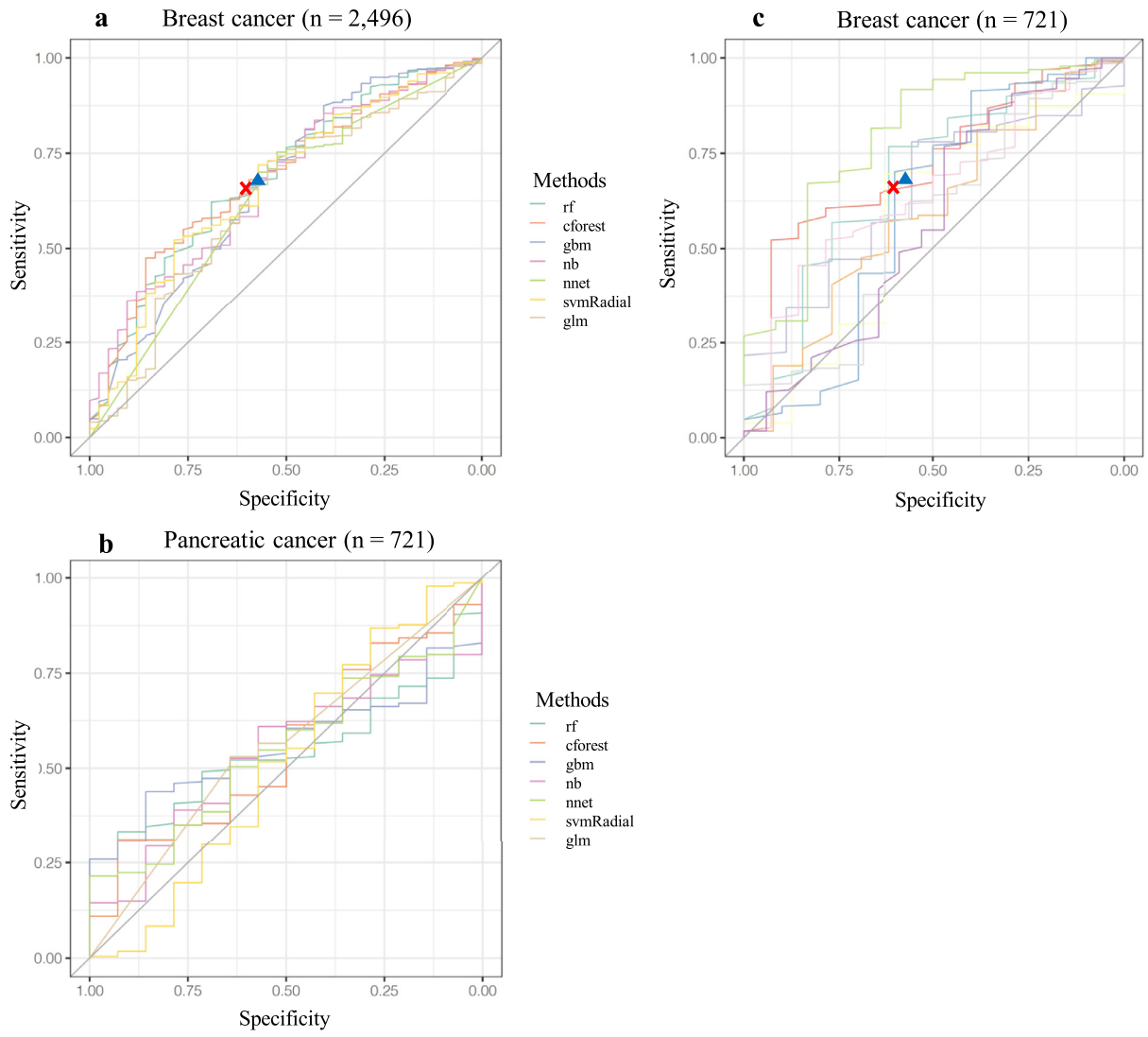

Fig 3. Predictive performance of machine learning for identifying patients with pathogenic variants. ROC curves derived from six supervised machine learning classifiers and multiple logistic regression classifiers for segregating cancer cases with germline pathogenic variants in four genes (*BRCA1, BRCA2, TP53,* and *PALB2*) in breast cancer (A), and in three genes (*BRCA1, ATM* and *BRCA2*) in pancreatic cancer (B). ROC curves using random forest and first ten sets from 100 reduced breast cancer data (C). The sensitivity and specificity (Supplementary Table 10) of predicting the carrier status of breast cancer patients based on the NCCN guideline criteria for germline genetic testing (Supplementary Table 9) are indicated as follows: ×, prediction based on criteria described as "Testing is clinically indicated" in the NCCN guideline; ▲, prediction based on criteria described as "Testing may be considered" as well as "Testing is clinically indicated". rf, random forest; cforest, conditional random forest; gbm, gradient boosting machine; nb, naive bayes; nnet, neural network; svmRadial, support vector machine; glm, multiple logistic regression.

the NCCN guideline for germline genetic testing criteria for pancreatic cancer [7]. In this study, 5.1% of patients with pancreatic cancer were carriers of pathogenic variants in these three genes, comparable to the frequency (4.8%) in the largest investigation in the European population [7]. Because 3.4% of patients harbored pathogenic variants in either *BRCA1* or *BRCA2*, PARP inhibitors are expected to help many Japanese patients with pancreatic cancer [9,30]. Patients with pathogenic variants in *ATM* may be expected to respond to PARP inhibitor treatment because *ATM* is one of homologous recombination DNA damage repair genes [31]. To our knowledge, there is only a case study reporting PARP inhibitor treatment for a patient with a germline pathogenic variant in *ATM* in pancreatic cancer [32], and it would thus be important to evaluate the effect of PARP inhibitors on a large number of pancreatic cancer patients with pathogenic variants in *ATM*. The three other genes (*CDKN2A, TP53*, and *MLH1*) that were shown to be cancer-associated in a previous investigation [7] were not observed to be significantly cancer-associated in this study. Because only one or two carriers of a pathogenic variant in *CDKN2A, TP53,* and *MLH1* were identified in Japanese pancreatic cancer patients and their proportions in this study were comparable to those in the previous study [7], the association of these genes need to be carefully evaluated based on larger investigation in this population. Therefore, *BRCA1, BRCA2*, and *ATM* were identified as the key genes commonly contributing to the risk of pancreatic cancer in Japanese and European populations.

