TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Title page 1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Abstract page 2
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Introduction pages 5 and 6
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	Introduction pages 6 and 7
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	"PaCaOmics patient's cohort and PDX" subsection page 8; "Ethic statements" subsection pages 8-9; "RNA-seq analysis and gene selection" subsection pages 9-10; "Data Sharing Statement" section pages 27-28; "Validation on public datasets" subsection pages 10-11; Supplementary table S1 "Survival data of classical subtype patients of PaCaOmics cohort" sheet; "References" section (number 10, 12, 28-32) pages 30-32
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	"PaCaOmics patient's cohort and PDX" subsection page 8; "Ethic statements" subsection pages 8-9; "Validation on public datasets" subsection pages 10-11; "Data Sharing Statement" section pages 27-28; "Survival analyses" subsection page 11; Supplementary table S1 "Survival data of classical subtype patients of PaCaOmics cohort" sheet "References" section (number 10, 12, 28-32) pages 30-32
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	"PaCaOmics patient's cohort and PDX" subsection page 8; "Ethic statements" subsection pages 8-9; "Validation on public datasets" subsection pages 10-11; "Data Sharing Statement" section pages 27-28 "References" section (number 10, 12, 28-32) pages 30-32
	5b	D;V	Describe eligibility criteria for participants.	"PaCaOmics patient's cohort and PDX" subsection page 8; "Data Sharing Statement" section pages 27-28 "References" section (number 10, 12, 28-32) pages 30-32
	5c	D;V	Give details of treatments received, if relevant.	Not applicable for this study
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Not applicable for this study. We have highlighted a new PDAC stratification based on glycosyltransferase expression profile allowing to distinguish different groups of patients with specific molecular profiles and distinct clinical features. No prediction algorithm has been developed for a clinical use in this study.
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	Not applicable for this study. We have highlighted a new PDAC stratification based on glycosyltransferase expression profile allowing to distinguish different groups of patients with specific molecular profiles and distinct clinical features. No prediction algorithm has been developed for a clinical use in this study.
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	"RNA-seq analysis and gene selection" subsection, pages 9-10; "Hierarchical Clustering on Principal Component (HCPC) analysis and glyco-signature definition" subsection, page 10; "Validation on public datasets" subsection, pages 10-11;

				Supplementary Table S1 "Statistical data of HCPC analysis on PaCaOmics patient's cohort" sheet
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	The prognostic value of the glyco-signature was validated on independent cohorts by performing HCPC analyses (allowing to identify systematically three clusters with significant differences in OS), without any classification based on the outcome prior to these analyses. "Validation of the glyco-signature prognostic value on independent cohorts" subsection page 17
Sample size	8	D;V	Explain how the study size was arrived at.	Not applicable for this study. This work is a data mining of PaCaOmics data and public datasets previously published; "References" section (number 10, 12, 28-32) pages 30-32.
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	Management of missing data (i.e. transcriptomic and survival data) were specified in the following subsections: "RNA-seq analysis and gene selection" subsection, pages 9-10; "Validation on public datasets" subsection, pages 10-11 "Survival analyses" subsection page 11
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	"RNA-seq analysis and gene selection" subsection, pages 9-10; "Hierarchical Clustering on Principal Component (HCPC) analysis and glyco-signature definition" subsection, page 10; Supplementary Table S1 "Statistical data of HCPC analysis on PaCaOmics patient's cohort" sheet
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	"Hierarchical Clustering on Principal Component (HCPC) analysis and glyco-signature definition" subsection, page 10; "Comparison with previously established classification" subsection, page 11; "Survival analyses" subsection, pages 11-12; "Statistical analysis" subsection, pages 12-13; Supplementary Table S1 "Statistical data of HCPC analysis on PaCaOmics patient's cohort" sheet
	10c	V	For validation, describe how the predictions were calculated.	Not applicable for this study. This study has used public datasets previously described to stratify PDAC through glycosyltransferase expression profile; "References" section (number 10, 12) page 30; "Validation on public datasets" subsection page 10-11
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	The glyco-signature prognostic value was validated by performing HCPC analyses on independent datasets followed by survival analyses: "Validation on public datasets" subsection page 10-11 and "Survival analyses" subsection page 11. Its performance was also compared with established PDAC classification: "Comparison with previously established classification" subsection page 11 and "Statistical analyses" subsection page 12.
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	"Validation on public datasets" subsection pages 10-11; "Survival analyses" subsection, pages 11-12; "Statistical analysis" subsection, pages 12-13
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	Not applicable for this study. This study has used PaCaOmics data and public datasets previously described to stratify PDAC through glycosyltransferase expression profile; "References" section (number 10, 12, 28-32) pages 30-32.
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	The glyco-signature was identified on RNA-seq data of PDX tumour (epithelial compartment) from resected and biopsied tumors, whereas the validation was performed on RNA-seq and microarrays datasets including resected whole tumour tissues: "Validation of the glyco-

