nature portfolio

Corresponding author(s): Prof Ajay Goel, PhD.

Last updated by author(s): Jul 31, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All statistical analyses were conducted within the R environment (version 4.4). Data ingestion and tidyverse-based preprocessing were handled using 'tidyverse' (v2.0) and 'readxl' (v1.5). Cohort partitioning was performed using 'splitTools' (v1.0) for balanced stratification. Machine learning processes were based on 'xgboost' (v3.4), 'caret' (v7.0), and 'DAAG' (v1.25). ROC curve analysis, AUROC estimation, and diagnostic threshold completed using 'pROC' (v1.18) and 'cutpointr' (v1.2). Survival analyses, time-to-event modeling, and longitudinal data tracking were conducted using 'survival' (v3.8) and 'SwimmeR' (v0.14). Restricted cubic splines were modeled with 'plotRCS' (v0.1). High-resolution plots and figures were generated and exported using 'ggplot2' (v4.0), 'ggpubr' (v0.6), and 'export' (v0.3).
Data analysis	R. The code utilized for XGBoost machine learning data analysis and visualization can be accessed at: https://zenodo.org/records/15041773

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data collected for the study, including de-identified participant data, is available at https://doi.org/10.17605/OSF.IO/V5MTA

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Biological sex was determined based on self-report by participants at the time of enrollment. Information on gender was not collected. During the training and validation phases, the cohort was balanced by biological sex to ensure robust biomarker discovery. Furthermore, sub-analyses were performed to evaluate the diagnostic performance of the miRNA signature across biological sexes, as detailed in the Results
Reporting on race, ethnicity, or other socially relevant groupings	Reported in Extended Data Table 1
Population characteristics	Reported in Extended Data Table 1 and Extended Data Table 2
Recruitment	Reported in the "statistics and reproducibility" section within the methods. Consecutive enrollment of patients and contorls across multiple centers.
Ethics oversight	Reported in the "Ethics approval" section within the methods. The prospective collection of the patient datasets in each cohort was approved by the institutional review board (IRB) at each institution: the Translational Genomics Research Institute (TGen) IRB (23-000161), Hoag Family Center Institute IRB (20181575), Medical College of Wisconsin (MCW) IRB (PRO#12151), HonorHealth Research Institute IRB (1618742), Nagoya University Graduate School of Medicine IRB (2014-0042), Mie University Graduate School of Medicine IRB (H2022-185), Samsung Medical Center IRB (2020-07-012), Ochsner Clinic Foundation IRB (00000152), Asan Medical Center IRB (2016-1077 and 2025-0259), Allegheny Health Network (AHN) IRB (2020-258), Baylor Scott & White Research Institute IRB (017-199). The study protocol for the collection of prospective cohorts in the Pancreatic Cancer Detection Consortium (PCDC) was registered at clinicaltrials.gov (Study Registration: NCT06271291).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N=1785, as depicted in Figure 1A. No statistical method was used to predetermine sample size, as reported in the methods. Instead, sample sizes were chosen based on the maximal availability of well-characterized clinical specimens within our institutional biobanks and international collaborative cohorts. These sample sizes are scientifically sufficient because they match or exceed the cohort volumes typical for high-throughput biomarker discovery and validation studies in this field. Furthermore, the robust statistical significance, highly constrained confidence intervals, and strong effect sizes achieved across our independent validation cohorts retrospectively confirm that the study was adequately powered to yield reproducible and statistically sound conclusions.
Data exclusions	N=28 due to suboptimal RNA quality, as reported in Figure 1A
Replication	N=1757 across three independent clinical cohorts. N=440 for the testing cohort. N=137 for the pre-/post-surgery cohort. N=168 for the cross-reactivity cohort. Results were replicable (across all the cohorts included in this study)
Randomization	N=1012 randomized 70:30 to the training:validation cohorts, as stated in the methods "The participants were randomly assigned to the training and validation cohorts in a 7:3 ratio, ensuring balance by biological sex, country, TNM stage, and ethnicity using the "splitTools" package in R"

Blinding	Analyses in the replication cohorts (N=482) were conducted after model and threshold locking. See: "Because all statistical parameters, computational models, and classification thresholds were fully frozen and locked at the conclusion of the initial training and validation phases, investigators were not blinded to allocation during experiments and outcome assessment"

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Study Registration: NCT06388967
Study protocol	Registered at Clinical Trial.gov
Data collection	Reported in the second paragraph of the "Statistics and reproducibility". See :"For miRNA signature training and validation, we obtained plasma samples from 635 patients with PDAC and 377 NDCs. All the 1012 participants in both the training and validation cohorts were consecutively enrolled at TGen (Phoenix, Arizona, USA), Hoag (Newport Beach, California, USA), MCW (Milwaukee, Wisconsin, USA), HonorHealth (Scottsdale, Arizona, USA), AHN (Pittsburgh, PA, USA), Baylor Scott & White Research Institute (Dallas, TX, USA), Nagoya University Graduate School of Medicine (Nagoya, Japan), Mie University Graduate School of Medicine (Mie, Japan), Samsung Medical Center (Seoul, Korea), and Asan Medical Institute (Seoul, Korea) between 2021 and 2023. No statistical method was used to predetermine sample size. The participants were randomly assigned to the training and validation cohorts in a 7:3 ratio, ensuring balance by biological sex, country, TNM stage, and ethnicity using the "splitTools" package in R. The miNRA training phase included 434 patients with PDAC and 273 NDCs, while the validation cohort comprised plasma specimens from 201 patients with PDAC and 104 NDCs. Subsequently, we tested the performance of the final miRNA signature in an independent cohort consisting of 90 patients with PDAC, 14 HGD, low risk controls (serous cysts and NDCs, n=93), and high risk controls (n=243, 48 PFC, 45 HPC, 33 FPC/HPC with low-risk IPMN, 30 low-risk IPMN, 34 CP and 53 high-risk cyst [40 IPMN and 13 MCN]), who were enrolled at the Ochsner Clinic Foundation (New Orleans, Louisiana, IRB #00000152) and IRCCS San Raffaele Hospital (Milan, Italy, IRB #BIOGASTRO/2011) between 2020 and 2025. The detailed clinical characteristics of training, validation, and testing cohorts are provided in Extended Data Table 1-2"
Outcomes	Reported in the "Definitions, inclusion and exclusion criteria, and study endpoints" paragraph of the methods. Please see: "The primary outcome of this study was to assess the sensitivity for detecting patients with PDAC. Secondary outcome measures included sensitivity for stage I-II PDAC. The exploratory analyses included sensitivity for detecting PDAC across different locations, ethnicities, and countries, as well as specificity during treatment"

Plants

Seed stocks	No plants were used in this study
Novel plant genotypes	No plants were used in this study
Authentication	No plants were used in this study