In this study, we identified three characteristics significantly associated with carrier status of pathogenic variants among pancreatic cancer patients: family history of ovarian cancers, family history of gastric cancers, and lymphovascular invasion. These associations have not been previously reported. Supplementary Table 12 shows demographic and clinical characteristics associated with pathogenic variants in cancer-predisposing genes in five investigations [5-7,11,33]. The association between carrier status and family history of gastric cancer was not described anywhere. Pathogenic variants in *ATM* mainly explained the association in family history of gastric cancer, indirectly reflecting the association between *ATM* and gastric

cancer [34]. The association between carrier status and family history of ovarian cancer was examined in a previous investigation of pancreatic cancer [7]. In the investigation, although the proportion of patients having family history of ovarian cancer was larger in carrier patients than in non-carrier patients, the association was not significant. A significant association between carrier status and lymphovascular invasion was not observed in pancreatic cancer patients in a previous investigation [6]. Other significant cancer-associated characteristics have been reported in previous investigations (age of onset, stage of cancer, personal history of cancers other than pancreatic cancer, family history of breast cancer and other cancers, and location of cancer), although the association was not always consistent among those investigations [5-7,11,33]. There was no obvious tendency for carrier frequencies to change depending on age groups, making age-based prediction of carrier patients impractical (Supplementary Fig. 1). The lower incidence rate of breast cancer in Japan than in Western countries might have resulted in the lack of a significant association between the carrier status and family history of breast cancer in this study [3]. Given the lack of common associations identified across investigations to date (including the present study), it would appear that individual demographic and clinical characteristics will not be useful to identify pancreatic cancer patients who should undergo germline genetic testing.

