Can Machine Learning Models Predict Asparaginase-associated Pancreatitis in Childhood Acute Lymphoblastic Leukemia

Nielsen, Rikke L.

2022-04

http://hdl.handle.net/10138/342783
https://doi.org/10.1097/MPH.0000000000002292

ORIGINAL ARTICLE

OPEN

Can Machine Learning Models Predict Asparaginase-associated Pancreatitis in Childhood Acute Lymphoblastic Leukemia

Rikke L. Nielsen, PhD,*†‡ Benjamin O. Wolthers, MD, PhD,‡
Marianne Helenius, MSc,* Birgitte K. Albertsen, MD, PhD,§
Line Clemmensen, PhD,‖ Kasper Nielsen, PhD,¶ Jukka Kanerva, MD, PhD,#
Riitta Niinimäki, MD, PhD,** Thomas L. Frandsen, MD, PhD,‡
Andishe Attarbaschi, MD, PhD,†† Shlomit Barzilai, MD,‡‡
Antonella Colombini, MD,§§ Gabriele Escherich, MD,‖‖
Derya Aytan-Aktug, MSc,¶¶ Hsi-Che Liu, MD,## Anja Möricke, MD,***
Sujith Samarasinghe, MD, PhD,††† Inge M. van der Sluis, MD, PhD,‡‡‡
Martin Stanulla, MD, PhD,§§§ Morten Tulstrup, MD,‡
Rachita Yadav, PhD,¶ Ester Zapotocka, MD, PhD,‖‖‖‖
Kjeld Schmiegelow, MD, PhD,‡¶¶¶ and Ramneek Gupta, PhD*

Summary: Asparaginase-associated pancreatitis (AAP) frequently affects children treated for acute lymphoblastic leukemia (ALL) causing severe acute and persisting complications. Known risk factors such as asparaginase dosing, older age and single nucleotide polymorphisms (SNPs) have insufficient odds ratios to allow personalized asparaginase therapy. In this study, we explored machine learning strategies for prediction of individual AAP risk. We integrated information on age, sex, and SNPs based on Illumina Omni2.5exome-8 arrays of patients with childhood ALL (N = 1564, 244 with AAP 1.0 to 17.9 yo) from 10 international ALL consortia into machine learning models including regression, random forest, AdaBoost and artificial neural networks. A model with only age and sex had area under the receiver operating characteristic curve (ROC-AUC) of 0.62. Inclusion of 6 pancreatitis candidate gene SNPs or 4 validated pancreatitis SNPs boosted ROC-AUC somewhat (0.67) while 30 SNPs, identified through our AAP genome-wide association study cohort, boosted performance (0.80). Most predictive features included rs10273639 (*PRSS1-PRSS2*), rs10436957 (*CTRC*), rs13228878 (*PRSS1/PRSS2*), rs1505495 (*GALNTL6*), rs4655107 (*EPHB2*) and age (1 to 7 y). Second AAP

---

following asparaginase re-exposure was predicted with ROC-AUC: 0.65. The machine learning models assist individual-level risk assessment of AAP for future prevention trials, and may legitimize asparaginase re-exposure when AAP risk is predicted to be low.

Key Words: pediatric hematology/oncology, acute lymphoblastic leukemia, treatment toxicity, translational research, artificial intelligence

(*J Pediatr Hematol Oncol* 2022;44:e628–e636)

Asparaginase is an essential drug in childhood acute lymphoblastic leukemia (ALL) treatment associated with increased survival rates.[1] By depleting circulating asparagine levels, malignant lymphoblasts are targeted for apoptosis, due to limited capacity for resynthesis of asparagine.[2] Asparaginase use is, however, associated with significant treatment related toxicities[3] of which pancreatitis (asparaginase-associated pancreatitis [AAP]) occurs in 2% to 18% of patients,[4] mostly in older children and adults.[5] Frequently, AAP leads to truncation of therapy, potentially increasing the risk of relapse.[1,6,7] Re-exposure to asparaginase after AAP has been associated with an almost 50% risk for a second AAP, but only after several doses of pegylated asparaginase (PegAsp).[8] Previous studies have identified older age at diagnosis,[4,5] and host genome variants[4,9,10] as AAP risk factors. However, these risk factors are currently not used to individualize treatment with asparaginase, as they only have modest effect sizes for clinical decision support. The study demonstrates how far machine learning methodologies can guide identification of AAP risk patients in childhood ALL.

MATERIALS AND METHODS

To address this challenge, we integrated germline single nucleotide polymorphisms (SNPs) from a childhood ALL AAP case-control cohort (N = 1564, including 244 AAP cases) into machine learning models to predict patients at very high risk of AAP and low risk of second AAP. We and others have previously applied machine learning modeling to identify patients with childhood ALL at high risk of relapse.[11,12] Individual level predictions across several machine learning models can be compared with improve understanding of relevant risk factors that appear in combination associated with AAP, and potential patient subgroups that are predictable by separate models. By identifying patients at high confidence for risk of AAP guided by the individual probabilities provided by the machine learning models, this analysis may facilitate research on targeted preemptive measures (eg, somatostatin) for patients at high risk of AAP. We also tested if the use of AAP SNP machine learning models performed equally well for predicting the risk of developing a second AAP following re-exposure to asparaginase. The need for identification of low-risk patients who are likely to tolerate re-exposure to PegAsp after their first AAP episode are even more important for avoiding the risk of relapse by guiding decision on whether patients are safe to re-expose after their first AAP given germline predisposition.

Patients

To map AAP phenotypes and identify significant host genome variants associated with AAP risk, the international Ponte di Legno (PdL) toxicity working group (PTWG) collected post remission blood samples from 1564 children (1.0 to 17.9 yo) with newly diagnosed t(9;22)-negative ALL between June 1, 1996, and January 1, 2016.[10] Oral and written consent was obtained. The database containing phenotype data was

approved by the regional ethical review board of The Capital Region of Denmark (H-2-2010-022), the Danish Data Protection Authorities (j.nr.: 2012-58-0004), and by relevant regulatory authorities in all participating countries.

