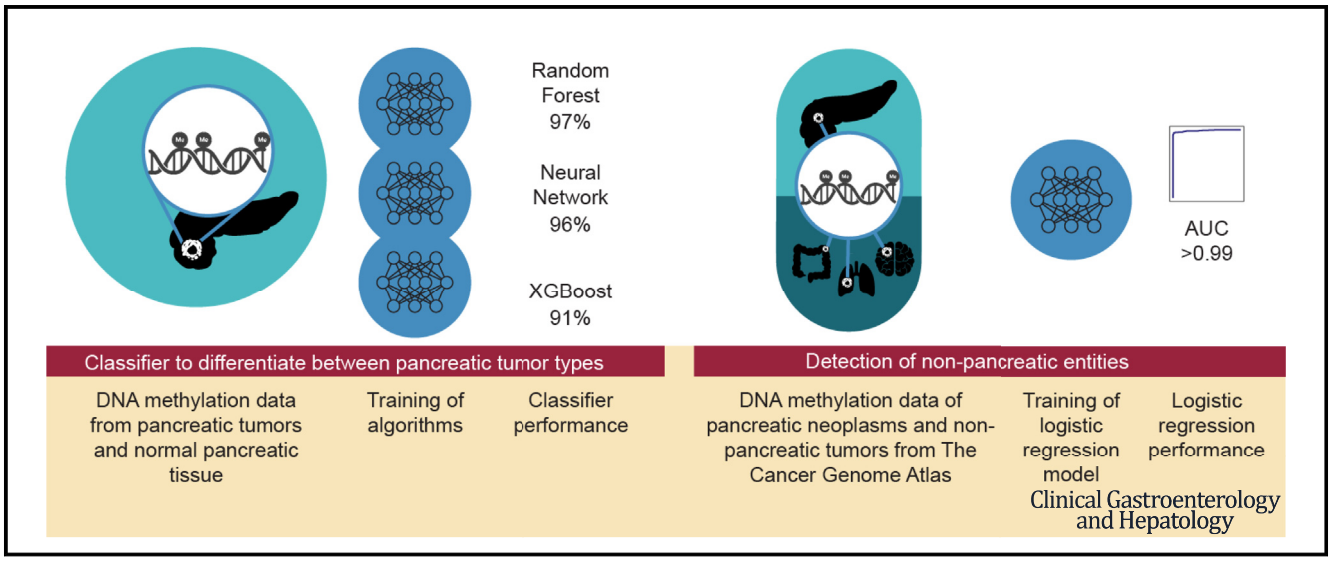
Clinical Gastroenterology and Hepatology 2024;22:1245–1254

PANCREATOBILIARY

DNA Methylation Profiling Enables Accurate Classification of Nonductal Primary Pancreatic Neoplasms

Anna Vera D. Verschuur,[1] Wenzel M. Hackeng,[1] Florine Westerbeke,[1] Jamal K. Benhamida,[2] Olca Basturk,[2] Pier Selenica,[2] G. Mihaela Raicu,[3] I. Quintus Molenaar,[3,4] Hjalmar C. van Santvoort,[4] Lois A. Daamen,[4] David S. Klimstra,[5] Shinichi Yachida,[6] Claudio Luchini,[7] Aatur D. Singhi,[8] Christoph Geisenberger,[9] and Lodewijk A. A. Brosens[1,10]

[1]Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [2]Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, New York; [3]Department of Pathology, St Antonius Hospital and Pathology DNA, Nieuwegein, The Netherlands; [4]Department of Surgery, Regional Academic Cancer Center Utrecht, University Medical Center Utrecht Cancer Center and St. Antonius Hospital, Nieuwegein, The Netherlands; [5]Paige.AI, New York, New York; [6]Department of Cancer Genome Informatics, Graduate School of Medicine, Osaka University, Osaka, Japan; [7]Department of Diagnostics and Public Health, Section of Pathology, University of Verona, Verona, Italy; [8]Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; [9]Institute of Pathology, LMU University, Munich, Germany; and [10]Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands



BACKGROUND & AIMS: **Cytologic and histopathologic diagnosis of non-ductal pancreatic neoplasms can be challenging in daily clinical practice, whereas it is crucial for therapy and prognosis. The cancer methylome is successfully used as a diagnostic tool in other cancer entities. Here, we investigate if methylation profiling can improve the diagnostic work-up of pancreatic neoplasms.**

METHODS: **DNA methylation data were obtained for 301 primary tumors spanning 6 primary pancreatic neoplasms and 20 normal pancreas controls. Neural Network, Random Forest, and extreme gradient boosting machine learning models were trained to distinguish between tumor types. Methylation data of 29 nonpancreatic neoplasms (n = 3708) were used to develop an algorithm capable of detecting neoplasms of non-pancreatic origin.**

Abbreviations used in this paper: ACC, acinar cell carcinoma; FNA, fine-needle aspirate; MiNEN, mixed neuroendocrine–nonneuroendocrine neoplasm; NN, neural network; PB, pancreatoblastoma; PDAC, pancreatic ductal adenocarcinoma; PNEC, pancreatic neuroendocrine carcinoma; PNET, pancreatic neuroendocrine tumor; RF, random forest; SPN, solid pseudopapillary neoplasm; UMAP, uniform manifold approximation and projection; XGB, extreme gradient boosting.

Most current article

1542-3565
https://doi.org/10.1016/j.cgh.2024.02.007

RESULTS: **After benchmarking 3 state-of-the-art machine learning models, the random forest model emerged as the best classifier with 96.9% accuracy. All classifications received a probability score reflecting the confidence of the prediction. Increasing the score threshold improved the random forest classifier performance up to 100% with 87% of samples with scores surpassing the cutoff. Using a logistic regression model, detection of nonpancreatic neoplasms achieved an area under the curve of >0.99. Analysis of biopsy specimens showed concordant classification with their paired resection sample.**

CONCLUSIONS: **Pancreatic neoplasms can be classified with high accuracy based on DNA methylation signatures. Additionally, non-pancreatic neoplasms are identified with near perfect precision. In summary, methylation profiling can serve as a valuable adjunct in the diagnosis of pancreatic neoplasms with minimal risk for misdiagnosis, even in the pre-operative setting.**

Keywords: Pancreatic Neuroendocrine Tumor; Acinar Cell Carcinoma; Solid Pseudopapillary Neoplasms; Pancreatoblastoma; Pancreatic Ductal Adenocarcinoma; DNA Methylation; Tumor Classification.

Around 90% of pancreatic neoplasms are ductal adenocarcinomas (PDACs); the remainder is of non-ductal origin, including pancreatic neuroendocrine neoplasms (pancreatic neuroendocrine tumor [PNET] and pancreatic neuroendocrine carcinoma [PNEC]), acinar cell carcinoma (ACC), solid pseudopapillary neoplasms (SPN), and pancreatoblastoma (PB), accounting for 5%, 2%, 1%, and <1% of cases, respectively.[1] These tumors differ significantly in biology and clinical outcomes. For instance, SPNs rarely metastasize and have a 95% 5-year survival rate,[2] whereas PNETs, ACCs, and PBs show 5-year survival rates of 38%–59%.[3–5] Additionally, these tumors require different treatment. ACCs are typically treated by a combination of surgical resection, chemotherapy, and radiotherapy[6]; SPNs are cured by surgical resection[7]; and PNET management varies from resection to active surveillance depending on such factors as hormone production, tumor size, grade, and stage.[8,9] Hence, accurate diagnosis is crucial for therapeutic decision making, but ACCs, SPNs, and PNETs can present diagnostic challenges because of overlapping cytologic, histologic, and immunohistochemical features, particularly when dealing with limited tissue samples, such as preoperative fine-needle aspirates (FNAs) or fine-needle biopsies (FNBs).[1,10–12]

To improve diagnosis, there is a growing trend in using whole genome methylation-based diagnostic models for tumor classification,[13–17] with the World Health Organization recommending its routine application in central nervous system tumors.[18] DNA methylation, a covalent modification of cytosine residues regulating gene expression, is known for its high cell-type specificity, making it effective in differentiating between tumor types.[17,19,20] Previous studies identified the origin of different gastrointestinal (neuroendocrine) tumors based on methylation profiles, including those originating in the pancreas, indicating potential utility in non-ductal pancreatic neoplasms.[21–23]

This study aimed to facilitate and improve preoperative and postoperative diagnosis of non-ductal pancreatic neoplasms, by developing a machine learning classifier on a well-curated collection of pancreatic neoplasms. Additionally, we aimed to make the classifier capable of detecting non-pancreatic entities.

Materials and Methods

Study Overview

Methylation data from 20 normal pancreatic tissues and 301 major primary pancreatic tumor types were collected to develop a classification algorithm (Table 1).[22–28] Samples were randomly split into training and test cohorts stratified by tumor type (50:50 ratio; Figure 3*A*). Neural Network (NN), Random Forest (RF), and extreme gradient boosting (XGB) models were trained on the training cohort and evaluated in the test cohort (Figure 1*A*). Considering that the models are specifically trained to classify only the tumor types present in their training data, we enhanced the model with the ability to identify neoplasms of non-pancreatic origin (eg, tumors types it is not trained on). To do so, a logistic regression model using the pancreatic dataset (n = 321) supplemented with non-pancreatic neoplasms (n = 3708) was trained based on the RF output (Figure 1*B*). Furthermore, we performed 3 additional analyses with (1) mixed neuroendocrine-non-neuroendocrine neoplasm (MiNEN, n = 7), (2) matched primary and metastatic cases (n = 5), and (3) matched FNA and resection specimens (n = 4). Details on patients and samples, DNA extraction and whole genome DNA methylation analysis, preprocessing, quality control and batch effect, unsupervised analysis, classifier development and validation, and additional analysis are described in Supplementary Methods, Supplementary Table 1, and Supplementary Figure 1.

Code Availability

All analyses were performed in RStudio version 2023.03.1 based on the statistical language R version

4.3.0, using the tidyverse software collection for data wrangling and ggplot for visualization. Code used for the generation of results and figures can be accessed at https://github.com/averschuur/pancreas.

Results

Pancreatic Dataset

The entire study cohort comprises 301 primary pancreatic tumors including 124 PNETs, 13 PNECs, 46 ACCs, 20 SPNs, 16 PBs, 82 PDACs, and 20 normal pancreas controls (Table 1). Dimensionality reduction with uniform manifold approximation and projection (UMAP) showed clustering predominantly by tumor type (Figure 2). Analysis of average methylation distribution, bisulfite conversion efficiency, and tumor purities based on ESTIMATE and ABSOLUTE scores across tumor types revealed lower purity in PDACs and lower genome-wide methylation in PBs (Supplementary Figure 2*A-H*). Projecting different sample preservation types onto the UMAP demonstrated no difference in clustering between fresh-frozen and formalin-fixed, paraffin embedded tissue (Supplementary Figure 3). Batch effect analyses revealed small but consistent differences regarding beta values distribution and conversion scores across tumor types (Supplementary Figure 4*A-T*).

Classifier Development and Performance

When assessing the classifiers without any cutoffs and assigning each sample to the category with the highest probability score, classifiers achieved accuracies of 96.9% (RF), 91.3% (XGB), and 95.7% (NN) in the test cohort (Figure 3*B*). The RF model consistently outperformed other models across all tumor types, except for SPN (Figure 3*C*). The accuracies of the models were not significantly different ($P = .40$; 1-way analysis of variance).

