**Manuscript title**

Non-ductal pancreatic tumor classification by whole genome DNA methylation profiling

**Authors and affiliations**

Anna Vera D. Verschuur\*, MD, Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands  
Florine Westerbeke, Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands  
Jamal K Benhamida, Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA  
Olca Basturk, Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA  
Shinichi Yachida, MD PhD, Department of Cancer Genome Informatics, Graduate School of Medicine, Osaka University, Osaka, Japan. syachida@cgi.med.osaka-u.ac.jp  
Pier Selenica, MD PhD, Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA.  
Wenzel M. Hackeng, MD PhD, Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands  
Christoph Geisenberger\*, Ludwig-Maximilians-University, Munich, Germany  
Lodewijk A.A. Brosens\*, MD PhD, Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

C. Geisenberger and L.A.A. Brosens contributed equally to this article.

***\*Corresponding author***

Lodewijk Brosens, Internal Mail Box H04-312, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands. Fax: 31-30-2544990; Phone 31-88-7556565; E-mail: [l.a.a.brosens@umcutrecht.nl](mailto:l.a.a.brosens@umcutrecht.nl)  
Anna Vera Verschuur, Internal Mail Box G02.5.26, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands. Fax: 31-30-2544990; Phone 31-88-7556565; E-mail: [annaveraverschuur@gmail.com](mailto:annaveraverschuur@gmail.com)

**Funding support:**

A.V.D. Verschuur is supported by a research grant of the Dutch Cancer Society (KWF)

**Conflict of interest:**

The authors declare that they have no conflict of interest.   
  
**Acknowledgements:**

**Word count:**

/5000

References:   
/50

# Statement of translational relevance (120-150)

# Abstract (250)

**Background and aim:** Histopathological diagnosis of non-ductal pancreatic tumors including acinar cell carcinoma’s (ACC), solid pseudopapillary neoplasm (SPN), pancreatoblastoma (PB) and pancreatic neuroendocrine neoplasms (PanNENs) may be challenging in daily clinical practice while it is crucial for therapeutic decision making. The cancer methylome harbors characteristics reflecting the cell of origin allowing identification of tumor origin. To facilitate diagnosis within non-ductal pancreatic tumors, here we have built a methylation-based classification model.  
**Methods:** DNA methylation profiling was obtained from 202 non-ductal pancreatic tumor (ACC, n = 44; PanNENs, n = 129; PB, n = 16; SPN, n = 13), 82 ductal pancreatic adenocarcinoma and 20 normal pancreas samples and we trained three different machine learning methods including neural network (NN), Random Forest (RF), Gradient Boosting (xGB). Methylation data of 28 tumors obtained from The Cancer Genome Atlas (TCGA) was used to develop a second random forest model that helps to differentiate between pancreatic and non-pancreatic tumors.  
**Results:** The NN, RF and xGB models achieved high accuracies of 96%, 94% and 92% respectively. After applying a threshold of >99% prediction accuracy, the NN model achieved classification in 78% of cases, outperforming the RF and xGB model that achieved classification in 0% and 2% of cases respectively.

With an AUC of 0.9992 the second random forest was perfectly able to identify pancreatic tumors and thus contributes to an increased probability of the tumor prediction **Conclusion:** This approach of molecular tumor classification can provide valuable information to guide the diagnostic workup of non-ductal pancreatic tumors.

**Key words:**

When applying a threshold of >99% prediction accuracy, classification was achieved in 78%, 0% and 2% when using NN, RF and xGB respectively.

# Introduction (361)

Around 90% of pancreatic cancers are pancreatic ductal adenocarcinomas (PDACs), while the remainder (10%) is derived from non-ductal structures. The latter include acinar cell carcinoma (ACC), solid pseudopapillary neoplasms (SPN) and pancreatic neuroendocrine tumors (PanNENs) and pancreatoblastoma’s (PB) and comprise 1%, 2%, 5% and .. of all non-ductal pancreatic neoplasms in adults respectively (1, 2).