All patients received asparaginase according to their respective treatment protocols. Patients treated by the Nordic Society of Pediatric Hematology and Oncology (NOPHO) ALL-2008 treatment protocol (92 AAP cases and 1024 controls) received 30 weeks of asparaginase exposure with pegylated asparaginase at 2 weeks intervals (15 doses in total), although a subset were randomized to pegylated asparaginase at 6 weeks intervals after the first 5 doses (8 doses in total).[13] Patients treated by other protocols received asparaginase for <30 weeks. The applied diagnostic criteria for AAP stated that 2 of the 3 following international consensus criteria must be fulfilled: (i) amylase, pancreatic amylase, or pancreatic lipase > 3× upper normal limit, (ii) abdominal pain, or (iii) imaging compatible with AAP.[14] DNA was genotyped on Illumina Omni2.5exome-8 BeadChip arrays. After quality control as previously described,[10] the genotype data consisted of 1,401,908 SNPs. Information on age and sex was available for all patients while patients treated with the NOPHO ALL-2008 protocol had more clinical features available including country, weight, length, immunophenotype, risk stratification group, white blood cell count at diagnosis, minimal residual disease (measured with flow cytometry and polymerase chain reaction at day 29) and asparaginase dosage information (dosage interval [2 or 6 wk intervals] and total number of asparaginase dosages) (Supplementary Methods S1, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474).

Machine Learning Training, Feature Importance, and Validation

Logistic regression, random forest, AdaBoost, and artificial neural networks (including 1 and 2 hidden layers) models were fitted using python (version 3.6.8)[15] with Scikit-learn (version 0.21.3).[16] Several multivariate machine learning models were trained to capture both linear and nonlinear interactions between genetic features. Feature importance was evaluated by the area under the receiver operating characteristic curve (ROC-AUC) loss using a "leave-one-out" approach on correlated features. To eliminate population substructure, only patients of European ancestry were included for training of the machine learning models in the study (N = 1390) whereof 205 patients developed AAP and 1185 did not. Performance was evaluated in 2 independent hold-out validation data sets that was stratified according to patients' genetic ancestry. A hold-out validation data set with European ancestry included 100 patients (50 AAP cases) that included all the 37 patients re-exposed to asparaginase after truncation extracted before training of the machine learning models leaving 1290 patients for model training (155 AAP cases). The second hold-out validation data set included 174 patients (39 AAP cases) with a non-European genetic ancestry. Performance was also evaluated on the 37 re-exposed patients to evaluate the predictive performance on a second AAP (13 cases) (Supplementary Methods S1, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474).

Genetic Feature Representation and Selection Strategies

To maximize learning from genetic data, SNPs were represented by additive, dominant and recessive genetic encodings as well as a nonadditive genetic encoding according to the presence of the major allele or minor allele (binary allele)
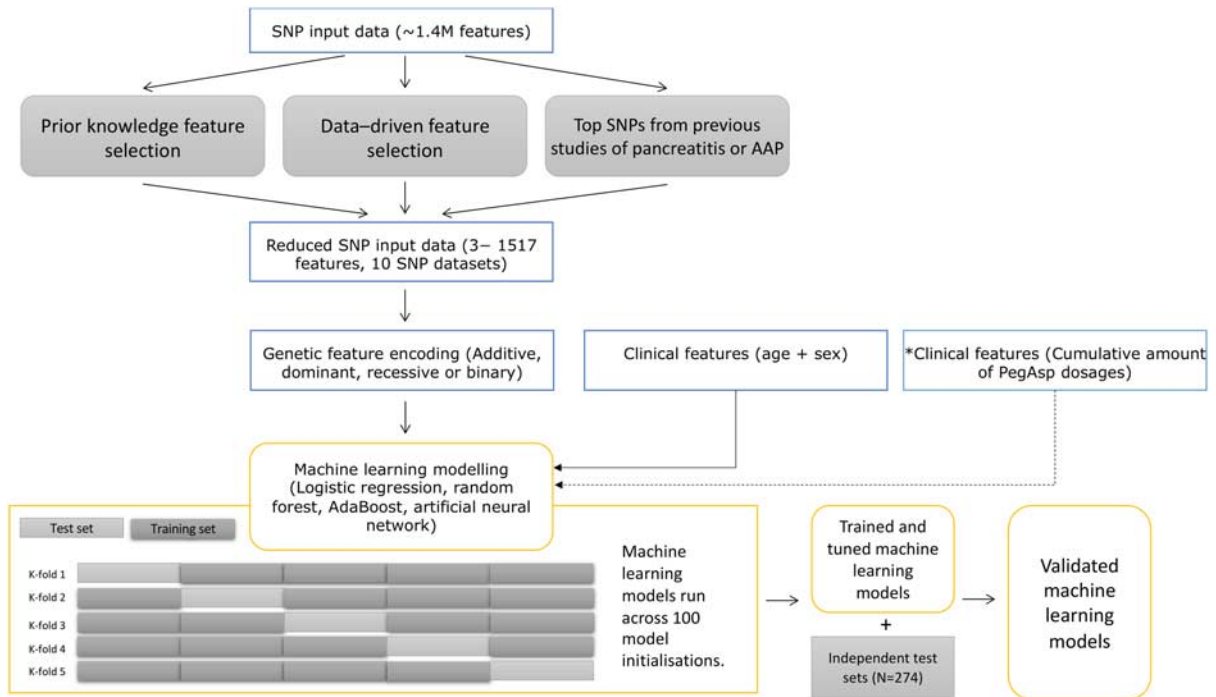
FIGURE 1. Overview of the feature selection and machine learning strategies used in the study. *A future model would benefit from inclusion of the cumulative dosage of pegylated asparaginase (PegAsp). In this study, it was only available on a subset of patients and was thus not fully explored. Age and sex were always included in modeling. AAP indicates asparaginase-associated pancreatitis; SNP, single nucleotide polymorphism.