Applying a threshold with minimum classification scores of 0.45 (RF), 0.95 (XGB), and 0.80 (NN), improved accuracies to 100%, 99%, and 97%, respectively (Figure 3*D-F*). With the same threshold, classification was possible for 87% (RF), 62% (XGB), and 99% (NN) of samples. Adjusting the RF score cutoff to 0.30 resulted in

What You Need to Know

Background

Non-ductal pancreatic neoplasms have overlapping cytological, histological and immunohistochemical features, which often makes correct histopathological diagnosis challenging, while it is crucial for prognosis and therapeutic decision making.

Findings

Methylation profiling enables highly accurate classification of pancreatic neoplasms and near perfect identification of non-pancreatic entities, and can also be applied in fine-needle-aspiration / biopsy specimens, providing a valuable adjunct for pre-operative diagnostics.

Implications for patient care

This DNA methylation-based classifier can reduce (pre-operative) misdiagnoses and thereby assist in pre-operative clinical decision making, enhancing therapy accuracy, and improving patient selection for studies. Prospective studies are needed to further validate and potentially extend the model.

Table 1. Overview of Pancreatic Dataset Including Source, Dataset Number, Tissue Type, and Array Type

				Tumor type							
Source	Dataset	Tissue type	Array	PDAC	PNET	PNEC	ACC	SPN	PB	NORM	Total
Benhamida et al[22]	NA	FFPE	EPIC	0	0	0	7	0	16	0	23
Jäkel et al[23]	EGAS00001002533	FFPE	450K	0	17	0	37	0	0	6	60
Chan et al[24]	GSE117852	FF	450K	0	23	0	0	0	0	0	23
Di Domenico et al[25]	EMTAB-7924	FFPE	450K	0	46	0	0	0	0	14	60
Endo et al[26]	GSE155353	FF	450K	82	0	0	0	0	0	0	82
Selenica et al[27]	NA	FFPE	EPIC	0	0	0	0	13	0	0	13
Yachida et al[28]	JGAS000359	Unknown	EPIC	0	30	13	0	0	0	0	43
Local cohort	EGAS00001004878	FFPE	EPIC	0	8	0	2	7	0	0	17
				82	124	13	46	20	16	20	321

ACC, acinar cell carcinoma; FF, fresh frozen; FFPE, formalin-fixed, paraffin-embedded; NA, not applicable; NORM, normal pancreatic tissue; PB, pancreatoblastoma; PDAC, pancreatic ductal adenocarcinoma; PNEC, pancreatic neuroendocrine carcinoma; PNET, pancreatic neuroendocrine tumor; SPN, solid pseudopapillary neoplasm.

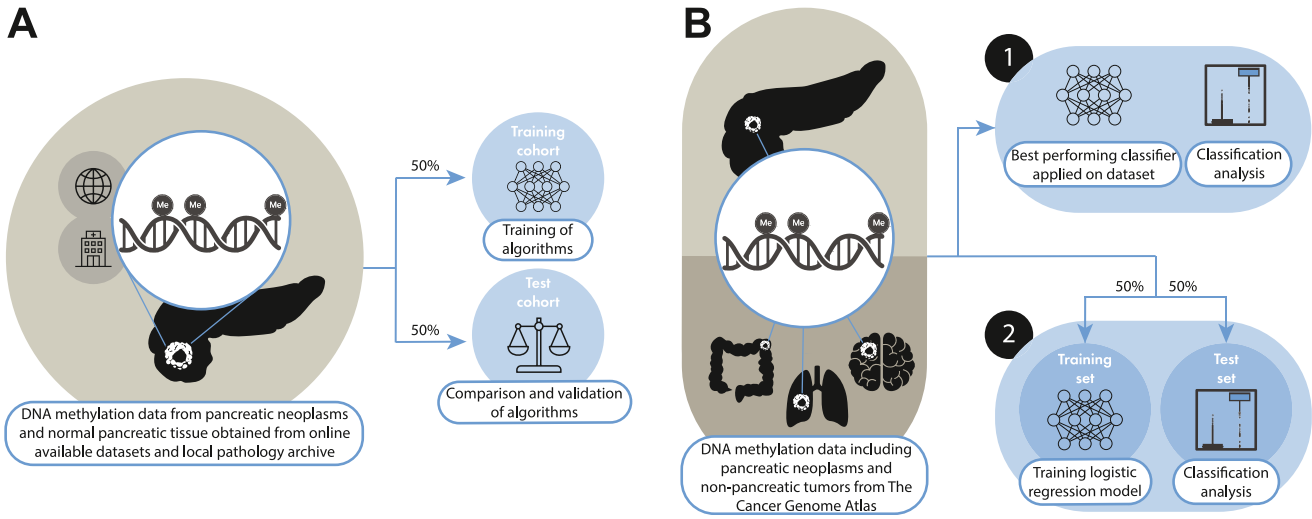


Figure 1. Study methods for classifier development and detection of non-pancreatic entities. A cohort of pancreatic neoplasms and pancreatic controls, comprising publicly available and in-house DNA methylation data, was randomly divided into training and test cohort stratified by tumor type (50:50 ratio). Neural Network, Random Forest, and extreme gradient boosting models were trained on the training cohort and evaluated in the test cohort (*A*). A cohort of pancreatic and non-pancreatic neoplasms was used to (1) evaluate the Random Forest's performance in detecting non-pancreatic tumors and to (2) optimize nonpancreatic tumor detection using logistic regression (*B*).

97.5% accuracy while retaining 98.8% of samples, comparable with the maximum accuracy of NN (97.4%) while retaining a slightly lower proportion of samples (96.9%). Taken together, the RF model without any cutoffs demonstrated the highest accuracy in the highest proportion of samples.

Furthermore, the confusion matrices of the RF, XGB, and NN classifiers give an overview of misclassified cases per tumor type (Figure 3*G-I*). Projecting RF classification results within the test cohort onto UMAP demonstrates that 2 out of 5 cases cluster with the tumor type assigned by the classifier (Figure 3*J*). In summary, RF achieved almost 97% accuracy in an independent test cohort with similar results across different tumor types, with potential for further improvement using score cutoffs.

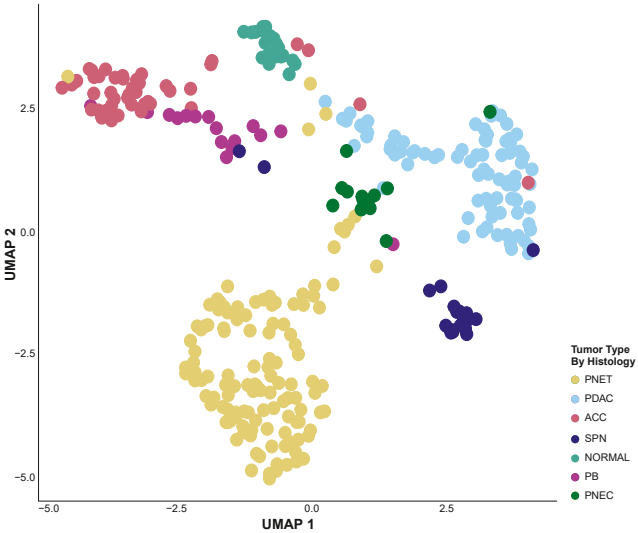


Figure 2. Visualization of methylation profiles of the pancreatic dataset using UMAP dimensionality reduction.

Detection of Non-pancreatic Neoplasms

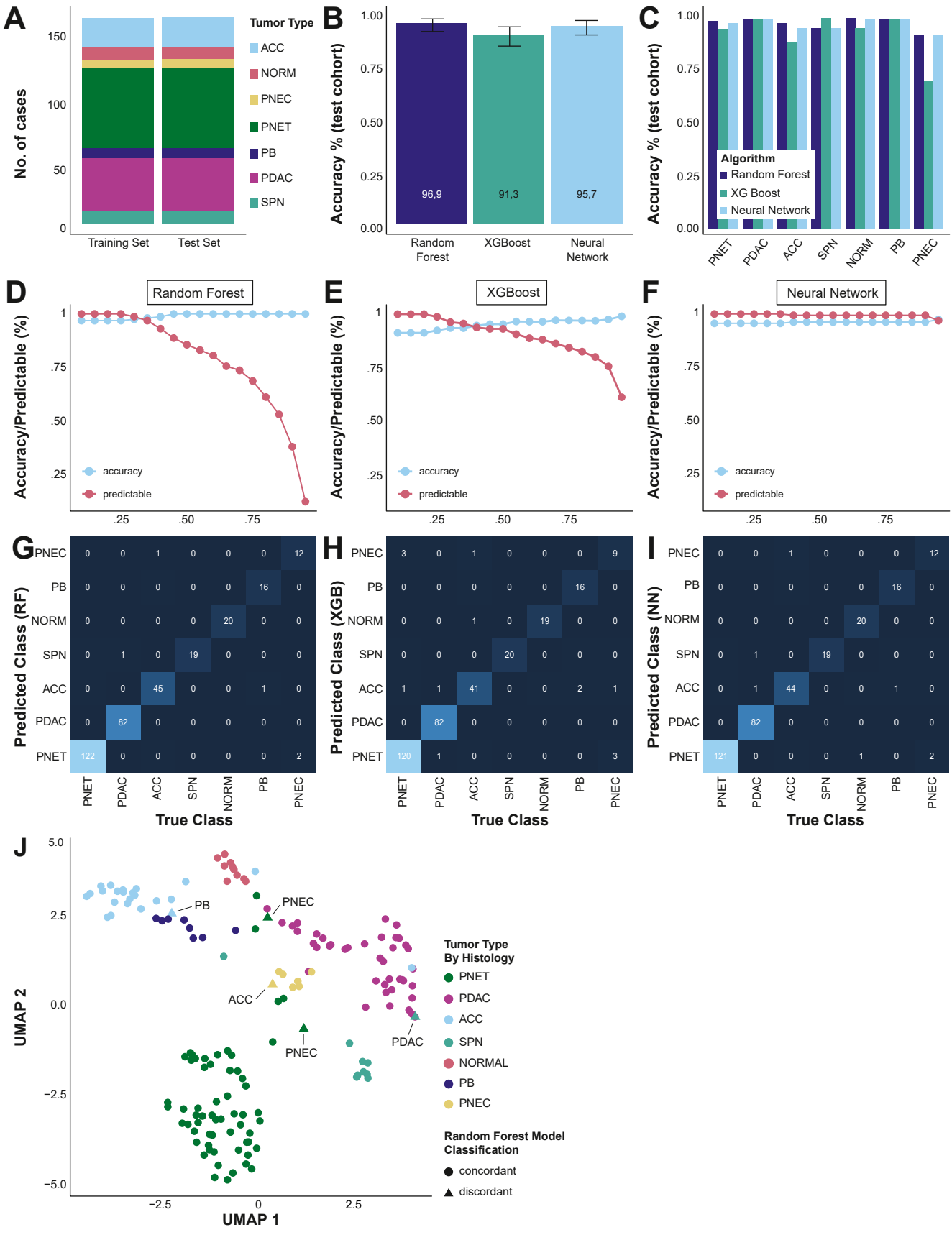
Considering that the models are specifically trained to classify pancreatic neoplasms, for diagnosis it is crucial to detect neoplasms of non-pancreatic origin. First, we applied the RF model in an extended cohort of 3708 samples encompassing 29 nonpancreatic neoplasms including those known to metastasize to the pancreas (Figure 4*A*). Out of 7 possible categories, the RF model classified all non-pancreatic tumors as either PDAC, PNET, ACC, SPN, or PNEC (Figure 4*B*). Non-pancreatic neoplasms received significantly lower probability scores compared with pancreatic neoplasms and the receiver operating characteristic analysis of RF probability scores revealed an area under the curve of >0.95 (Figure 4*C* and *D*). Logistic regression using RF scores and classification results as input was used to identify non-pancreatic neoplasms. The resulting probability scores were lower for non-pancreatic neoplasms (Figure 4*E*) and receiver operating characteristic analysis showed an improved area under the curve of >0.99 compared with RF scores alone (Figure 4*F*). At a cutoff of 0.5, 7% of pancreatic tumors were classified as non-pancreatic, whereas <0.1% non-pancreatic neoplasms were classified as originating from the pancreas (Figure 4*G* and *H*). This additional safety measure can be used to flag samples of non-pancreatic origin. Excluding flagged samples from the test cohort improved classifier performance to 97.9% (RF), 93.6% (XGB), and 97.9% (NN).