We also examined the potential for predicting pathogenic variant carrier status using machine learning, a suitable technique for interactive high-dimensional data. In this study, we verified the applicability of machine learning for this purpose by using machine learning to analyze breast cancer data and comparing the results to the predictivity of the criteria used in the NCCN germline genetic testing guideline. It should be noted that machine learning was conducted using unselected variables without any adjustment based on existing knowledge, whereas NCCN guideline criteria have been polished several times according to the findings of various investigations [35]. On the other hand, the predictive performance of all classifiers was found to be poor for pancreatic cancer, although the variables used for pancreatic cancer analyses were collected by a biobank project like the one used for breast cancer analyses. The most reasonable explanation is that the association between each characteristic and carrier status is not so strong in pancreatic cancer, and any combinations of those characteristics is still inadequate for predicting carrier status.

This study had some limitations. First, because DNA samples in this study were extracted from peripheral blood, somatic variants caused by clonal hematopoiesis may be reported as germline mutations especially in *TP53* [36]. Compared with germline variants, mosaic variants usually have a lower variant allele fraction (VAF). The mean VAF of pathogenic variants in *TP53* ($N = 12$) in this study was 0.354 (min: 0.254, max: 0.482). The median VAF from clonal hematopoiesis was 0.03 [37]. Therefore, the variants reported in this study are most likely to be a germline variant. Second, we allowed for the presence of comorbidities in the cases and controls recruited in this study, which might have affected the analysis of the association between pathogenic variants and pancreatic cancer. We analyzed the association between pathogenic variants and the personal history of chronic pancreatitis and diabetes mellitus, which are known to be risk factors of pancreatic cancer [38]. As no significant association was observed ($P = 0.619$ in chronic pancreatitis, $P = 0.297$ in diabetes mellitus [two-sided Fisher's exact test]), at least these diseases were unlikely to interfere with the analysis of pancreatic cancer and pathogenic variants. Third, although this study examined the largest number of non-European pancreatic cancer patients to our knowledge, we could not systematically analyze demographic and clinical characteristics associated with carrier status of each individual gene separately because the number of carriers for each gene was 9–25. A larger number of samples is required to investigate if there are novel clinically useful associations between characteristics and carrier status of each gene. Finally, the number of samples used for machine learning modeling might be limited. The predictive performance may be improved by increasing the number of samples in the training cohorts. However, the predictive performance for over 80% of reduced breast cancer datasets was higher than that for the pancreatic cancer data, suggesting a limited possibility of notably improving predictive performance for pancreatic cancer cases by increasing the number of samples in the training cohort.

Conclusion

The genetic characterization of patients with pancreatic cancer revealed that pathogenic variants in *BRCA1*, *BRCA2*, and *ATM* were significantly associated with pancreatic cancer and that at least 3.4% of Japanese patients would show a favorable response to PARP inhibitor treatment, indicating the significance of germline genetic testing. Predicting whether a pancreatic cancer patient was a carrier of a pathogenic variant in a cancer-predisposing gene based on clinical and demographic characteristics was found to be difficult, even when using machine learning classifiers. We thus concluded that offering germline genetic testing for all pancreatic cancer patients would be a reasonable strategy.

Contributors

YuM conceived and designed the study. MH, KoM, and KS recruited patients. KeM, YI, ME, and YuM performed the sequencing and data analyses. KeM, YK, TY, YoM, HN, and YuM performed the statistical analyses and data interpretation. KeM and EK performed the machine learning analyses. KeM, AS, and YuM wrote the manuscript and all authors reviewed the final version.

Declaration of Competing Interests

YuM received a grant from Japan Agency for Medical Research and Development (AMED). AS received a grant from Australian National Health and Medical Research funding. The other authors declare no competing interests.