in the machine learning models. To guide training of the models, SNPs were selected by different strategies to test their predictiveness of AAP (Fig. 1). Feature selection included SNPs previously associated with pancreatitis[17–19] or AAP.[4,9,10]

rs13228878 (*PRSS1/PRSS2*) was previously validated in the Children's Oncology Group's AALL0232 cohort,[10] while rs10273639 (*PRSS1-PRSS2*), rs17107315 (*SPINK1*) and rs12688220 (*CLDN2/MORC4*) have been identified and
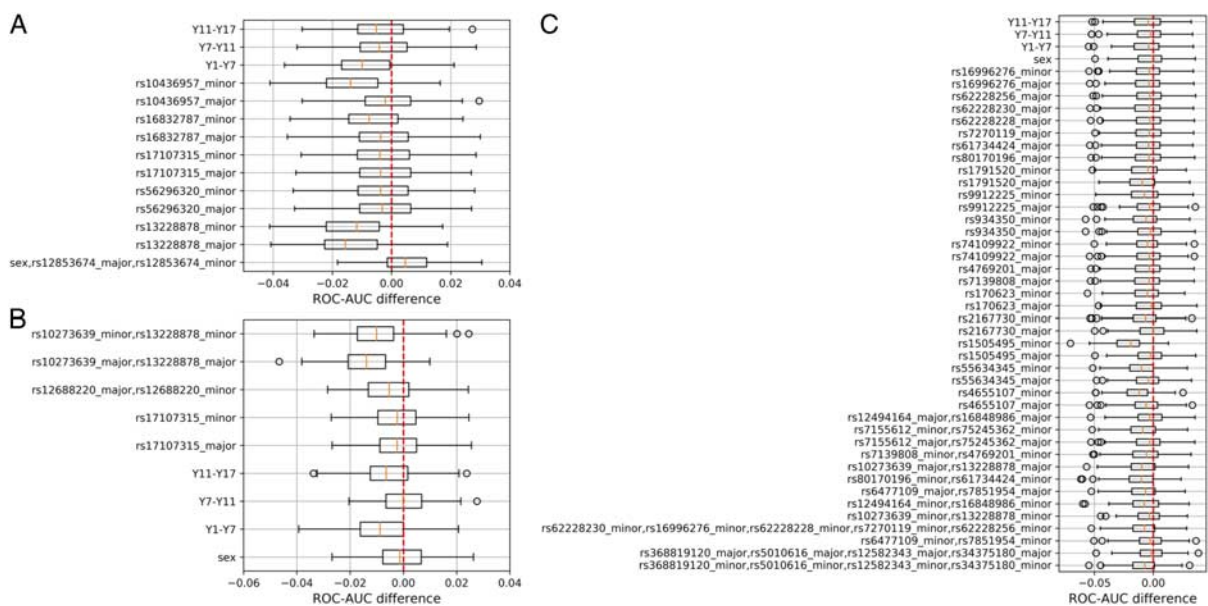
FIGURE 2. Leave-one-out area under the receiver operating characteristic curve (ROC-AUC) feature importance for asparaginase-associated pancreatitis risk models. Models were trained on N = 1290 patients using artificial neural networks with 1 hidden layer (A) using age, sex and 6 candidate single nucleotide polymorphisms (SNPs) as features. B, Using age, sex and 4 previously validated SNPs as features. C, Using age, sex and top 30 SNPs associated with asparaginase-associated pancreatitis from Wolthers et al[10] genome-wide association study as features.

validated in alcoholic and nonalcoholic pancreatitis studies.[17–19] Recent AAP genome-wide association study (GWAS) discovered shared genetic predisposition between AAP and non-AAP pathways.[10] We thus also explored predictability using SNPs annotated to 8 candidate genes involved in development of pancreatitis, that is, *PRSS1*, *PRSS2*, *SPINK1*, *CTRC*, *CASR*, *CFTR*, *CPA1*, and *CLDN2*[20] and their expression quantitative trait loci (eQTLs) from the GTEx biobank (v6)[21,22] using prior knowledge on the most severe functional SNPs consequences to subset data sets with prioritized SNPs for modeling. SNPs annotated to the 8 candidate genes (with minor allele frequency > 5%) were in addition reduced to 3 principal components to model the variance explained by all variants. Furthermore, 6 SNPs; rs10436957 (*CTRC*), rs12853674 (*CLDN2*), rs13228878 (*PRSS1/PRSS2*), rs16832787 (*CASR*), rs17107315 (*SPINK1*), and rs56296320 (*CFTR*) were identified as most significant SNPs in candidate genes of pancreatitis[20] in the PdL AAP GWAS[10] (Supplementary Methods S1, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474).

Ensemble Model

An ensemble of prediction models optimized for prediction of AAP was created without increasing the complexity of any individual model. The ensemble model was scored by 3 different approaches (i) average mean scoring, (ii) majority voting or (iii) average mean scoring of confident individual predictions, that is, the score should be ≤ 0.30 or ≥ 0.70 to count in the final prediction (Supplementary Methods S1, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474).

RESULTS

Clinical Baseline Model

The PdL AAP cohort included 1564 children whereof 1390 patients had European ancestry (205 AAP cases) that were considered for training of the machine learning models. The AAP cases (N = 1390, 205 AAP cases) had older age (case: 8.7 ± 4.8, controls: 6.3 ± 4.5, *P* = 1.4e−12) and no difference in sex (*P* = 0.82) (Supplementary Table S.2, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). A genome-wide association analysis on the patients with only European background (N = 1390) is available in Supplementary Table S.3 (Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). No SNPs reached genome-wide significance. A clinical baseline model of AAP was established using only age and sex as features which resulted in ROC-AUCs of 0.62 ± 0.01 (Supplementary Table S.4, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). SNPs were integrated to the clinical baseline models using prior information of pancreatitis pathways and SNPs or previous AAP studies to investigate how predictive germline predisposition are of AAP.

Integration of Genetic Risk Variants in Pancreatitis Pathways

Given the shared genetic predisposition between AAP and pancreatitis,[10] 6 different data sets were used to test the predictiveness of SNPs annotated to 8 genes involved in pancreatitis pathways; *PRSS1*, *PRSS2*, *SPINK1*, *CTRC*, *CASR*, *CFTR*, *CPA1*, and *CLDN2* (Supplementary Methods S1, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474: *6 data sets of SNPs annotated to 8 genes involved in pancreatitis pathway*). The predictive performance for models with age, sex and the selected SNP data sets ranged with ROC-AUC: 0.47 to 0.67 (Supplementary Table S.5, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). Overall,

the best performance was achieved from the 6 candidate SNPs, age and sex model with ROC-AUC: 0.67 (ROC curve in Supplementary Fig. S.6A, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The performance was robust across 100 model initializations compared with a permuted outcome label of AAP with significantly higher ROC-AUC for the true AAP models (*P* < < 1e−6, Supplementary Fig. S.6B, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The most predictive features were rs13228878 (*PRSS1/PRSS2*), the minor allele of rs10436957 (*CTRC*) and the age group of 1 to 7 years (Fig. 2A).