Classifier Application in Histopathologic Challenging and Misdiagnosed Clinical Cases

To investigate the added value of methylation profiling, we retrospectively applied the RF classifier on

3 cases that were misdiagnosed (cases A1 and B) or difficult to classify (case C) during initial histopathologic assessment (Figure 5).

The first case, a ^{68}Ga-DOTATOC PET/CT, revealed a moderately radioactive deposition in the pancreas and spine, suggesting metastasized PNET. Histopathologically,



the differential diagnoses of the spinal cord biopsy were PNET and ACC. Because neuroendocrine markers were positive and ACC marker BCL10 was negative (Figure 5, case A1), the case was signed out as metastasis of a PNET grade 3 and the patient underwent resection of the primary tumor in the pancreas. Histopathologic analysis of the resection specimen showed comparable immunohistochemistry with the biopsy but, surprisingly, strong BCL10 expression, changing the diagnosis to ACC (Figure 5, case A2). Methylation profiling by the RF model classified both the bone biopsy and the pancreatic lesion as ACC.

The second case was initially classified as PNET but histopathologic revision at a tertiary expert center raised doubt about this diagnosis. Based on additional immunohistochemistry of SPN markers (β-catenin, vimentin, and CD10), the case was reclassified as SPN (Figure 5, case B). The RF model also classified the case as SPN.

The third case was a liver biopsy that showed a tumor with varying neuroendocrine marker expression and BCL10 positivity in about half of the tumor cells (Figure 5, case C). The differential diagnosis included a pancreatic neuroendocrine neoplasm (PNET G3 or PNEC), ACC with neuroendocrine differentiation, or mixed acinar-neuroendocrine neoplasm. Molecular immunohistochemistry showed, among other things, ATRX loss and TSO500 sequencing TruSight Oncology (Illumina) revealed mutations in *NRAS* (p.Q61R), *RAD50*, and *ATM*. Taken together, although noted that this is a difficult tumor to classify, the case was signed out as a MiNEN (PNET G3-ACC). The RF classification yielded divergent results, with high probability scores for both ACC and PNET, aligning with the suspicion of a MiNEN. Because of these divergent probability scores, we analyzed 7 additional primary MiNENs that obtained corresponding divergent classification results and higher entropy compared with pancreatic tumors (Supplementary Figure 5*A-E*).

Additional Analysis on Paired Cases

First, 4 primary tumors with matched metastases were analyzed: 3 liver metastases from 2 primary PNETs, 1 liver metastasis from a MiNEN, and 1 bone metastases from an ACC (Figure 5, cases A1 and A2). High correlations were observed between the matched samples compared with other cases, and metastases and primary lesions were classified identically (Supplementary Figure 6*A-C*). Second, 4 paired preoperative FNA cytology and resection specimens were analyzed. The EPIC array successfully processed all FNA cases, and there were strong correlations between matched samples compared with other cases. Identical classifications were observed in all matched cases (Supplementary Figure 7*A-C*).

Discussion

This proof-of-concept study effectively uses DNA methylation profiling to address the diagnostic challenge of distinguishing between different primary pancreatic tumors. After benchmarking the performance of 3 state-of-the-art machine learning models (NN, XGB, RF) on a comprehensive cohort spanning all relevant pancreatic tumor types, the RF model emerged as the best classifier with 96.9% accuracy. Because the models' classification is limited to classify tumor types present in their training data, and to account for metastases to the pancreas, we developed a second algorithm based on the pancreatic dataset supplemented with non-pancreatic neoplasms. This approach reduced the probability of a false pancreatic neoplasm diagnosis to less than 1 in 1000 cases, whereas 7% of true pancreatic tumors were classified as non-pancreatic. Considering the rare occurrence of metastases to the pancreas and the artificial case mix, the actual risk of a false pancreatic tumor diagnosis is likely much lower.

The pancreatic-tumor classifier in this study is highly accurate, comparable with methylation-based classifiers for other tumor types (95.1% for brain tumors,[13] 99.9% for sarcoma,[15] 96% for squamous cell carcinoma,[14] and 89% for head and neck squamous cell carcinoma[16]). Although RF models have traditionally been successful and widely used algorithms, recent studies suggest superiority of other models, such as NN, Support Vector Machines, and LOGREG.[13–16,21] In this study, the performance of the 3 models was not significantly different. Nonetheless, adjusting cutoffs for minimum probability scores demonstrated that RF achieved higher accuracies for most of the cases. Its probability score distribution makes it especially well-suited for detecting nonpancreatic entities, making RF the preferred model in this study.

This study underscores the benefits of methylation profiling in diagnosing pancreatic tumors in several ways. First, the algorithm presented here effectively distinguishes between different pancreatic tumors, even when they exhibit overlapping signatures in unsupervised analysis, such as PB and ACC, which is likely caused by their similar epigenetic profiles linked to acinar differentiation.[22] Our machine learning model correctly classified 30 out of 31 ACC and PB cases in the test

Figure 3. Development, comparison, and validation of DNA methylation–based classifier. Sample counts for each tumor type in the training and test sets (*A*). The classifiers achieved accuracies of 96.9% (Random Forest), 91.3% (XGB), and 95.7% (Neural Network) in the test cohort (*B*). Evaluation of classifier accuracies by tumor type shows that the Random Forest classifier consistently outperforms the Neural Network and XGB algorithms (*C*). Threshold analysis of Random Forest (*D*), XGB (*E*), and Neural Network (*F*) probability scores demonstrates that the Random Forest model achieves the highest accuracy in most cases. Confusion matrices for Random Forest (*G*), XGB (*H*), and Neural Network (*I*) showing discordant classifications per tumor type. *UMAP* visualization of the random forest classification results within the test cohort demonstrates that 2 out of 5 cases actually cluster with the tumor type assigned by the classifier (*J*).

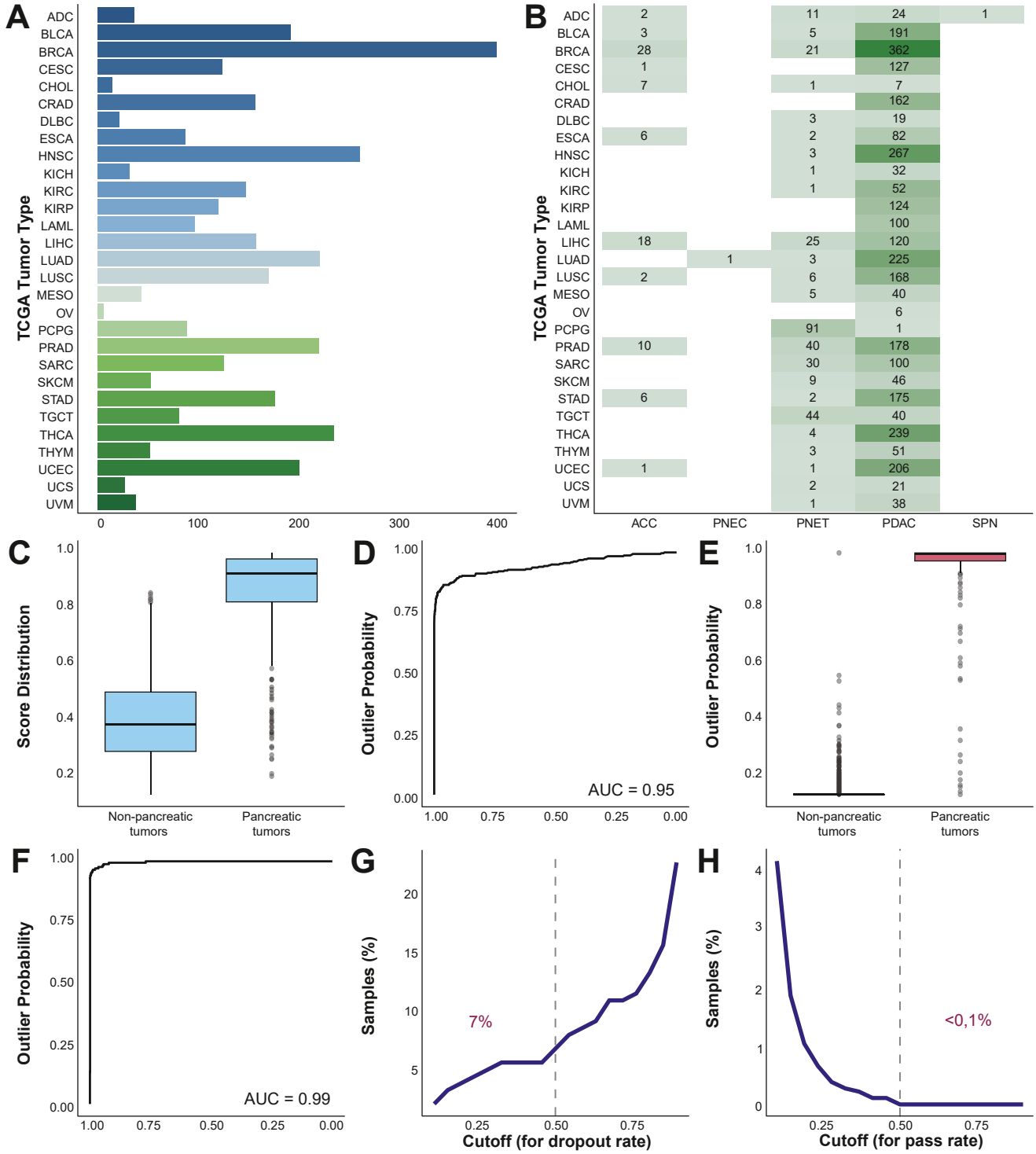


Figure 4. Detection of non-pancreatic neoplasms. Sample counts per tumor type in the non-pancreatic dataset (*A*). The Random Forest model classified all non-pancreatic tumors as PDAC, PNET, ACC, PNEC, or SPN (*B*). Random Forest score distribution of non-pancreatic and pancreatic neoplasms (*C*). Receiver operating characteristic (ROC) curves based on Random Forest distribution scores show an outlier probability with area under the curve (AUC) of 0.95 (*D*). A second algorithm (logistic regression) was developed to identify non-pancreatic neoplasms, and its probability scores were lower for non-pancreatic neoplasms (*E*). ROC curve based on logistic regression shows an improved AUC of >0.99 (*F*). Threshold analysis shows real pancreatic samples considered as non-pancreatic neoplasms (dropouts) (*H*) and non-pancreatic neoplasms classified as pancreatic neoplasms (pass rate) (*I*). At a cutoff of 0.5, dropout is 7% and pass rate <0.1%. The Cancer Genome Atlas (TCGA) abbreviations: ADC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; CRAD, colorectal adenocarcinoma; DLBC, lymphoid neoplasm diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma.