Acknowledgments

This research was supported by Japan Agency for Medical Research and Development (AMED) under Grant No. JP19kk0305010. We thank the individuals who participated in this study. We acknowledge the staff of the Laboratory for Genotyping Development in RIKEN, the RIKEN-IMS sequence platform and the BioBank Japan project. The work of AS was supported by Australian National Health and Medical Research funding (ID177524).

Data sharing

The sequence data are under submission at NDBC human database (NBDC Research ID: hum0014.v19) (https://humandbs.biosciencedbc.jp/en/hum0014-v19). The data will be available under NBDC Data Sharing Policy (controlled-access data Type-1) after this paper is accepted.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2020.103033.

References

[1] GBD 2017 Pancreatic Cancer Collaborators. The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and

territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet Gastroenterol Hepatol 2019;4(12):934–47.
[2] Center for Cancer Control and Information Services NCC. Monitoring of cancer incidence in Japan - survival 2006-2008 report. 2016.
[3] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global cancer observatory: cancer today [Internet]. Lyon, France: International Agency for Research on Cancer; 2018. [cited 2020 Aug 15]. Available from: https://gco.iarc.fr/today.
[4] Takai E, Yachida S, Shimizu K, Furuse J, Kubo E, Ohmoto A, et al. Germline mutations in Japanese familial pancreatic cancer patients. Oncotarget 2016;7(45):74227–35.
[5] Brand R, Borazanci E, Speare V, Dudley B, Karloski E, Peters MLB, et al. Prospective study of germline genetic testing in incident cases of pancreatic adenocarcinoma. Cancer 2018;124(17):3520–7.
[6] Yurgelun MB, Chittenden AB, Morales-Oyarvide V, Rubinson DA, Dunne RF, Kozak MM, et al. Germline cancer susceptibility gene variants, somatic second hits, and survival outcomes in patients with resected pancreatic cancer. Genet Med 2019;21(1):213–23.
[7] Hu C, Hart SN, Polley EC, Gnanaolivu R, Shimelis H, Lee KY, et al. Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. JAMA 2018;319(23):2401–9.
[8] Del Chiaro M, Verbeke CS, Kartalis N, Pozzi Mucelli R, Gustafsson P, Hansson J, et al. Short-term results of a magnetic resonance imaging-based swedish screening program for individuals at risk for pancreatic cancer. JAMA Surg 2015;150(6):512–8.
[9] Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ, et al. Maintenance olaparib for germline BRCA-mutated metastatic pancreatic cancer. N Engl J Med 2019;381(4):317–27.
[10] Wood LD, Yurgelun MB, Goggins MG. Genetics of familial and sporadic pancreatic cancer. Gastroenterology 2019;156(7):2041–55.
[11] Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. J Clin Oncol 2017;35(30):3382–90.
[12] Lowery MA, Wong W, Jordan EJ, Lee JW, Kemel Y, Vijai J, et al. Prospective evaluation of germline alterations in patients with exocrine pancreatic neoplasms. J Natl Cancer Inst 2018;110(10):1067–74.
[13] Daly MB, Pilarski R, Yurgelun MB, Berry MP, Buys SS, Dickson P, et al. NCCN guidelines insights: genetic/familial high-risk assessment: breast, ovarian, and pancreatic, version 1.2020. J Natl Compr Canc Netw 2020;18(4):380–91.
[14] Hu C, LaDuca H, Shimelis H, Polley EC, Lilyquist J, Hart SN, et al. Multigene hereditary cancer panels reveal high-risk pancreatic cancer susceptibility genes. JCO Precis Oncol 2018;2.
[15] Huang BZ, Stram DO, Le Marchand L, Haiman CA, Wilkens LR, Pandol SJ, et al. Interethnic differences in pancreatic cancer incidence and risk factors: the Multiethnic Cohort. Cancer Med 2019;8(7):3592–603.
[16] Momozawa Y, Iwasaki Y, Hirata M, Liu X, Kamatani Y, Takahashi A, et al. Germline pathogenic variants in 7,636 Japanese patients with prostate cancer and 12,366 controls. J Natl Cancer Inst 2019.