We also explored the impact of integration of previously validated pancreatitis SNPs in studies of alcoholic and non-alcoholic pancreatitis[17–19] (Supplementary Methods S1, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474: *4 data sets of SNPs previously associated with AAP or pancreatitis*). Integration of previously validated pancreatitis SNPs (rs13228878 (*PRSS1/PRSS2*), rs17107315 (*SPINK1*), rs10273639 (*PRSS1/PRSS2*), rs12688220 (*CLDN2/MORC4*) resulted in ROC-AUC: 0.67 with age and sex (Table 1, ROC curve in Supplementary Fig. S.6C, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474) where the ROC-AUC of the models was robust across 100 model initializations compared with a permuted outcome label of AAP with significantly higher ROC-AUC for the true AAP models (*P* < < 1e−6, Supplementary Fig. S.6D, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The most important features were rs10273639 (*PRSS1/PRSS2*), rs13228878 (*PRSS1/PRSS2*) and the age group of 1 to 7 years (Fig. 2B).

Integration of AAP-associated Genetic Variants

Finally, SNPs previously associated with AAP were obtained from 3 previous genetic studies by; Liu et al,[4] Abaji et al,[9] and Wolthers et al.[10] SNPs identified by Liu and colleagues and Abaji and colleagues did not change ROC-AUC compared with the clinical baseline model (ROC-AUC: 0.60 to 0.63, Supplementary Table S5, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The top 30 AAP-associated SNPs reported from an AAP GWAS by Wolthers and colleagues resulted in ROC-AUC: 0.80 (Table 1, ROC curve and permutation test in Supplementary Figs. S.6E, F, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The performance appeared to be independent of the type of machine learning model used, as well as of additive, dominant or binary allele encoding of the genetic variants. However, using the recessive encoding of genetic features resulted in multiple near-zero variance predictors due to very few homozygous recessive alleles, and thus lower ROC-AUC: 0.67 to 0.70 (Supplementary Table S.5, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474).

The most important features were the minor alleles of rs1505495 (*GALNTL6*) and rs4655107 (*EPHB2*) (Fig. 2C). We tested with a forward selection algorithm if age, sex, rs1505495 and rs4655107 were just as predictive on their own. Approximately 25 features were selected by the algorithm supporting a combination of SNPs is required for prediction of AAP (Supplementary Table S.7, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The model was robust across 100 model initializations compared with a permuted outcome label with significantly higher ROC-AUC for the true AAP-labeled models (*P* < < 1e−6, Supplementary Fig. S.6F, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). This model was significantly more confident in AAP patients with higher age (6 y and above, mean [95% confidence interval] of individual

TABLE 1. ROC-AUC Performances Reported as Mean ± SD for the Training Data Set (N = 1290, 155 AAP Cases), Hold-out Validation Data Set With N = 100 Patients With European Ancestry (50 AAP Cases and Controls), Hold-out Validation With N = 174 Patients With Non-European Ancestry (39 AAP Cases) and a Subset of the 37 Patients With European Ancestry That Were Re-exposed to Asparaginase (13 AAP Cases) Across 100 Model Initializations in 5-fold Cross-validation

Data Type	Model	ROC-AUC (N = 1290)	ROC-AUC CEU Validation (N = 100)	ROC-AUC Non-CEU Validation (N = 174)	ROC-AUC Validation 2nd AAP (N = 37)
Six candidate SNPs	ANN (1 hidden layer), binary allele encoding	0.67 ± 0.01	0.64 ± 0.01	0.62 ± 0.01	0.60 ± 0.01
Validated pancreatitis SNPs	ANN (1 hidden layer), binary allele encoding	0.67 ± 0.01	0.64 ± 0.01	0.63 ± 0.01	0.57 ± 0.01
Top 30 *P*-value SNPs	ANN (1 hidden layer), binary allele encoding	0.80 ± 0.01	0.84 ± 0.01	0.72 ± 0.01	0.55 ± 0.04

All models are trained with down-sampling on the control group within the cross-validation folds.

AAP indicate asparaginase-associated pancreatitis; ANN, artificial neural network; CEU, Utah Residents (CEPH) with Northern and Western European ancestry; ROC-AUC, area under the receiver operating characteristic curve; SNPs, single nucleotide polymorphisms.

risk of AAP: 0.74 [0.69 to 0.78]) compared with children younger than 6 years (mean [95% confidence interval] of individual risk of AAP: 0.58 [0.52 to 0.65], *P* = 0.0001). Addition of 2 previously validated SNPs that were not part of the 30 SNPs data set (rs12688220 and rs17107315) did not improve the performance further (Supplementary Table S.7, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). As an attempt, we tried to redo the PTWG GWAS within a 30% hold-out setup (Supplementary Methods S1, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). This however reduced ROC-AUC to 0.59 (Supplementary Table S.7, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474).

Validation of Models

Data sets left out from the original training of the models was used for validation of the AAP risk models in patients with a European ancestry (N = 100, 50 cases of AAP) or a non-European ancestry (N = 174, 39 cases of AAP) as well as validation of the ability to predict risk of second AAP (N = 37, 13 cases of second AAP). The most successful models validated the risk of AAP (Table 1). However, SNPs identified by Wolthers and colleagues had reduced capability of predicting non-European patients (ROC-AUC: 0.72) as compared with European patients (ROC-AUC: 0.84). The trained models predicted risk of second AAP with reduced performance (ROC-AUC: 0.55 to 0.60). The most predictive SNPs from the models (rs13228878, rs10436957, rs10273639, rs1505495, rs4655107) are summarized in Table 2.

Personalized Artificial Intelligence (AI) Ensemble Model

The most predictive models compared with the clinical baseline model (ROC-AUC ≥ 0.62) with different genetic encoding and features capturing different subsets of patients were integrated into an ensemble model based on sensitivity to improve prediction in AAP cases. This ensemble model was composed of 50 models capturing different individuals across predictions across a total of 18,000 possible. For establishing a joint prediction score on each patient, the scores of the individual models within the ensemble were combined via (a) averaging, (b) majority voting, and (c) averaging only on confident scores.