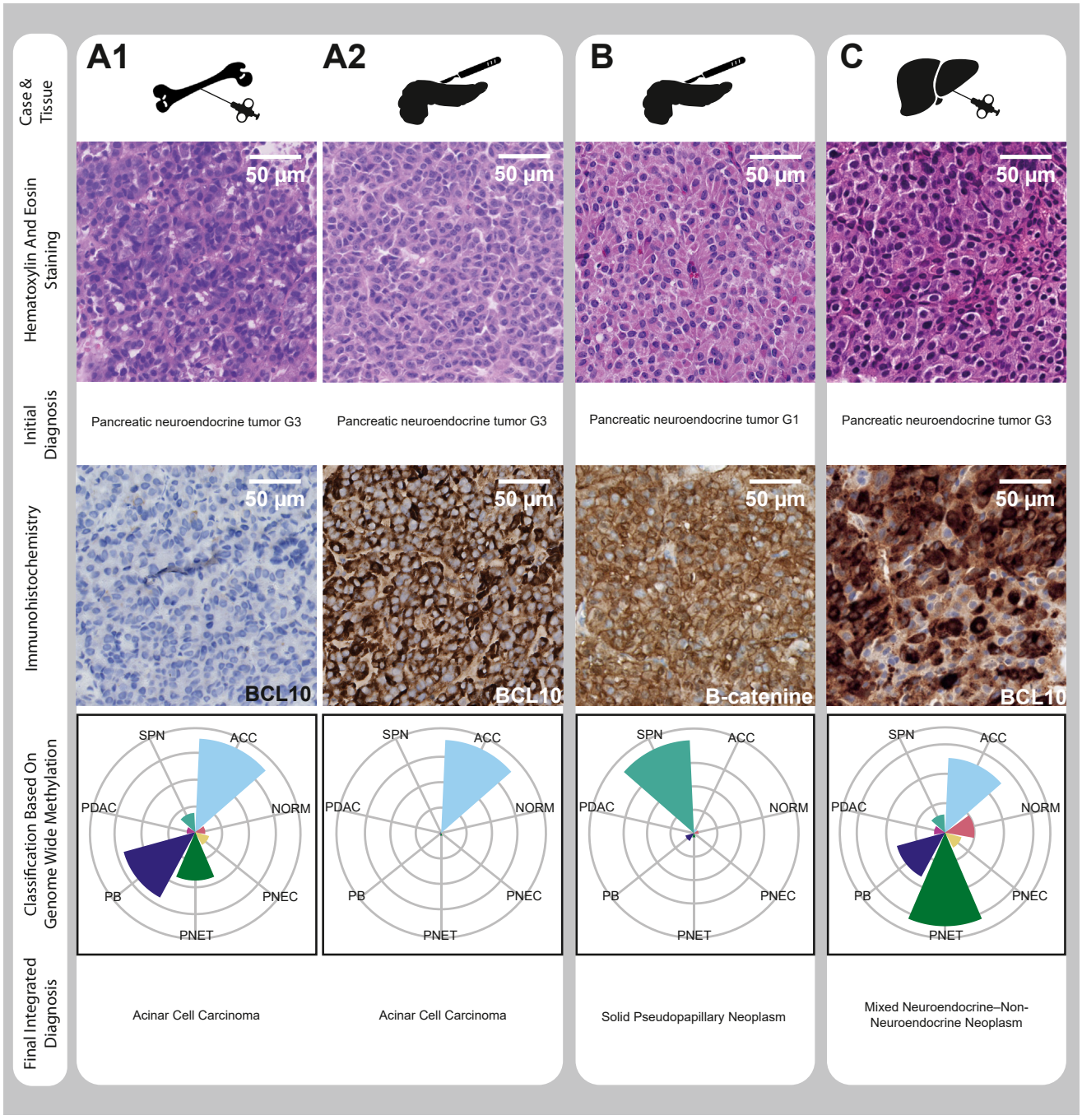


Figure 5. Classifier application in histopathologically misdiagnosed and challenging clinical cases. Case A includes a spine biopsy (case A1) and resection specimen of the primary pancreatic lesion (case A2) of a patient with moderate radioactive depositions in the pancreas and spine on ^{68}Ga-*DOTATOC* PET/CT suggesting metastasized PNET. Based on positive neuroendocrine markers and negative ACC marker BCL10, the case was signed out as metastatic PNET grade 3. Histopathologic analysis of the subsequent pancreatic resection specimen showed positivity of BCL10 changing the diagnosis to ACC. The Random Forest model classified both the bone biopsy and the pancreatic lesion as ACC. Case B, initially misdiagnosed as a PNET, was later reevaluated as a SPN on revision at a tertiary referral center. This reclassification was based on additional immunohistochemistry studies, which showed positivity for SPN markers (β-catenin, vimentin, and CD10). The Random Forest model confirmed the case as SPN. Case C is a liver biopsy classified as MiNEN (mixed ACC-PNET) based on expression of both neuroendocrine markers and ACC marker BCL10 in approximately half of the tumor cells. The Random Forest classification was divergent and showed a high probability score for ACC and PNET, consistent with the histopathologic suspicion of MiNEN.

cohort, indicating that dissimilarities between these tumors are significant enough to enable differentiation through DNA methylation. Of note, because of existing diagnostic challenges and the impact of sampling bias, methylation profiling alone is currently not considered suitable for classification of potential MiNEN cases.

Second, this classifier assists in excluding other tumor types in the differential diagnosis of pancreatic lesions. The non-pancreatic dataset includes primary neoplasms known to metastasize to or directly invade the pancreas, such as renal cell, lung, breast, colonic, and gastric cancer.[29,30] This is valuable in excluding the possibility of metastasis to the pancreas. Additionally, the presence of cholangiocarcinoma in the non-pancreatic dataset helps exclude periampullary tumors from consideration. Although it would have been interesting to include more periampullary tumors, such methylation data were unavailable from public resources.

Lastly, DNA methylation profiling can aid diagnosis of FNA/FNB samples of primary tumors and metastases, particularly in challenging cases with insufficient material for additional immunohistochemistry, technical artefacts (ie, caused by decalcification, or cellular distortion). This study demonstrates that the RF model accurately classified all FNA/FNB samples. Thereby, it correctly classified the preoperative biopsy of a case affected by technical artifacts (case A), highlighting the potential robustness of DNA methylation profiling in various sample conditions, mitigating potential diagnostic errors.

Nonetheless, using publicly available DNA methylation data has limitations. First, our dataset had limited data on PB, SPN, and PNEC from only a few studies, possibly contributing to slightly lower model performance in these tumor types (Figure 3*C*). Second, we were unable to verify the histopathologic diagnoses from the publicly available DNA methylation data. Most samples that received a discordant classifier classification clustered with the corresponding tumor type. Although UMAP should not be considered a definitive means to classify samples, it suggests that at least in several cases, methylation profiling may have identified misdiagnosed samples. This has been noted previously for PDACs of The Cancer Genome Atlas cohort, which were later shown to actually represent PNETs.[31,32] Because of a lack of additional data this hypothesis cannot be proven. Nonetheless, these dissimilarities highlight the potential value of methylation profiling as a second layer of information.

This study again demonstrates the potential of epigenetic tumor classification and shows that DNA methylation-based classification can be achieved with limited tumor cells, such as FNA cytology samples. This suggests that the previously reported 13% misclassification rate of pancreatic neuroendocrine neoplasms using FNA cytology may be reduced with methylation profiling.[12] The ability to assist in preoperative clinical decision making makes this diagnostic tool highly appealing.

DNA methylation is now routinely used to aid diagnostics of central nervous and soft tissue tumors.[15,18] EPIC array can be performed on histologic and cytologic specimens, is routinely available in most academic centers or accessible on a consultation basis, and results are obtained in 1–2 weeks. Ongoing optimizations of brain tumor classifiers aim for a diagnosis within 90 minutes,[33] offering promising prospects for broader clinical use. Considering our model is a proof-of-concept classification model, further extension and validation are necessary before considering its widespread accessibility for daily practice.

To conclude, this study presents a highly accurate DNA methylation-based classifier capable of differentiating pancreatic neoplasms. At the same time, non-pancreatic entities are identified with near perfect precision, thereby preventing false pancreatic tumor diagnoses excellently. The successful classification of FNA samples makes the technique applicable in the preoperative setting.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at http://doi.org/10.1016/j.cgh.2024.02.007.

References

1. Hackeng WM, Hruban RH, Offerhaus GJ, et al. Surgical and molecular pathology of pancreatic neoplasms. Diagn Pathol 2016;11:47.
2. Papavramidis T, Papavramidis S. Solid pseudopapillary tumors of the pancreas: review of 718 patients reported in English literature. J Am Coll Surg 2005;200:965–972.
3. Schmidt CM, Matos JM, Bentrem DJ, et al. Acinar cell carcinoma of the pancreas in the United States: prognostic factors and comparison to ductal adenocarcinoma. J Gastrointest Surg 2008;12:2078–2086.
4. Sonbol MB, Mazza GL, Mi L, et al. Survival and incidence patterns of pancreatic neuroendocrine tumors over the last 2 decades: a SEER Database analysis. Oncologist 2022; 27:573–578.
5. Wisnoski NC, Townsend CM Jr, Nealon WH, et al. 672 patients with acinar cell carcinoma of the pancreas: a population-based comparison to pancreatic adenocarcinoma. Surgery 2008; 144:141–148.
6. Calimano-Ramirez LF, Daoud T, Gopireddy DR, et al. Pancreatic acinar cell carcinoma: a comprehensive review. World J Gastroenterol 2022;28:5827–5844.
7. Cruz MAA, Moutinho-Ribeiro P, Costa-Moreira P, et al. Solid pseudopapillary neoplasm of the pancreas: unfolding an intriguing condition. GE Port J Gastroenterol 2022;29:151–162.
8. Hofland J, Falconi M, Christ E, et al. European Neuroendocrine Tumor Society 2023 guidance paper for functioning pancreatic neuroendocrine tumour syndromes. J Neuroendocrinol 2023;35: e13318.
9. Kos-Kudla B, Castano JP, Denecke T, et al. European Neuroendocrine Tumour Society (ENETS) 2023 guidance paper for nonfunctioning pancreatic neuroendocrine tumours. J Neuroendocrinol 2023;e13343.
10. Dhillon J. Non-ductal tumors of the pancreas. Monogr Clin Cytol 2020;26:92–108.
11. Manfrin E, Parisi A, Stefanizzi L, et al. Bcl-10, trypsin and synaptophysin helps recognize acinar cell and mixed acinar