Behavior varies widely: while SPNs are with an 95% 5-year overall survival predominantly indolent, PanNETs and ACCs have respectively a 5-years overall survival of 59% and 45% come close to the aggressiveness of PDACs (3-6). With regard to histomorphology and immunophenotype ACCs, SPNs and PanNETs are similar (1, 2). Immunohistochemistry and molecular genetics have contributed to an improved classification of these tumors. Still, differentiation sometimes remains challenging while it is crucial for therapeutic decision making.

Whole genome methylation-based tumor classification is increasingly used for tumor classification, and the WHO recommends the routine application in tumor of the central nervous system (7). DNA methylation is a covalent modification of cytosine residues and is involved in gene expression regulation. Hypermethylation of specific gene promotor regions can lead to transcriptional inactivation, including tumor suppressor genes (8, 9). Besides somatically acquired DNA methylation changes, the cancer methylome harbors characteristics reflecting the cell of origin (10). This tissue specificity is what makes DNA methylation profiling well suited for the identification of tumor origin (10, 11). Furthermore, archival formalin-fixed, paraffin embedded (FFPE) tissues can be used for DNA methylation analysis. Based on this rationale numerous classifiers have been developed for cancer classification and some are routinely used in daily practice (11-16). Similarly, Hackeng et al. developed a classifier for distinguishing neuroendocrine tumors from different locations, including pancreatic NETs (17). Previous studies including Jäkel et al. and Benhamida et al., showed methylation profiles of pancreatic tumors are well suited for the differentiation of these tumors (18, 19). Together, these data suggest potential applicability of methylation profiling for classification of non-ductal pancreatic tumors.

To facilitate diagnosis within non-ductal pancreatic tumors ACC, SPN, PanNEN and PB, here we have built a methylation-based prediction model. We evaluated random forest classifiers, neural networks and gradient boosting machines and present an approach to distinguish between non-ductal pancreatic cancers with almost perfect accuracy.

# Materials and methods

**Study design**We build a diagnostic classifier able to distinguish PanNENs, ACCs, SPNs, PBs and PDACs based on their DNA methylation profiles. DNA methylation array data from PanNETs (n= 87), ACCs (n=37), SPNs (n= 13) and PDACs (n=..) that where publicly available was split into a training (76 primaries) and test cohort (139 primaries and metastases). We trained three machine learning algorithms (Random Forest (RF), artificial neural network (ANN), gradient boosting machines (xgBoost) on the training cohort. The three classifiers were evaluated using the test cohort. The three algorithms were compared and the best performing method was chosen for further analysis. We further validated our classifier on a clinical cohort compiled of DNA methylation array data obtained from FFPE tissue from PanNETs (n= 87), ACCs (n=37) and SPNs (n= 13) primaries and metastases obtained from the University Medical Center Utrecht and Radboud Medical Center pathological archives.

**Patients and samples**The University Medical Center Utrecht (UMCU) Biobank Research Ethics Committee approved the use of archival material for this study. Data used for the test and training cohort were retrieved from previously published datasets GSE117852 and GSE155353 from Gene Expression Omnibus (GEO) (20, 21), E-MTAB-7924 from EMBL-EBI (22), EGAD00010001298 from European Genome-Phenome Archive (EGA) (19), and SPN methylation data used in Selenica et al.(23) was requested by the authors. The clinical cohort included previously published data from dataset EGAS00001004878 from European Genome-Phenome Archive (EGA) (17), and additional cases selected from the University Medical Center Utrecht and Radboud Medical Center pathological archives. These methylation data have been deposited at the EGA, which is hosted by the European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG), under accession number EGA…..