TABLE 2. Overview of the Most Predictive AAP Single Nucleotide Polymorphisms

SNPs	Chromosome	Position	Minor Allele Frequency in Training (N = 1290), CEU Validation (N = 100), Non-CEU Validation (N = 174), Second AAP Validation (N = 37)	Model	Odds Ratio	*P* (Original Study)	Gene
rs13228878	7	142473466	0.40, 0.42, 0.44, 0.39	6 candidate SNPs*	0.6261	1.275e−05	*PRSS1 (+12.54 kb)/ PRSS2 (−6.441 kb)*
				Previously validated SNPs†	NA	0.03	
rs10436957	1	15768304	0.23, 0.22, 0.19, 0.16	6 candidate SNPs*	0.6643	0.00199	*CTRC (0 kb)*
rs10273639	7	142456928	0.41, 0.40, 0.45, 0.36	Previously validated SNPs‡	1.4	2.0e−14	*PRSS1 (−0.39 kb)/ PRSS2 (−22.98 kb)*
rs1505495	4	172973580	0.16, 0.13, 0.17, 0.11	Top 30 PTWG*	0.4974	1.856e−05	*GALNTL6 (0 kb)*
rs4655107	1	23094454	0.24, 0.22, 0.11, 0.18	Top 30 PTWG*	0.5573	3.972e−05	*EPHB2 (0 kb)*

*Odds ratio and *P*-value is obtained from the PTWG AAP GWAS by Wolthers et al.[10]

†Odds ratio and *P*-value reported for validated variant from the PTWG AAP GWAS 2019 by Wolthers et al.[10]

‡Odds ratio and *P*-value reported as in Table S.3 (Supplemental Digital Content 1, http://links.lww.com/JPHO/A474), Rosendahl et al.[19]

AAP indicates asparaginase-associated pancreatitis; CEU, Utah Residents (CEPH) with Northern and Western European ancestry; GWAS, genome-wide association study; NA, not applicable; PTWG, Ponte di Legno toxicity working group; SNPs, single nucleotide polymorphisms.
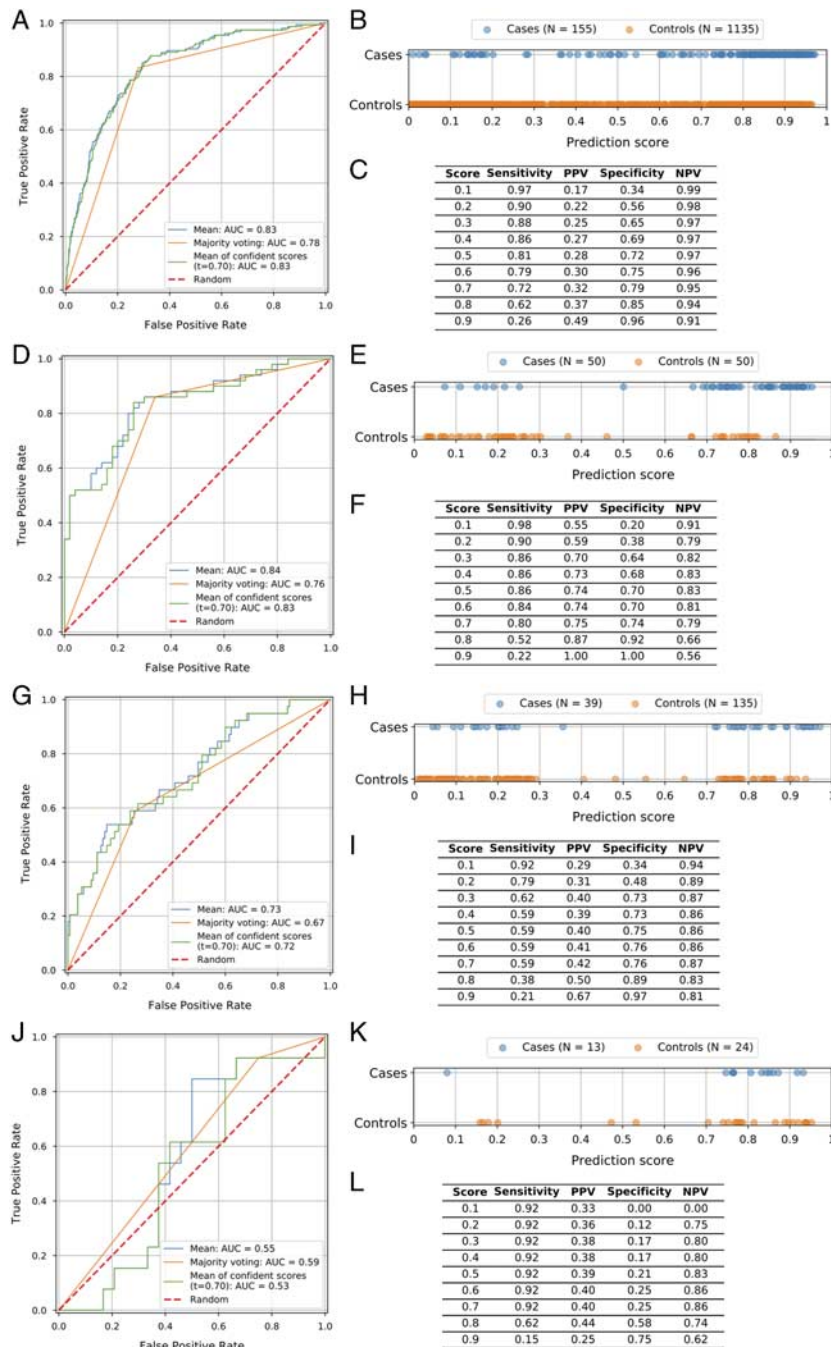
FIGURE 3. Personalized artificial intelligence ensemble models based on mean of scores, majority voting and mean of confident scores (*t* = 0.7). A, ROC curve for the ensemble when predicting on the training data set (N = 1290). B and C, Plot of prediction scores vs true class and table of performance metrics for different score thresholds when scoring the predictions on the training data set (N = 1290) model ensemble with the mean of confident scores (score threshold of ≤ 0.30 or ≥ 0.70). D, ROC curve for the ensemble when predicting on the European hold-out data set (N = 100). E and F, Plot of prediction scores versus true class and table of performance metrics for different score thresholds when scoring the predictions on the European hold-out data set (N = 100) model ensemble with the mean of confident scores (score threshold of ≤ 0.30 or ≥ 0.70). G, ROC curve for the ensemble when predicting on the non-European hold-out data set (N = 174). H and I, Plot of prediction scores versus true class and table of performance metrics for different score thresholds when scoring the predictions on the non-European hold-out data set (N = 174) model ensemble with the mean of confident scores (score threshold of ≤ 0.30 or ≥ 0.70). J, ROC curve for the ensemble when predicting secondary AAP cases. K and L, Plot of prediction scores versus true class and table of performance metrics for different score thresholds when scoring the predictions on the second AAP phenotype (N = 37) model ensemble with the mean of confident scores (score threshold of ≤ 0.30 or ≥ 0.70). AAP indicates asparaginase-associated pancreatitis; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic; Score, applied prediction score threshold for classification (≥ score).
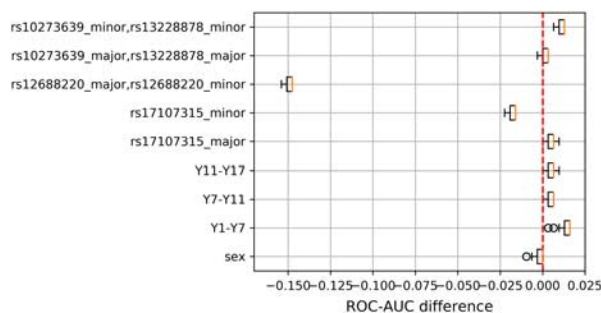
FIGURE 4. Leave-one-out area under the receiver operating characteristic curve (ROC-AUC) feature importance for asparaginase-associated pancreatitis re-exposure model using a logistic regression trained to predict second cases of asparaginase-associated pancreatitis when re-exposed to asparaginase (N = 37, 13 cases). The model used age, sex, and previously validated single nucleotide polymorphisms trained with leave-one-out cross-validation.