neuroendocrine cell carcinoma of the pancreas on both preoperative cytological samples and needle biopsy specimens. Pathol Res Pract 2021;226:153593.
12. Sigel C, Reidy-Lagunes D, Lin O, et al. Cytological features contributing to the misclassification of pancreatic neuroendocrine tumors. J Am Soc Cytopathol 2016;5:266–276.
13. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. Nature 2018; 555:469–474.
14. Jurmeister P, Bockmayr M, Seegerer P, et al. Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. Sci Transl Med 2019;11:eaaw8513.
15. Koelsche C, Schrimpf D, Stichel D, et al. Sarcoma classification by DNA methylation profiling. Nat Commun 2021;12:498.
16. Leitheiser M, Capper D, Seegerer P, et al. Machine learning models predict the primary sites of head and neck squamous cell carcinoma metastases based on DNA methylation. J Pathol 2022;256:378–387.
17. Moran S, Martinez-Cardus A, Sayols S, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol 2016;17:1386–1395.
18. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro Oncol 2021;23:1231–1251.
19. Fernandez AF, Assenov Y, Martin-Subero JI, et al. A DNA methylation fingerprint of 1628 human samples. Genome Res 2012;22:407–419.
20. Loyfer N, Magenheim J, Peretz A, et al. A DNA methylation atlas of normal human cell types. Nature 2023;613:355–364.
21. Hackeng WM, Dreijerink KMA, de Leng WWJ, et al. Genome methylation accurately predicts neuroendocrine tumor origin: an online tool. Clin Cancer Res 2021;27:1341–1350.
22. Benhamida JK, Vyas M, Tanaka A, et al. Pancreatoblastomas and mixed and pure acinar cell carcinomas share epigenetic signatures distinct from other neoplasms of the pancreas. Mod Pathol 2022;35:956–961.
23. Jakel C, Bergmann F, Toth R, et al. Genome-wide genetic and epigenetic analyses of pancreatic acinar cell carcinomas reveal aberrations in genome stability. Nat Commun 2017;8:1323.
24. Chan CS, Laddha SV, Lewis PW, et al. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. Nat Commun 2018;9:4158.
25. Di Domenico A, Pipinikas CP, Maire RS, et al. Epigenetic landscape of pancreatic neuroendocrine tumours reveals distinct cells of origin and means of tumour progression. Commun Biol 2020;3:740.
26. Endo Y, Fujimoto M, Ito N, et al. Clinicopathological impacts of DNA methylation alterations on pancreatic ductal adenocarcinoma: prediction of early recurrence based on genome-wide DNA methylation profiling. J Cancer Res Clin Oncol 2021;147:1341–1354.
27. Selenica P, Raj N, Kumar R, et al. Solid pseudopapillary neoplasms of the pancreas are dependent on the Wnt pathway. Mol Oncol 2019;13:1684–1692.
28. Yachida S, Totoki Y, Noe M, et al. Comprehensive genomic profiling of neuroendocrine carcinomas of the gastrointestinal system. Cancer Discov 2022;12:692–711.
29. Adsay NV, Andea A, Basturk O, et al. Secondary tumors of the pancreas: an analysis of a surgical and autopsy database and review of the literature. Virchows Arch 2004;444:527–535.
30. Hiotis SP, Klimstra DS, Conlon KC, et al. Results after pancreatic resection for metastatic lesions. Ann Surg Oncol 2002; 9:675–679.
31. Peran I, Madhavan S, Byers SW, et al. Curation of the pancreatic ductal adenocarcinoma subset of the Cancer Genome Atlas is essential for accurate conclusions about survival-related molecular mechanisms. Clin Cancer Res 2018;24:3813–3819.
32. Zhu T, Liu J, Beck S, et al. A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. Nat Methods 2022;19:296–306.
33. Vermeulen C, Pages-Gallego M, Kester L, et al. Ultra-fast deep-learned CNS tumour classification during surgery. Nature 2023; 622:842–849.

Correspondence
Address correspondence to: Lodewijk A.A. Brosens, MD, PhD, Internal Mail Box H04-312, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands. e-mail: l.a.a.brosens@umcutrecht.nl. or Anna Vera D. Verschuur, MD, Internal Mail Box G02.5.26, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands. e-mail: annaveraverschuur@gmail.com.

Acknowledgments
Christoph Geisenberger and Lodewijk A.A. Brosens contributed equally to this article. The results published here are in part based on data generated by the Heidelberg Cancer Genome Atlas Research Network: https://www.cancer.gov/tcga.

CRediT Authorship Contributions
Anna Vera Ditte Verschuur, MD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Methodology: Equal; Validation: Equal; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Supporting)
Wenzel M. Hackeng, MD, PhD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Supporting; Methodology: Equal; Writing – review & editing: Equal)
Florine Westerbeke, MD (Data curation: Equal; Formal analysis: Equal; Writing – review & editing: Equal)
Jamal K. Benhamida, MD (Data curation: Equal)
Olca Basturk, MD, PhD (Data curation: Equal)
Pier Selenica, MD (Data curation: Lead)
G. Mihaela Raicu, MD, PhD (Data curation: Equal)
I. Quintus Molenaar, MD, PhD (Writing – review & editing: Equal)
Hjalmar C. van Santvoort, MD, PhD (Writing – review & editing: Equal)
Lois A. Daamen, MD, PhD (Methodology: Equal; Writing – review & editing: Equal)
David S. Klimstra, MD (Data curation: Equal)
Shinichi Yachida, MD, PhD (Data curation: Equal; Writing – review & editing: Equal)
Claudio Luchini, MD, PhD (Writing – review & editing: Equal)
Aatur D. Singhi, MD, PhD (Writing – review & editing: Equal)
Christoph Geisenberger, MD (Conceptualization: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Software: Lead; Supervision: Lead; Validation: Equal; Writing – original draft: Supporting; Writing – review & editing: Lead)
Lodewijk A.A. Brosens, MD, PhD (Conceptualization: Equal; Funding acquisition: Lead; Methodology: Equal; Project administration: Lead; Resources: Lead; Supervision: Lead; Writing – original draft: Equal; Writing – review & editing: Lead)

Conflicts of interest
The authors disclose no conflicts.

Funding
Anna Vera D. Verschuur is supported by a research grant of the Dutch Cancer Society (KWF 12978 / 2020-1).

Supplementary Methods

Patients and Samples

Methylation data for the pancreatic dataset (n = 321) were obtained from 7 previously published datasets (n = 304)[1–7] and from resection specimens collected at the University Medical Center Utrecht (UMCU) and Radboud University Medical Center (Radboud UMC), The Netherlands (n = 17) (Table 1). Dataset details can be found in Table 1. Datasets used in Selenica et al[6] and Behamida et al[1] were obtained from the authors on request. Data collected from the UMCU and Radboud UMC pathologic archives were obtained with approval by the UMCU and Radboud UMC Biobank Research Ethics Committees. The data that have not been deposited at EGA, which is hosted by the European Bioinformatics Institute and the Centre for Genomic Regulation under accession number EGAS00001004878,[8] are available on reasonable request. Methylation data of the pancreatic dataset were generated on Infinium HumanMethylation450 (n = 225) and Infinium HumanMethylationEPIC v1.0 (n = 96) platforms.

To facilitate detection of nonpancreatic entities, nonpancreatic tumor samples (n = 3708), spanning 29 different tumor types, were obtained from The Cancer Genome Atlas (TCGA). Three cases, referred for a second opinion or treatment at our tertiary center, which were histopathologically difficult and/or misdiagnosed, were used to illustrate the added value of the classifier (Figure 5). The cases comprise 2 primary tumors included in the pancreatic dataset and 2 additional metastatic cases, 1 of which was a matched metastasis from case 1 (Supplementary Table 1).

The additional analysis on (1) MiNENs (n = 7) included 1 additional local MiNEN case and 6 samples from Benhamida et al[1]; (2) matched primary and metastatic cases (n = 5) included 3 primary tumors included in the pancreatic dataset, 1 primary tumor used for MiNEN analysis, and 5 additional matched metastatic cases, all were local cases; and (3) matched FNA and resection specimens (n = 4) of which all 4 resection specimens were included in the pancreatic dataset, all were local cases (Supplementary Table 1).

Methylation Array Processing

DNA extraction and bisulfite conversion were conducted at our in-house core facility. Freshly cut unstained (4 μm) sections were used and tumor tissue was macrodissected for DNA extraction. DNA extraction was performed using the Maxwell RSC FFPE Plus DNA kit (Promega) after 2-hour proteinase K digestion. Subsequent steps adhered to protocols from the DNA Methylation BeadChip manufacturer (Illumina). In brief, EZ DNA Methylation Kit (Zymo Research) was used for bisulfite conversion, and FFPE Restore Kit (Illumina) addressed formalin fixation-induced DNA damage. Whole-genome amplification on the bisulfite converted and restored DNA was performed, followed by fragmentation, precipitation, and hybridization to CpG Site–specific Beads (Illumina). After clearing unhybridized and nonspecifically hybridized DNA, BeadChips were extended, stained, coated, and scanned using the iScan Software (Illumina). Raw data were exported as IDAT files for further processing.

Preprocessing

IDAT files were analyzed using the software package *minfi*[9] and normalized using the *preprocessNoob* function.[10] Next, probes were removed that (1) were located on sex chromosomes, (2) contain single-nucleotide polymorphisms in the CpG or single-base extension site, or (3) were found to exhibit cross-reactivity previously.[11] Finally, data were filtered for probes available on both platforms (n = 400,962), and available beta values were merged into a data table. IDAT files for TCGA samples were downloaded from https://portal.gdc.cancer.gov; processed as outlined previously; and filtered for primary, nonpancreatic neoplasms (n = 3708 remaining).

Quality Control and Batch Effect

Average methylation (beta value) was calculated for all filtered probes. Bisulfite conversion scores were calculated by dividing the minimal intensity for the red channel by the maximum intensity of the green channel for Bisulfite Conversion II control probes. Tumor purity was predicted using the Bioconductor package *RFpurify*.[12] Batch effect was evaluated by analyzing tissue types that were obtained from 2 or more studies. Average methylation values, conversion scores, and tumor purities were compared, and graphically inspected by *UMAP*.[13]

Unsupervised Analysis

To improve clustering and because of computational considerations, unsupervised analysis was performed on a reduced dataset filtered for the n = 5000 most variable CpG sites across the training set (as ranked by variance). Unsupervised analysis using all probes did not reveal major differences in clustering (Supplementary Figure 1). The R package *umap* was used for dimensionality reduction with *UMAP* with n_neighbours = 15 and min_dist = 0.2.

Classifier Development and Validation

Figure 1*A* shows an overview of the classifier development. Samples were randomly assigned to the training and test cohort stratified by tumor type (50/50 ratio; Figure 3*A*) using the R package *caret*.[14] All models were trained on a reduced dataset using the 5000 most

variable probes as input with 10-fold cross-validation and n = 3 repetitions per fold. To account for uneven distribution across the different tumor types, the *upSample* function was used to upsample all classes to the same number of cases. Neural network, random forest and XGBoost algorithms were trained on the training cohort using the R package *Keras*,[15] *randomForest*,[16] and *xgboost*,[17] respectively. Samples of the test cohort were at no point used for variable selection or model building. Random forest model was trained with default parameters: *trees* = 500, *mtry* = sqrt(5000). XGBoost machines were trained with the following parameters: *nrounds* = 100, *max_depth* = 6, *eta* = 0.3, *gamma* = 1, *colsample_bytree* = 1, *min_child_weight* = 1, *subsample* = 1. The neural network model was trained for n = 50 epochs with a 4-layer structure. Layers 1–3 used 64 neurons with *relu* activation, the last layer *softmax* activation. *RMSProp* was used as the optimizer, *categorical crossentropy* as the loss function, and *accuracy* as the optimization metric. The highest score was used to assign samples to a class (*majority vote*). One-way analysis of variance was used to assess the difference in model accuracies, considering a $P < .05$ statistically significant. Classifier performance for different probability score thresholds was assessed by calculating the proportion of predictable cases and accuracy for every cutoff. Confusion matrices were calculated for all 3 models, and the clustering of misclassification of the best performing classifier was visualized using UMAP.