				signature prognostic value on independent cohorts" subsection page 17-18.
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Not applicable for this study. This study has used RNA-seq data of PaCaOmics patient's cohort (flow of participants described in reference number 30 page 32) and public datasets to stratify PDAC through glycosyltransferase expression profile
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Figure 1; Supplementary Figure S1 and S4 "Expression profiles of GT genes predict overall survival of PDAC patients" subsection, page 14; "Clinical features of patients and their PDAC molecular profiles" subsection, pages 15, 16; "Validation of the glyco-signature prognostic value on independent cohorts" subsection, page 17
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	This study has used public datasets previously described to stratify PDAC through glycosyltransferase expression profile "References" section (number 10, 12,) page 30. Comparison with predictors and outcome in Supplementary Figure S6, S7; "Validation of the glyco-signature prognostic value on independent cohorts" subsection, pages 17, 18;
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	Figure 1; Supplementary Figure S2; Supplementary table S1 "Survival data of classical subtype patients of PaCaOmics cohort" sheet "Expression profiles of GT genes predict overall survival of PDAC patients" subsection, page 14; "Identification of GT genes as prognostic markers" subsection, pages 14-15; "Clinical features of patients and their PDAC molecular profiles" subsection, pages 15, 16
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	Not applicable for this study. This study has identified a glycosyl-transferase gene signature to highlight a new PDAC stratification
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Not applicable for this study. No prediction algorithm has been developed for an individual in clinical use. We have highlighted a new PDAC stratification based on glycosyltransferase expression profile allowing to distinguish different groups of patients with specific molecular profiles and distinct clinical features.
	15b	D	Explain how to the use the prediction model.	Not applicable for this study. No prediction algorithm has been developed for an individual in clinical use. We have highlighted a new PDAC stratification based on glycosyltransferase expression profile allowing to distinguish different groups of patients with specific molecular profiles and distinct clinical features.
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	"Expression profiles of GT genes predict overall survival of PDAC patients" subsection, page 14; "Clinical features of patients and their PDAC molecular profiles" subsection, pages 15, 16, 17; "Validation of the glyco-signature prognostic value on independent cohorts" subsection, pages 17, 18; Figures 1, 2, 3, 4; Supplementary Figures S1, S2, S3, S6, S7.
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	Not applicable for this study. No prediction algorithm has been developed for an individual in clinical use.
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	Page 21 lines 12-14 and 20-22; Page 22 lines 10-13; Page 23 lines 22-25; Page 24 lines 18-22;

				Page 26 lines 8-9
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	Pages 21-22
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	Page 22 1st and 2nd paragraph and Page 23; Page 24; Page 25-26
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	Pages 22 lines10-13; Page 23 lines 22-25; Pages 25 lines 19-20; Page 26 lines 6-9 and 14-16
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	"Data Sharing Statement" section pages 27-28; Supplementary tables S1, S2 and S3
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	"Acknowledgements" section page 27; "Role of the funding source" subsection, page 13.

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.