[17] Momozawa Y, Iwasaki Y, Parsons MT, Kamatani Y, Takahashi A, Tamura C, et al. Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls. Nat Commun 2018;9(1):4083.
[18] Hirata M, Kamatani Y, Nagai A, Kiyohara Y, Ninomiya T, Tamakoshi A, et al. Cross-sectional analysis of BioBank Japan clinical data: a large cohort of 200,000 patients with 47 common diseases. J Epidemiol 2017;27(3S):S9–S21.
[19] Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, et al. Overview of the bioBank Japan project: study design and profile. J Epidemiol 2017;27(3S):S2–8.
[20] Fujita M, Liu X, Iwasaki Y, Terao C, Takata S, Inai C, et al. Population-wide screening for germline variants of hereditary cancer genes in 12K unselected Japanese colorectal cancers and 27K controls. bioRxiv 2020.
[21] Buys SS, Sandbach JF, Gammon A, Patel G, Kidd J, Brown KL, et al. A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. Cancer 2017;123(10):1721–30.
[22] Nakagawa H, Lockman JC, Frankel WL, Hampel H, Steenblock K, Burgart LJ, et al. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res 2004;64(14):4721–7.
[23] Momozawa Y, Akiyama M, Kamatani Y, Arakawa S, Yasuda M, Yoshida S, et al. Low-frequency coding variants in CETP and CFB are associated with susceptibility of exudative age-related macular degeneration in the Japanese population. Hum Mol Genet 2016;25(22):5027–34.
[24] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. Genet Med 2015;17(5):405–24.
[25] Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. Nat Rev Genet 2017;18(10):599–612.
[26] Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 2016;44(D1):D862–8.
[27] Max K. Building predictive models in R using the caret package. J Stat Softw 2008;28:1–26.
[28] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.
[29] Therneau T. A package for survival analysis in R.2019 [Internet]: 2020[cited 2020 March 27]. Available from:https://rdrr.io/cran/survival/f/inst/doc/survival.pdf.
[30] Jonsson P, Bandlamudi C, Cheng ML, Srinivasan P, Chavan SS, Friedman ND, et al. Tumour lineage shapes BRCA-mediated phenotypes. Nature 2019;571(7766):576–9.
[31] Heeke AL, Pishvaian MJ, Lynce F, Xiu J, Brody JR, Chen WJ, et al. Prevalence of homologous recombination-related gene mutations across multiple cancer types. JCO Precis Oncol 2018;2018.
[32] Martino C, Pandya D, Lee R, Levy G, Lo T, Lobo S, et al. ATM-mutated pancreatic cancer: clinical and molecular response to Gemcitabine/Nab-paclitaxel after genome-based therapy resistance. Pancreas 2020;49(1):143–7.
[33] Grant RC, Selander I, Connor AA, Selvarajah S, Borgida A, Briollais L, et al. Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. Gastroenterology 2015;148(3):556–64.
[34] Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic germline variants in 10,389 adult cancers. Cell 2018;173(2) 355-70.e14.
[35] Daly MB, Pilarski R, Berry M, Buys SS, Farmer M, Friedman S, et al. NCCN guidelines insights: genetic/familial high-risk assessment: breast and ovarian, version 2.2017. J Natl Compr Canc Netw 2017;15(1):9–20.
[36] Weitzel JN, Chao EC, Nehoray B, Van Tongeren LR, LaDuca H, Blazer KR, et al. Somatic TP53 variants frequently confound germ-line testing results. Genet Med 2018;20(8):809–16.
[37] Coombs CC, Gillis NK, Tan X, Berg JS, Ball M, Balasis ME, et al. Identification of clonal hematopoiesis mutations in solid tumor patients undergoing unpaired next-generation sequencing assays. Clin Cancer Res 2018;24(23):5918–24.
[38] Andersen DK, Andren-Sandberg Å, Duell EJ, Goggins M, Korc M, Petersen GM, et al. Pancreatitis-diabetes-pancreatic cancer: summary of an NIDDK-NCI workshop. Pancreas 2013;42(8):1227–37.