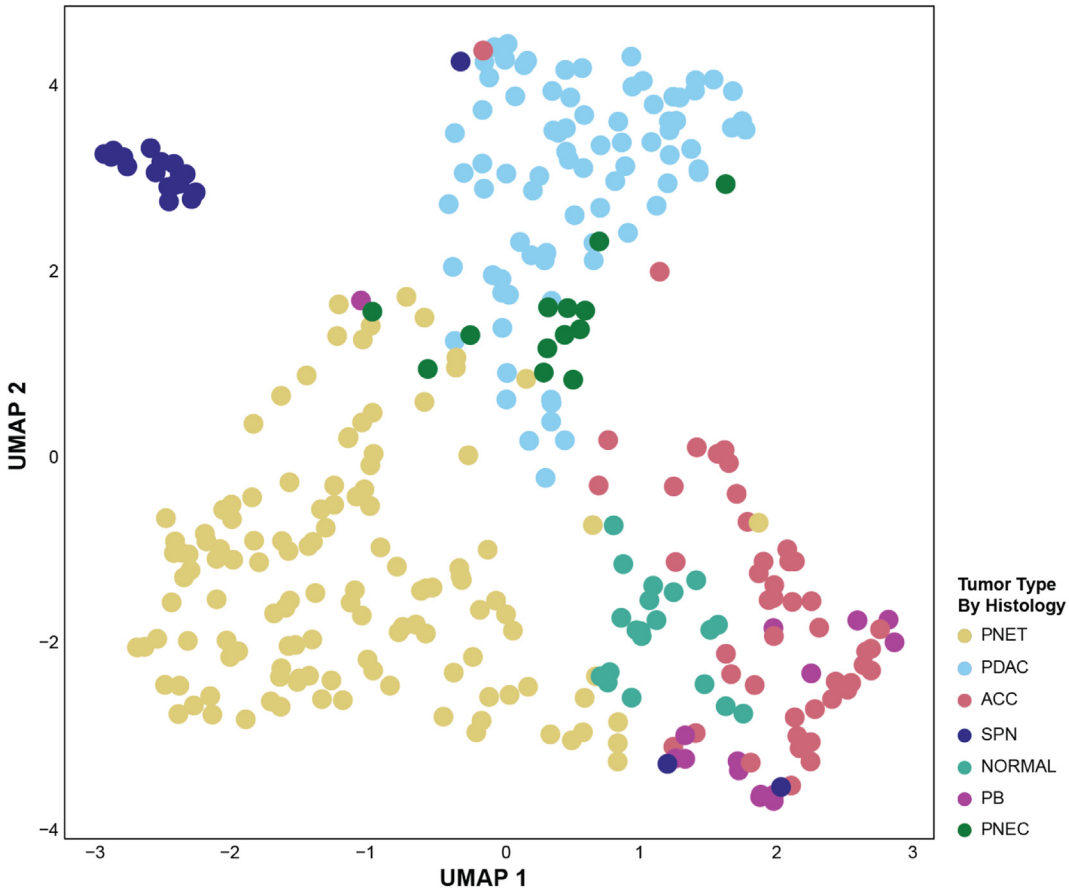
Figure 1*B* shows an overview of the detection of nonpancreatic entities. First, the best performing model's classification results were assessed in an extended cohort of n = 3708 nonpancreatic neoplasms. Second, all samples of the TCGA and pancreatic cohorts were randomly assigned to a training and test set (50/50 split). A logistic regression model using the score and label of the best performing model was trained to differentiate between pancreatic and nonpancreatic (ie, outlier) samples and evaluated in the test set. Finally, model performance in terms of accuracy and drop-out was calculated for different thresholds of logistic regression probability scores. Student *t* test was used to compare random forest model probability scores distribution between the pancreatic and nonpancreatic cohorts. A $P < .05$ was considered statistically significant.

Additional Analysis

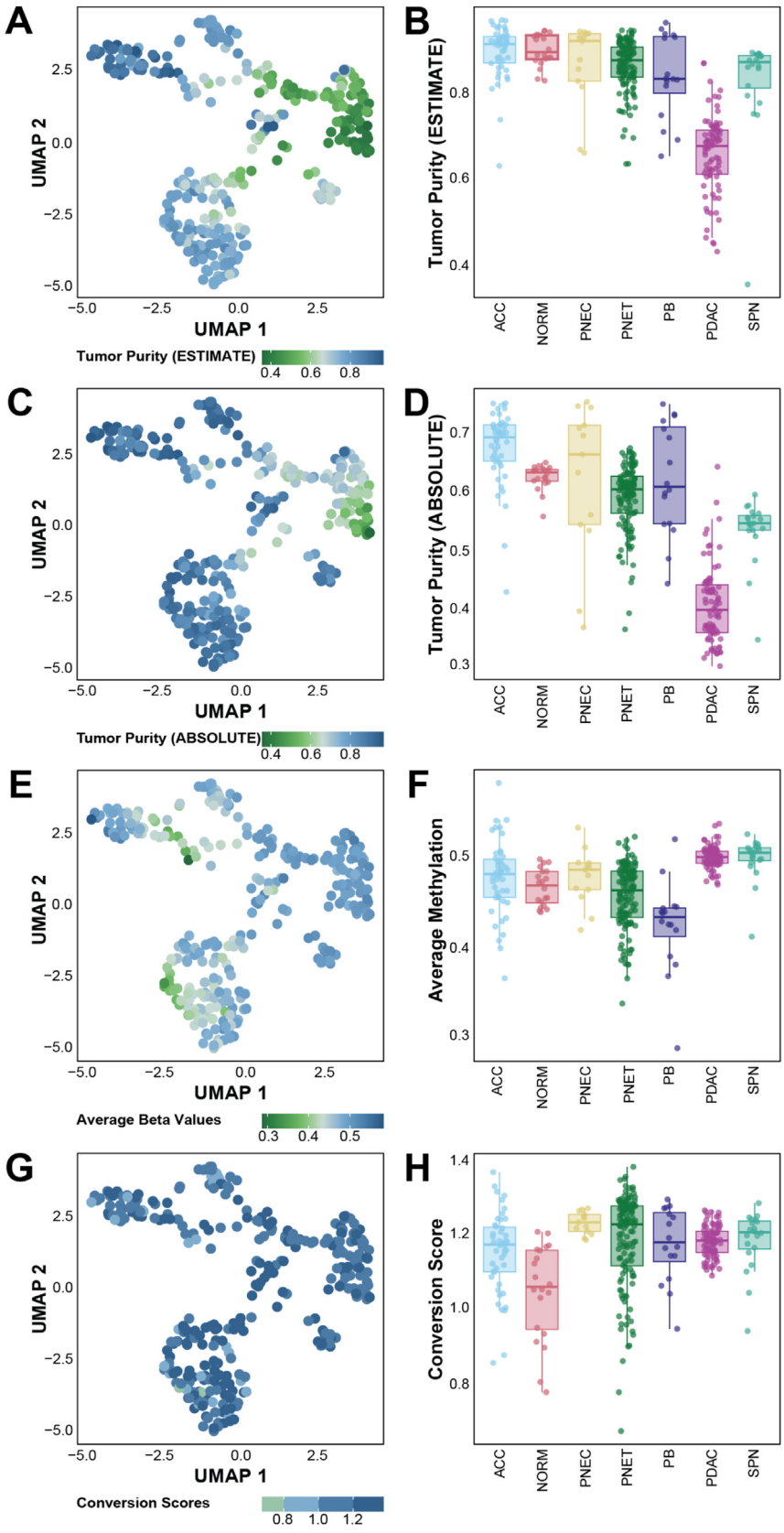
Furthermore, we performed 3 additional subanalyses with (1) MiNEN (n = 7), (2) matched primary and metastatic cases (n = 5), and (3) matched FNA and resection specimens (n = 4). The MiNEN tumors' relationship to other tumor types using reduced dataset (n = 5000 most variable CpG sites) was visualized using UMAP, and entropy scores and mean probability scores per tumor type were evaluated. For the matched primary and metastatic cases and matched FNA and resection specimens, correlations between classifier probability scores in comparison with nonmatched cases were evaluated.

References

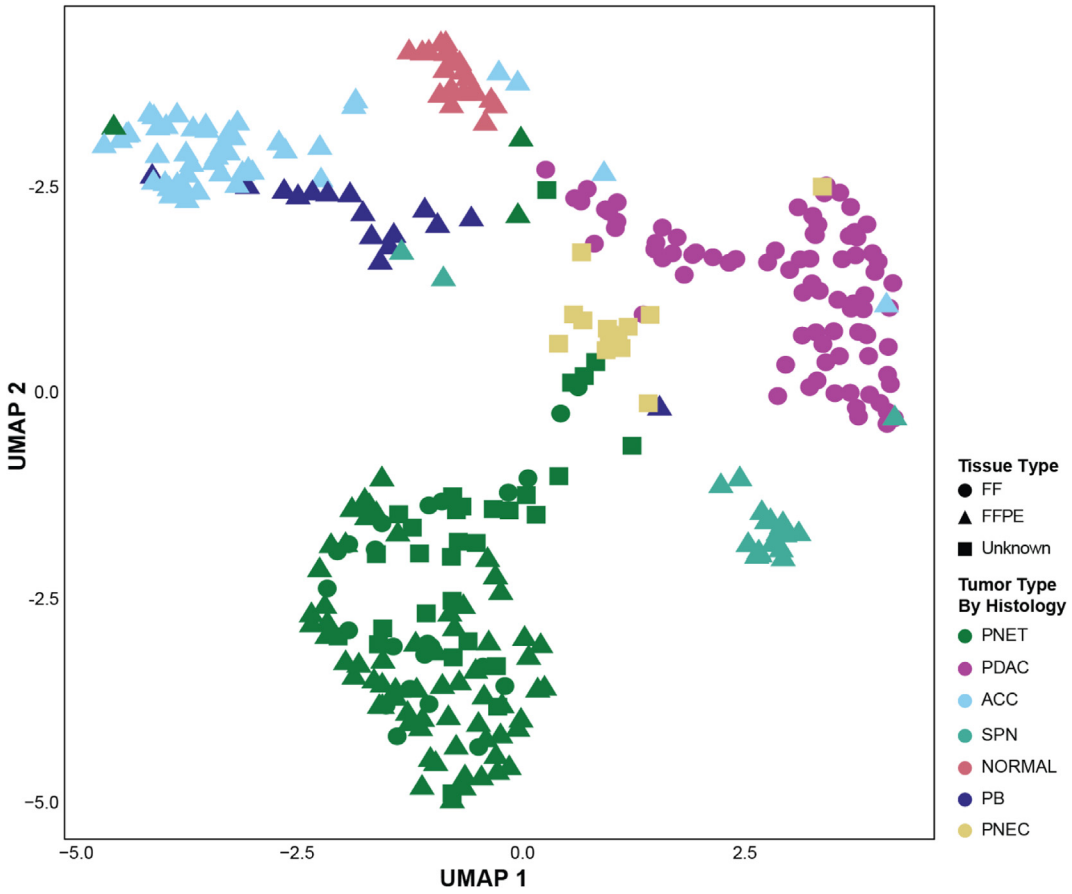
1. Benhamida JK, Vyas M, Tanaka A, et al. Pancreatoblastomas and mixed and pure acinar cell carcinomas share epigenetic signatures distinct from other neoplasms of the pancreas. Mod Pathol 2022;35:956–961.
2. Chan CS, Laddha SV, Lewis PW, et al. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. Nat Commun 2018;9:4158.
3. Di Domenico A, Pipinikas CP, Maire RS, et al. Epigenetic landscape of pancreatic neuroendocrine tumours reveals distinct cells of origin and means of tumour progression. Commun Biol 2020;3:740.
4. Endo Y, Fujimoto M, Ito N, et al. Clinicopathological impacts of DNA methylation alterations on pancreatic ductal adenocarcinoma: prediction of early recurrence based on genome-wide DNA methylation profiling. J Cancer Res Clin Oncol 2021;147:1341–1354.
5. Jakel C, Bergmann F, Toth R, et al. Genome-wide genetic and epigenetic analyses of pancreatic acinar cell carcinomas reveal aberrations in genome stability. Nat Commun 2017;8:1323.
6. Selenica P, Raj N, Kumar R, et al. Solid pseudopapillary neoplasms of the pancreas are dependent on the Wnt pathway. Mol Oncol 2019;13:1684–1692.
7. Yachida S, Totoki Y, Noe M, et al. Comprehensive genomic profiling of neuroendocrine carcinomas of the gastrointestinal system. Cancer Discov 2022;12:692–711.
8. Hackeng WM, Dreijerink KMA, de Leng WWJ, et al. Genome methylation accurately predicts neuroendocrine tumor origin: an online tool. Clin Cancer Res 2021;27:1341–1350.
9. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363–1369.
10. Fortin JP, Triche TJ Jr, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics 2017;33:558–560.
11. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013;8:203–209.
12. Johann PD, Jager N, Pfister SM, et al. RF_Purify: a novel tool for comprehensive analysis of tumor-purity in methylation array data based on random forest regression. BMC Bioinformatics 2019;20:428.
13. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv e-prints 2018 1802;03426.
14. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:1–26.
15. J. Allaire FC. R interface to keras. Available at: https://github.com/rstudio/keras. Accessed March 26, 2024.
16. Breiman L. Random forests. Mach Learn 2001;45:5–32.
17. Chen CQ, Guestrin C. Xgboost: A Scalable Tree Boosting System:Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 13-17 August 2016, 785-794; 2016.