Using only confident scores (score threshold of ≤ 0.30 or ≥ 0.70) resulted in ROC-AUC: 0.83 on the cross-validation data set (N = 1290) and on the European hold-out test data set (N = 100) (Figs. 3A, D). The ROC-AUC slightly improved from the best model with age, sex and 30 previously associated AAP SNPs (ROC-AUC = 0.80 to 0.83). For most of the individual predictions, models with the 30 SNPs associated with AAP were highly confident compared with other models in the personalized AI ensemble. However, the combined prediction in the ensemble helped correct previously false predictions or provide more confidence to many of the correct predictions (Supplementary Fig. S.8, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). Figures 3B, C, E, and F depict the prediction scores against their true class as well as estimations of sensitivity, specificity, positive predictive value (PPV) and negative predictive value for the personalized AI ensemble model and test data set performance. We assessed how good the model was in the extremes of its score distribution. Patients at a high risk of AAP were identified by setting the prediction threshold of 0.8 in the trained model and test performance reported in Figures 3B and C, respectively. In the cross-validation, the sensitivity was 62% with a PPV of 37%, where 262 patients (both case/controls) had prediction threshold ≥ 0.8, whereof 96 cases are correctly predicted in the trained models (Fig. 3C). On the hold-out European validation data set (N = 100), a prediction threshold of 0.8 resulted in sensitivity of 52% and a PPV of 87% where 30 patients (both cases/controls) had a prediction score ≥ 0.8 whereof 26 AAP cases are correctly predicted (Fig. 3F). Validation of the personalized AI ensemble model on the non-European hold-out data set (N = 174) resulted in ROC-AUC: 0.73 (Fig. 3G). A prediction threshold of 0.8 resulted in sensitivity of 38% and a PPV of 50% of this group (Fig. 3I). The personalized AI ensemble model trained on the first AAP event cases only predicted second AAP following re-exposure to asparaginase with limited performance (ROC-AUC: 0.55, Fig. 3J).

Second AAP Following Re-exposure to Asparaginase

A separate model for re-exposure risk was trained as the phenotype between first AAP and second AAP patients may differ. A logistic regression predicted second AAP risk with ROC-AUC: 0.65, sensitivity: 0.62 and specificity: 0.79 using age, sex and previously validated pancreatitis SNPs

(Supplementary Table S.9, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The model was robust when comparing the true performance to random second AAP classification labels ($P < < 1e-6$, Supplementary Fig. S.10, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). The most important feature was rs12688220 (*CLDN2/MORC4*) that reduced ROC-AUC with 0.15 when being left out (Fig. 4).

Impact of Clinical Information Integrated With Genetic Information on AAP Prediction

Patients treated with the NOPHO ALL-2008 protocol (N = 892, whereof 77 developed AAP) had more clinical features available that were integrated in 1 hidden layer artificial neural network model with the genetic features for prediction of AAP (Supplementary Table S.11, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). Patients that developed AAP received fewer asparaginase dosages compared with controls (mean ± SD; cases: 7.3 ± 4.4, controls: 12.8 ± 3.3). A model with the 5 most predictive AAP SNPs (Table 2, binary allele encoding), age and sex, boosted ROC-AUC from 0.59 (age and sex only) to 0.64 (Supplementary Table S.12, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). Additional clinical features and information of asparaginase randomization group in the NOPHO ALL-2008 treatment protocol (either 2 or 6 weekly-treatment intervals) and the total number of asparaginase dosages per patient, boosted the models' performance to ROC-AUC: 0.86. The most important features were the asparaginase treatment randomization group of either 2 or 6 weeks treatment intervals and the total number of asparaginase dosages per patient. Exclusion of asparaginase treatment intensity and dosing features reduced ROC-AUC with 0.24 when estimating the leave-one-group-out feature importance (Supplementary Fig. S.13A, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474). To confirm the large impact of the treatment intensity and the total number of asparaginase dosages on AAP risk prediction, we trained a LASSO regression model, where the ROC-AUC was: 0.80. More frequent treatment intervals of asparaginase and fewer asparaginase dosages were important to predict AAP (Supplementary Fig. S.13B, Supplemental Digital Content 1, http://links.lww.com/JPHO/A474).