**Whole genome DNA methylation analysis**On cases selected from our pathology archives, DNA extraction, bisulfite conversion and array processing were performed in-house at our core facility as previously described(17). For DNA extraction the DNA Methylation Bead-Chip manufacturer (Illumina) was used, formalin fixation–induced DNA damage was restored using the FFPE Restore Kit (Illumina) and DNA methylation data was obtained using the Infinium Methylation EPIC BeadChip (Illumina), according to protocols supplied by the manufacturer.

**Data analysis***Preprocessing* All analyses were performed using R version … (R Core Team. R: a language and environment for statistical computing. Available from: <https://www.R-project.org/>.). Raw 450K and EPIC data from all cohorts were separately imported using the minfi package (24) and data was normalized using ssNoob normalization (25). Failed probes with a detection P value >0.01 (n= 83712 (450K), 45443(EPIC)), probes targeting the sex chromosomes were removed (n= (450K), (EPIC)), probes containing a single-nucleotide polymorphism (dbSNP132 Common) within five base pairs of and including the targeted CpG site (n= (450K), (EPIC))/ SNP containing (n= (450K), (EPIC)) and cross reacting probes (n= (450K), (EPIC)) were removed. Betas values were obtained and both normalized 450K and EPIC beta values were merged resulting in … common probes.

*Batch effect* Batch effect was evaluated analyzing tissue types that occurred in two or more studies, these included normal tissue PanNET tissue. Average methylation values were compared, and unsupervised analysis (t-SNE and heatmaps) were made .

*Probe selection and unsupervised analysis*

5000 variable probes were selected for unsupervised analysis.

*Classifier development*

*Classifier validation*

# Results

Cohort description and characteristics (Figure 1)

* Number of samples

From 7 study we obtained 6 pancreatic tumor types and normal pancreatic tissue. (Figure 1B)

* mild batch effect (Figure S1)
* Tumor purity, Avg methylation etc  
  The average methylation was aprox … (Figure 1C, 1F)

Absolute tumor purity mostly above 0.5, exept for PDACs and some single ACC, PanNET, PanNEC, SPN cases (Figure 1D, Figure 1G)

* UMAP shows; shows some misclassifications; which can be explained by …

Classifier training (Figure 2)

* 10 cases from each study per tumor type (to reduce batch effect)
* Trained 3 algorithms
* Algorithm performance: NN/RF is best

Outlier detection (Figure 3)

Results UMCU dataset (Figure 4, Figure S2)

* Cohort description
* Biopsy vs resection
* Case report cases

# Discussion

This is the first study evaluating molecular classification of pancreatic tumors.

Batch effect

* small batch effect, unclear where it is caused by and what effect it will have on the classifier
* Other study do/do not evaluate batch effect

Verklaring noise region

* Misclassification
* Low tumor purity
* Bad samples

Classification performance

* the classifier correctly reclassified … samples initially diagnosed as … respectively. This reclassification is of profound clinical importance, as … show improved response to … while …. are associated with very poor prognosis.

SPN clusters with AML

* Corresponding markers CD56, CD1017, CD99
* We can not evaluate the expression of the Selenica SPN cases

Application of the algorithm

* Better not use for PDACS
* Outlier detection: more confidence in the DNA methylation-based classification

Limitations

* underlying biological mechanism of DNA methylation classification remain relatively unclear
* Online data; no central histopathological review, we cannot be 100% sure of the diagnosis of the online obtained data

# References (max 50)

1. Dhillon J. Non-Ductal Tumors of the Pancreas. Monogr Clin Cytol. 2020;26:92-108.

2. Hackeng WM, Hruban RH, Offerhaus GJ, Brosens LA. Surgical and molecular pathology of pancreatic neoplasms. Diagn Pathol. 2016;11(1):47.

3. Papavramidis T, Papavramidis S. Solid pseudopapillary tumors of the pancreas: review of 718 patients reported in English literature. J Am Coll Surg. 2005;200(6):965-72.