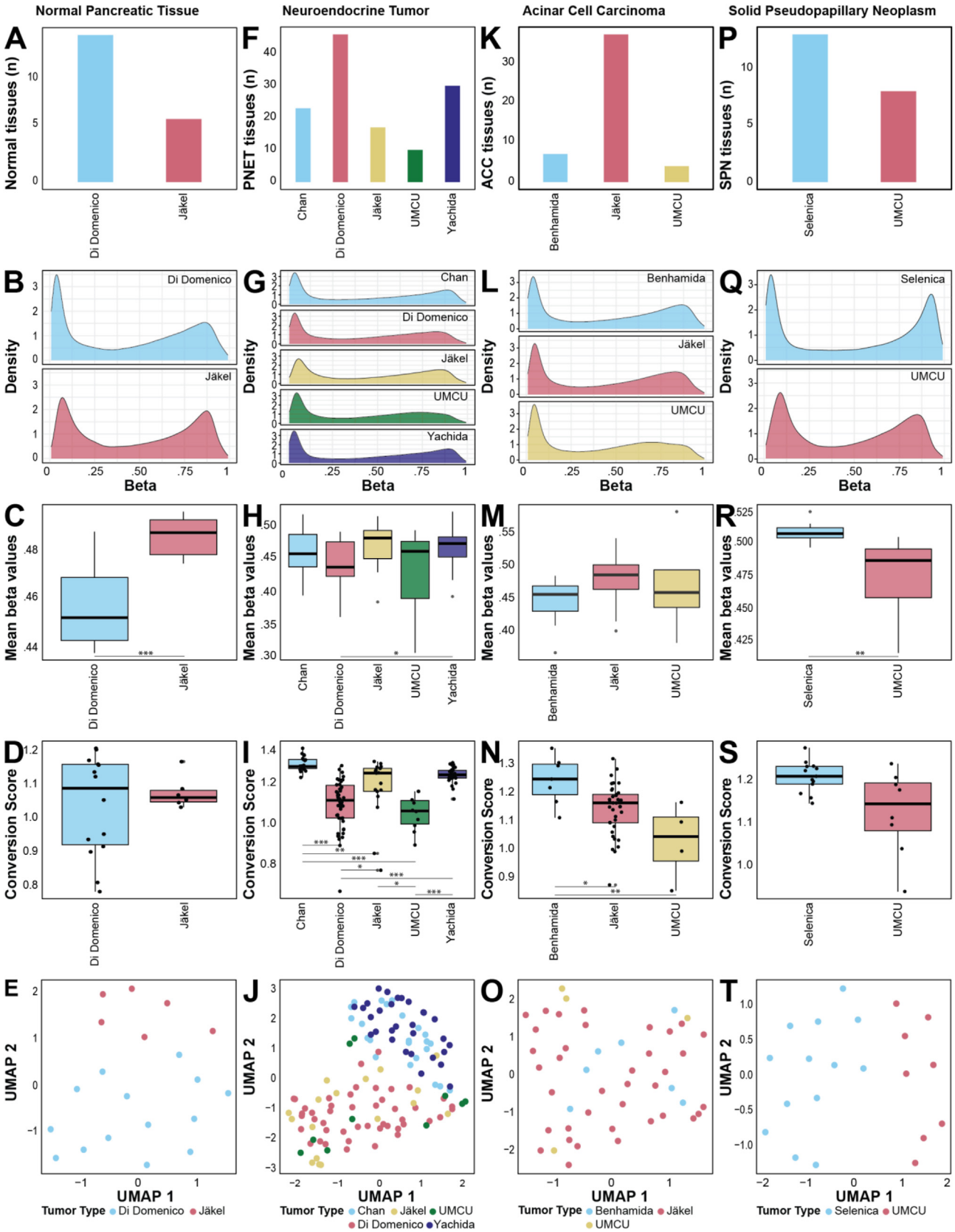
Supplementary Figure 1. Primary pancreatic dataset projected onto the UMAP using all probes.



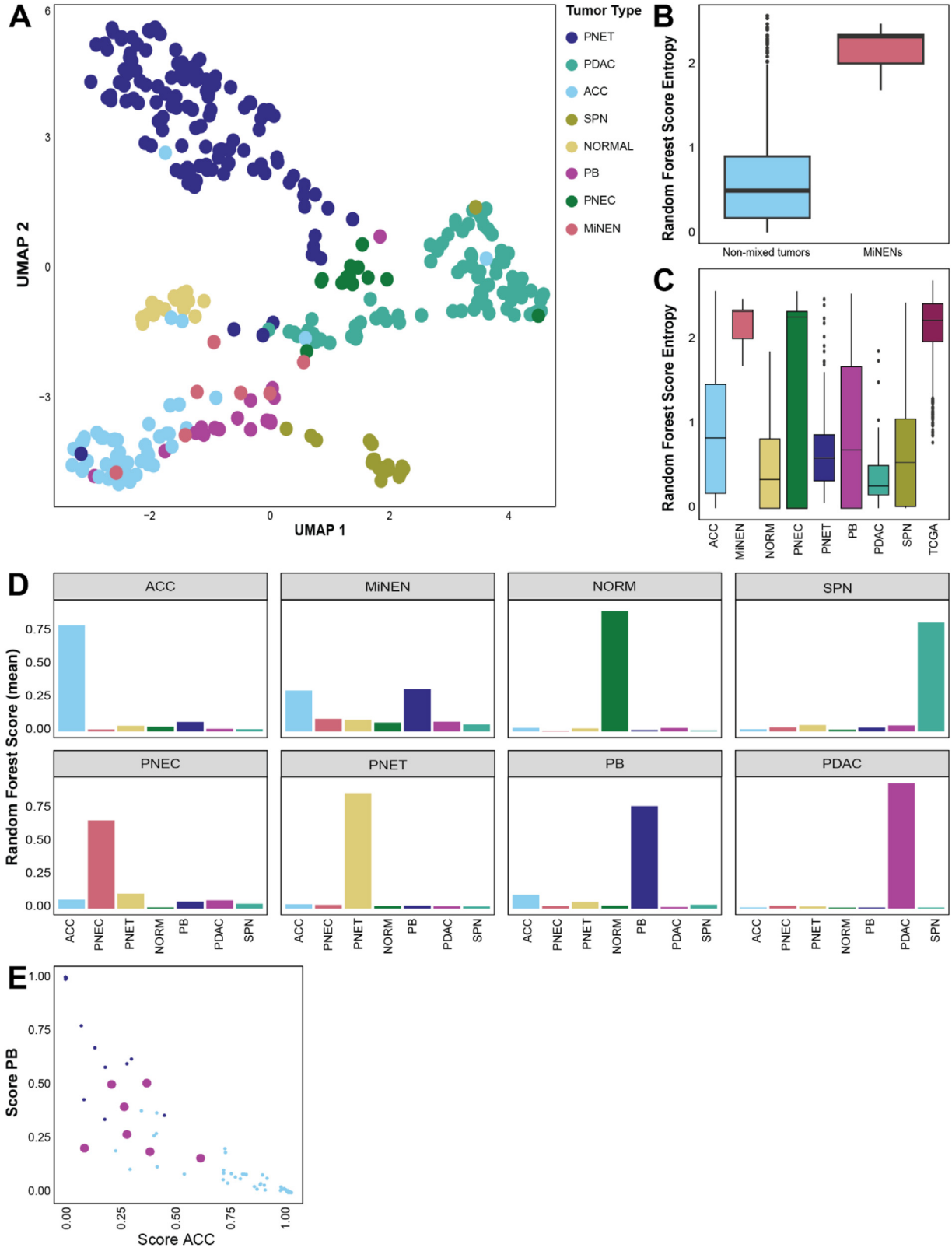
Supplementary Figure 2. Additional unsupervised analysis of pancreatic dataset. ESTIMATE and ABSOLUTE tumor cell purity using the DNA methylation-based "RFPurify" method projected over UMAP and per pancreatic tumor type showing lower tumor purity for pancreatic ductal adenocarcinoma compared with other tumor types (*A-D*). Average methylation and conversion score projected over UMAP and per pancreatic tumor type showing lower genome-wide methylation in pancreatoblastoma (*E-H*).



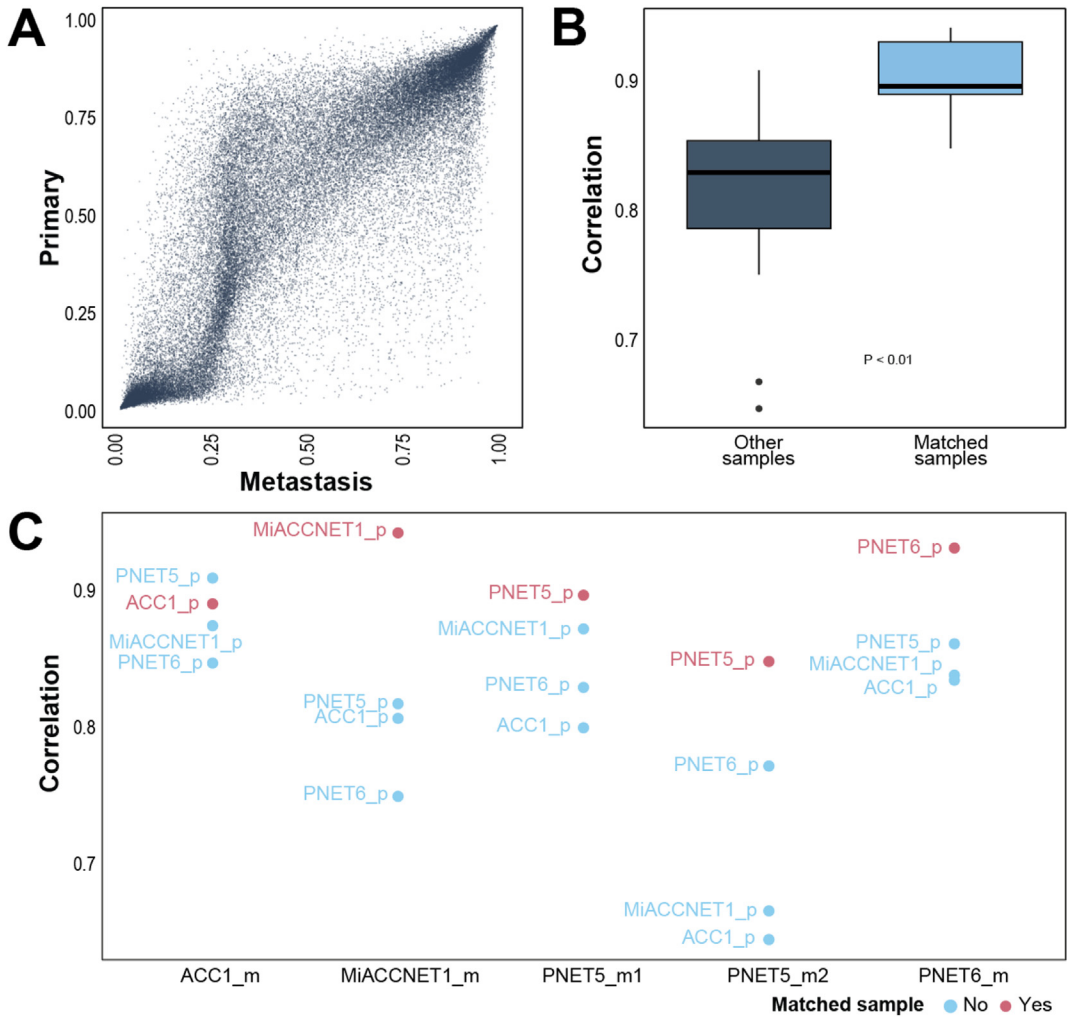
Supplementary Figure 3. Tissue types were projected onto the UMAP. Showing no difference in clustering between fresh-frozen (FF) and formalin-fixed, paraffin-embedded (FFPE) tissue.