DISCUSSION

Asparaginase is an essential drug for ALL therapy, and truncation of therapy due to, for example, pancreatitis, thrombosis, hypersensitivity, or silent inactivation has in several studies been associated with an increased risk of relapse.[1,6,7] Thus, there is currently an unmet need to identify patients at high risk of such adverse events which could guide patient selection for future AAP prevention trials or to guide clinicians in deciding when re-exposure to asparaginase is likely to be safe.

We applied machine learning algorithms and integrated multiple SNPs given a clinical baseline model of age and sex (ROC-AUC: 0.62). Several methodologies were employed to identify predictive features from a pool of ~1.4M SNPs to provide a sufficiently strong model that could, if validated by other groups, be clinically applicable for identification of patients with very high risk of AAP based on germline predisposition (minimum 80%). Improvements from 6 candidate SNPs or previously

validated SNPs in adult pancreatitis drove performance up from ROC-AUC: 0.62 to 0.67 and with 30 previously associated AAP SNPs to ROC-AUC: 0.80 indicating that germline genetic profiling can significantly assist in the prediction of some patients at risk of AAP.

The 30 SNPs identified in Wolthers and colleagues provides the strongest power to detect true findings whereof rs13228878 and rs10273639 were validated in another cohort previously.[10] A limitation of the feature selection for the best performing model is that the 30 AAP-associated SNPs were identified through a GWAS on the same data set used in this study. Despite this AAP prediction model being validated on a held-out test data set (ROC-AUC: 0.84), validation on an external data set must be used before adaption of such machine learning models in clinic. The presented AAP models could predict AAP risk for non-European patients with slightly reduced performance (ROC-AUC: 0.72) compared with European patients.

It was not possible to obtain similar performance with other genetic variants identified in prior studies by Liu et al[4] or Abaji et al[9] of AAP in childhood ALL and these models only obtained ROC-AUC of 0.60 to 0.63. This is possibly due to the cohorts of patients being very different, both in asparaginase exposure and diagnostic criteria of AAP, or more likely reflected false positive findings as none of the those SNPs reached GWAS significance.[4,9]

Feature importance of these AAP prediction models showed that a combination of all features was required to achieve clinically useful performance as each feature had minor impact when being left out on the ROC-AUC. Across the different models, the most predictive SNPs were rs10273639 (*PRSS1/PRSS2*), rs10436957 (*CTRC*), rs13228878 (*PRSS1/PRSS2*), rs1505495 (*GALNTL6*), and rs4655107 (*EPHB2*). *PRSS1/PRSS2* and *CTRC* are expressed in pancreatic tissue.[21] The variants rs10273639 and rs13228878 are located in the *PRSS1/PRSS2* locus which encode trypsinogens that can be cleaved into trypsin to activate digestive enzymes prematurely leading to cases of AAP.[20] The minor allele of rs13228878 was previously found to reduce risk of AAP.[10] rs10436957 is annotated to *CTRC* which encodes the enzyme chymotrypsin C that helps regulate the activation and degradation of trypsinogens.[10,20] *GALNTL6* encodes polypeptide N-acetylgalactosaminyl-transferase-like 6, which is a transferase-like enzyme involved in the posttranslational process of O-linked glycosylation responsible for transferring N-acetylgalactosamine to an exposed serine or threonine.[23,24] *EPHB2* encodes the transmembrane EPH receptor B2, which is part of the largest family of tyrosine kinase receptors and capable of bidirectional signaling through binding with ephrin ligands on neighboring cells. This signaling is involved in developmental processes such as cell and axon growth, as well as involved in cancers.[25,26]

As different models captured different learnings on AAP risk, the individual AAP predictions were improved by an ensemble model approach. This helped correctly re-classify patients as well as increased the confidence of the true prediction class. Prediction of AAP cases with high confidence can motivate increased monitoring of patients or supportive care. The clinical utility of these models should be evaluated in future pre-emptive trials. Wrong predictions, that is, non-AAP risk patients predicted to be at risk will not experience any negative repercussions in the clinic. Classification of AAP cases with minimum 80% confidence was applied when evaluating the model's predictive performance. This confidence threshold resulted in 62% of the true cases to be correctly predicted, with false positives of 63% on the cross-validated test data sets and limited to only 13% false positives on the European hold-out validation data set with 52% of the AAP cases being correctly identified by the model. It should be noted that the hold-out validation data set had an artificially high incidence of AAP (50%) as it was sampled to contain equal cases and controls. The hold-out validation data set was sampled and set aside before training of the models.

Prediction of patients who are likely to tolerate re-exposure of PegAsp after their first AAP episode is currently one of the most critical questions associated with asparaginase therapy, since truncation of therapy has been associated with an increased risk of relapse.[1,6,7] However, neither the AAP phenotype, including severity of the first AAP, or the age of patients nor SNPs are sufficiently strong risk factors to guide the re-exposure decision. Currently, consensus guidelines do not exist, and decisions to re-expose a patient will thus reflect physicians' attitudes and gut feeling, and the balance between anticipated risks of a second AAP versus leukemic relapse.

Thus, separate models were trained on re-exposure patients resulting in the most predictive model being a logistic regression with ROC-AUC: 0.65. rs12688220 (*CLDN2/MORC4*) was most predictive of a second AAP, which is previously associated with adult pancreatitis.[18,19]

The present study has limited power for prediction of second AAP, but the PTWG is currently collecting very detailed data on >100 patients re-exposed with PegAsp after AAP of whom ~40% are expected to develop AAP. Since a second AAP episode usually occurs after several doses of PegAsp, future developments of this tool could increase the number of patients that will be re-exposed to asparaginase.

The potential clinical utility of the models should be evaluated in the light of predictive performance as well as their interpretability of features which is an important challenge to address for adaptation into clinic.[27] The machine learning models learn patterns from data which can be complex and nonlinear and achieve good predictive performance, while the feature importance—especially with complex feature interaction—at the individual patient level can be harder to identify. Since asparaginase is an essential drug in the treatment of childhood ALL, the model should primarily identify patients with a very high risk of developing AAP, which could guide patient selection for future AAP prevention trials, and potentially also patient selection for asparaginase re-exposure. On the path towards clinical translation of an AAP prediction model, it is also important to know the time to AAP or information on additional clinical events such as severity or necrotizing at the first AAP event for prediction of second AAP. Other clinical features were only reported for AAP cases and not controls, and furthermore were inconsistently recorded and thus had a high level of missing values. This motivates the importance of more rigorous data collection to gain further insights of clinical features for prediction. With the available data, the main scope of this study was to identify predictive genetic predisposition to AAP and second AAP risk—however more rigorous clinical data collection across collaborative cohorts in the future would offer the opportunity to build richer models that can integrate a wider clinical context with genetics in building predictors.