4. Schmidt CM, Matos JM, Bentrem DJ, Talamonti MS, Lillemoe KD, Bilimoria KY. Acinar cell carcinoma of the pancreas in the United States: prognostic factors and comparison to ductal adenocarcinoma. J Gastrointest Surg. 2008;12(12):2078-86.

5. Sonbol MB, Mazza GL, Mi L, Oliver T, Starr J, Gudmundsdottir H, et al. Survival and Incidence Patterns of Pancreatic Neuroendocrine Tumors Over the Last 2 Decades: A SEER Database Analysis. Oncologist. 2022;27(7):573-8.

6. Wisnoski NC, Townsend CM, Jr., Nealon WH, Freeman JL, Riall TS. 672 patients with acinar cell carcinoma of the pancreas: a population-based comparison to pancreatic adenocarcinoma. Surgery. 2008;144(2):141-8.

7. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. 2021;23(8):1231-51.

8. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene. 2002;21(35):5400-13.

9. Shen H, Laird PW. Interplay between the cancer genome and epigenome. Cell. 2013;153(1):38-55.

10. Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, et al. A DNA methylation fingerprint of 1628 human samples. Genome Res. 2012;22(2):407-19.

11. Moran S, Martinez-Cardus A, Sayols S, Musulen E, Balana C, Estival-Gonzalez A, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol. 2016;17(10):1386-95.

12. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018;555(7697):469-74.

13. Jurmeister P, Bockmayr M, Seegerer P, Bockmayr T, Treue D, Montavon G, 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(509).

14. Koelsche C, Schrimpf D, Stichel D, Sill M, Sahm F, Reuss DE, et al. Sarcoma classification by DNA methylation profiling. Nat Commun. 2021;12(1):498.

15. Leitheiser M, Capper D, Seegerer P, Lehmann A, Schuller U, Muller KR, 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(4):378-87.

16. Maas SLN, Stichel D, Hielscher T, Sievers P, Berghoff AS, Schrimpf D, et al. Integrated Molecular-Morphologic Meningioma Classification: A Multicenter Retrospective Analysis, Retrospectively and Prospectively Validated. J Clin Oncol. 2021;39(34):3839-52.

17. Hackeng WM, Dreijerink KMA, de Leng WWJ, Morsink FHM, Valk GD, Vriens MR, et al. Genome Methylation Accurately Predicts Neuroendocrine Tumor Origin: An Online Tool. Clin Cancer Res. 2021;27(5):1341-50.

18. Benhamida JK, Vyas M, Tanaka A, Wang L, Bahrami A, Ozcan K, et al. Pancreatoblastomas and mixed and pure acinar cell carcinomas share epigenetic signatures distinct from other neoplasms of the pancreas. Mod Pathol. 2022;35(7):956-61.

19. Jakel C, Bergmann F, Toth R, Assenov Y, van der Duin D, Strobel O, et al. Genome-wide genetic and epigenetic analyses of pancreatic acinar cell carcinomas reveal aberrations in genome stability. Nat Commun. 2017;8(1):1323.

20. Chan CS, Laddha SV, Lewis PW, Koletsky MS, Robzyk K, Da Silva E, et al. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. Nat Commun. 2018;9(1):4158.

21. Endo Y, Fujimoto M, Ito N, Takahashi Y, Kitago M, Gotoh M, 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(5):1341-54.

22. Di Domenico A, Pipinikas CP, Maire RS, Brautigam K, Simillion C, Dettmer MS, et al. Epigenetic landscape of pancreatic neuroendocrine tumours reveals distinct cells of origin and means of tumour progression. Commun Biol. 2020;3(1):740.

23. Selenica P, Raj N, Kumar R, Brown DN, Arques O, Reidy D, et al. Solid pseudopapillary neoplasms of the pancreas are dependent on the Wnt pathway. Mol Oncol. 2019;13(8):1684-92.

24. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.

25. Triche TJ, Jr., Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res. 2013;41(7):e90.