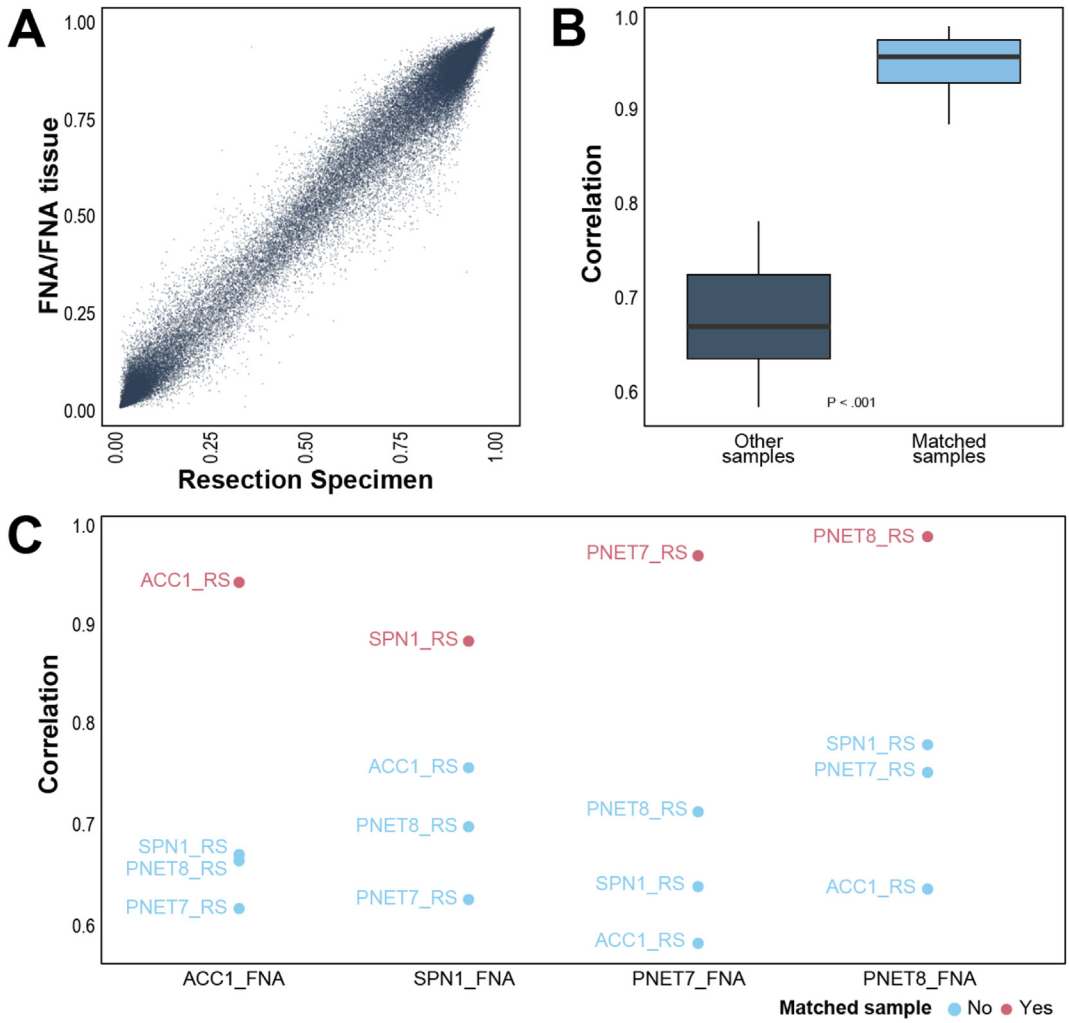
Supplementary Figure 4. Batch effect analysis. Number of samples per source, density plots, mean beta values, conversion scores, and dimensionality reduction through *UMAP* for normal pancreatic tissue (*A-E*), pancreatic neuroendocrine tumor (*F-J*), acinar cell carcinoma (*K-O*), and solid pseudopapillary neoplasm (*P-T*). Results show some differences between sources in the density plots, mean beta values, conversion scores, and UMAP across tumor types. Batch effect analyses revealed small but consistent differences regarding beta values distribution and conversion scores across tumor types. This effect was strongest in normal pancreatic and solid pseudopapillary neoplasm samples, and may suggest batch effects. However, classifier performance seems not to be influenced by the origin of a sample. This implies robustness of the classifier toward minor technical differences resulting from specimen selection and reagent handling. *$P < .05$; **$P < .01$; ***$P < .001$.



Supplementary Figure 5. Analysis of MiNEN. Dimensionality reduction through *UMAP* showing (n = 7) MiNENs samples tend to cluster with acinar cell carcinomas and pancreatoblastomas (*A*). Random **forest** model probability scores of MiNENs show higher entropy as compared with nonmixed tumors (*B*). Random forest model probability scores of MiNENs and TCGA samples show higher entropy as compared with primary pancreatic neoplasms (*C*). Distribution of the mean random forest model probability scores show high scores for the corresponding tumor type, and divergent scores for MiNEN (*D*). Comparing the MiNENs probability scores for acinar cell carcinoma and pancreatoblastoma with that of acinar cell carcinoma and pancreatoblastoma, shows that MiNENs fall between pancreatoblastoma and acinar cell carcinoma (*E*).



Supplementary Figure 6. Analysis of matched primary and metastatic cases. Comparison of the unfiltered beta values of the primary and matched metastasis cases show some correlation (*A*). When quantifying the correlation between the beta values a significantly higher correlation between beta values of matched cases as compared with the nonmatched cases is seen (paired T-test: $P < .01$; 95% confidence interval, -0.15 to -0.04) (*B*). When analyzing the correlation between the samples, it becomes apparent that in most cases the matched primary sample shows superior correlation to its matched metastases sample as compared with other cases. The metastatic acinar cell carcinoma (Figure 5, case 1) shows highest correlation with a primary PNET sample (*C*).



Supplementary Figure 7. Analysis of matched FNA and resection specimen cases. The unfiltered beta values of the FNA and resection specimen cases show high correlation (*A*). When quantifying the correlation between the beta values a significant higher correlation between beta values of matched cases as compared with the nonmatched cases is seen (paired T-test $P <$ .001; 95% confidence interval, -0.33 to -0.20) (*B*). When analyzing the correlation between the samples, it becomes apparent that in all instances the resection specimen samples show superior correlation to their matched FNA sample as compared with other cases (*C*).

Supplementary Table 1. Overview of Histopathologic Challenging and Misdiagnosed Clinical Cases and Samples Used Additional Analysis Including, Source, Array Type, Tumor Type, Location, and Sample Type

Source	Sample name	Array type	Tumor type	Location	Sample Type	In pancreatic dataset
Histopathologic challenging and misdiagnosed clinical cases (Figure 5)						
UMCU	UMCU_ACC1_M	EPIC	ACC	Bone metastasis	FNB	No
UMCU	UMCU_ACC1	EPIC	ACC	Primary	Resection	Yes
UMCU	UMCU_SPN1	EPIC	SPN	Primary	Resection	Yes
UMCU	UMCU_ACC2_M	EPIC	ACC	Liver metastasis	Resection	No
Analysis on MiNEN						
UMCU	UMCU_MiACCNET1	EPIC	MNEN	Primary	Resection	No
Benhamida	Mixed_ACC1	EPIC	MiNEN	Primary	Resection	No
Benhamida	Mixed_ACC2	EPIC	MiNEN	Primary	Resection	No
Benhamida	Mixed_ACC3	EPIC	MiNEN	Primary	Resection	No
Benhamida	Mixed_ACC4	EPIC	MiNEN	Primary	Resection	No
Benhamida	Mixed_ACC5	EPIC	MiNEN	Primary	Resection	No
Benhamida	Mixed_ACC6	EPIC	MiNEN	Primary	Resection	No
Analysis on matched primary and metastatic cases						
UMCU	UMCU_panNET6	EPIC	PNET	Primary	Resection	Yes
UMCU	UMCU_panNET6_M	EPIC	PNET	Liver metastasis	Resection	No
UMCU	UMCU_panNET5	EPIC	PNET	Primary	Resection	Yes
UMCU	UMCU_panNET5_M1	EPIC	PNET	Liver metastasis	FNB	No
UMCU	UMCU_panNET5_M2	EPIC	PNET	Liver metastasis	Resection	No
UMCU	UMCU_ACC1	EPIC	ACC	Primary	Resection	Yes
UMCU	UMCU_ACC1_M	EPIC	ACC	Bone metastasis	FNB	No
UMCU	UMCU_MiACCNET1	EPIC	MiNEN	Primary	Resection	No
UMCU	UMCU_MiACCNET1_M	EPIC	MiNEN	Metastasis	FNA	No
Analysis on matched FNA and resection specimens						
UMCU	UMCU_panNET7_b	EPIC	PNET	Primary	FNA	No
UMCU	UMCU_panNET7	EPIC	PNET	Primary	Resection	Yes
UMCU	UMCU_panNET8_b	EPIC	PNET	Primary	FNA	No
UMCU	UMCU_panNET8	EPIC	PNET	Primary	Resection	Yes
RB	RB_ACC1_b	EPIC	ACC	Primary	FNA	No
RB	RB_ACC1	EPIC	ACC	Primary	Resection	Yes
RB	RB_SPN1_b	EPIC	SPN	Primary	FNA	No
RB	RB_SPN1	EPIC	SPN	Primary	Resection	Yes

ACC, acinar cell carcinoma; FNA, fine-needle aspiration; FNB, fine-needle biopsy; MiNEN, mixed neuroendocrine–nonneuroendocrine neoplasm; PNET, pancreatic neuroendocrine tumor; RB, Radboud UMC; SPN, solid pseudopapillary neoplasm; UMCU, University Medical Center Utrecht.