For a subset of the patients treated with the NOPHO ALL-2008 protocol, more frequent treatment intervals of asparaginase and fewer asparaginase dosages was identified as the most important features to predict AAP. The fewer asparaginase dosages reported for the AAP cases reflects truncation of treatment as controls would receive further dosages. A suggested follow-up study is integrating the number of asparaginase doses with the identified predictive SNPs and account for the time to event to determine the timing of a patient's risk of AAP.

In conclusion, this study supports the impact of host-genome variants on risk of AAP and exemplifies strategies for applying predictive modeling on other severe acute toxicities to ALL therapy.

REFERENCES

1. Pieters R, Hunger SP, Boos J, et al. L-asparaginase treatment in acute lymphoblastic leukemia: a focus on Erwinia asparaginase. *Cancer.* 2011;117:238–249.
2. Müller HJ, Boos J. Use of L-asparaginase in childhood ALL. *Crit Rev Oncol Hematol.* 1998;28:97–113.
3. Hijiya N, van der Sluis IM. Asparaginase-associated toxicity in children with acute lymphoblastic leukemia. *Leuk Lymphoma.* 2016;57:748–757.
4. Liu C, Yang W, Devidas M, et al. Clinical and genetic risk factors for acute pancreatitis in patients with acute lymphoblastic leukemia. *J Clin Oncol.* 2016;34:2133–2140.
5. Rank CU, Wolthers BO, Grell K, et al. Asparaginase-associated pancreatitis in acute lymphoblastic leukemia: results from the NOPHO ALL2008 treatment of patients 1-45 years of age. *J Clin Oncol.* 2020;38:145–154.
6. Gupta S, Wang C, Raetz EA, et al. Impact of asparaginase discontinuation on outcome in childhood ALL: a report from the Children's Oncology Group (COG). *J Clin Oncol.* 2019;37 (suppl):10005.
7. Gottschalk Højfeldt S, Grell K, Abrahamsson J, et al. Relapse risk following truncation of pegylated-asparaginase in childhood acute lymphoblastic leukemia. *Blood.* 2021;137:2373–2382.
8. Wolthers BO, Frandsen TL, Baruchel A, et al. Asparaginase-associated pancreatitis in childhood acute lymphoblastic leukaemia: an observational Ponte di Legno Toxicity Working Group study. *Lancet Oncol.* 2017;18:1238–1248.
9. Abaji R, Gagné V, Xu CJ, et al. Whole-exome sequencing identified genetic risk factors for asparaginase-related complications in childhood ALL patients. *Oncotarget.* 2017;8:43752–43767.
10. Wolthers BO, Frandsen TL, Patel CJ, et al. Trypsin-encoding PRSS1-PRSS2 variations influence the risk of asparaginase-associated pancreatitis in children with acute lymphoblastic leukemia: a ponte di legno toxicity working group report. *Haematologica.* 2019;104:556–563.
11. Wesołowska-Andersen A, Borst L, Dalgaard MD, et al. Genomic profiling of thousands of candidate polymorphisms predicts risk of relapse in 778 Danish and German childhood acute lymphoblastic leukemia patients. *Leukemia.* 2015;29:297–303.
12. Pan L, Liu G, Lin F, et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Sci Rep.* 2017;7:1–9.
13. Albertsen BK, Grell K, Abrahamsson J, et al. Intermittent versus continuous PEG-asparaginase to reduce asparaginase-associated toxicities: a NOPHO ALL2008 randomized study. *J Clin Oncol.* 2019;37:1638–1646.
14. Schmiegelow K, Attarbaschi A, Barzilai S, et al. Consensus definitions of 14 severe acute toxic effects for childhood lymphoblastic leukaemia treatment: a Delphi consensus. *Lancet Oncol.* 2016;17:e231–e239.
15. Python Software Foundation. Python, version 3.6.8. 2018. Available at: www.python.org/. Accessed January 6, 2020.
16. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.
17. Whitcomb DC, Larusch J, Krasinskas AM, et al. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. *Nat Genet.* 2012;44:1349–1354.
18. Derikx MH, Kovacs P, Scholz M, et al. Polymorphisms at PRSS1-PRSS2 and CLDN2-MORC4 loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. *Gut.* 2015;64:1426–1433.
19. Rosendahl J, Kirsten H, Hegyi E, et al. Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut.* 2018;67:1855–1863.
20. Zator Z, Whitcomb DC. Insights into the genetic risk factors for the development of pancreatic disease. *Therap Adv Gastroenterol.* 2017;10:323–336.
21. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45:580–585.
22. Yates A, Beal K, Keenan S, et al. The Ensembl REST API: Ensembl Data for Any Language. *Bioinformatics.* 2015;31:143–145.
23. Van Den SP, Rudd PM, Dwek RA, et al. Concepts and Principles of O-Linked Glycosylation. *Crit Rev Biochem Mol Biol.* 1998;33:151–208.
24. UniProtKB-Q49A17 (GLTL6_HUMAN). Integrated into UniProtKB/Swiss-Prot: March 18, 2008. 2010. Available at: https://www.uniprot.org/uniprot/Q49A17. Accessed November 22, 2019.
25. Himanen J, Chumley MJ, Lackmann M, et al. Repelling class discrimination: ephrin-A5 binds to and activates EphB2 receptor signaling. *Nat Neurosci.* 2004;7:501–509.
26. UniProtKB-P29323 (EPHB2_HUMAN). Integrated into UniProtKB/Swiss-Prot: December 1, 1992. 2005. Available at: www.uniprot.org/uniprot/P29323. Accessed November 22, 2019.
27. Prosperi M, Min JS, Bian J, et al. Big data hurdles in precision medicine and precision public health. *BMC Med Inform Decis Mak.* 2018;18:1–15.