Supplement

Identification and validation of a multivariable prediction model based on blood plasma and serum metabolomics for the distinction of chronic pancreatitis subjects from non-pancreas disease control subjects

M. Gordian Adam#[1,2], Georg Beyer#[3], Nicole Christiansen[1,4], Beate Kamlage[1], Christian Pilarsky[5], Marius Distler[6], Tim Falbusch[7], Ansgar Chromik[8], Fritz Klein[9], Marcus Bahra[9], Waldemar Uhl[7], Robert Grützmann[5], F. Ulrich Weiss[10], Julia Mayerle#[3,10], and Markus M. Lerch#[10]*

[1] Metanomics Health GmbH, Tegeler Weg 33, 10589 Berlin, Germany
[2] Biocrates AG, Eduard-Bodem-Gasse 8, 6020 Innsbruck, Austria
[3] Department of Medicine II, University Hospital, Ludwig-Maximilians-Universität München, Marchioninistr. 15, 81377 Munich, Germany
[4] trinamiX GmbH, Industriestraße 31a, 67063 Ludwigshafen am Rhein, Germany
[5] Department of Surgery, University Hospital, Erlangen, Germany
[6] Clinic and Outpatient Clinic for Visceral, Thorax, and Vascular Surgery, Medical Faculty, TU Dresden, Dresden, Germany
[7] Ruhr-University Bochum, St. Josef Hospital, Department of Surgery, Bochum, Germany
[8] Asklepios Clinic Harburg, Department for General and Visceral Surgery, Hamburg, Germany
[9] Department of Surgery, Charité Campus Mitte and Charité Campus Virchow Klinikum, Charité-Universitätsmedizin Berlin, Berlin, Germany
[10] Department of Medicine A, University Medicine Greifswald, Ferdinand-Sauerbruchstrasse, 17475 Greifswald, Germany
* Correspondence:

Prof. Dr. med. Markus M. Lerch
Department of Medicine A
University Medicine Greifswald
Ferdinand-Sauerbruchstrasse
17475 Greifswald
Germany
Phone: +49 (0) 3834 86 7230
Fax: +49 (0) 3834 86 7234
Markus.lerch@med.uni-greifswald.de

Suppl. Material and Methods

Study details

We conducted a type 3 study for multivariable prediction for individual prognosis according to the TRIPOD guidelines [1]. Patients with chronic pancreatitis, liver cirrhosis, healthy blood donors and preoperative patients with non-pancreatic or liver disease were consecutively recruited from university referral centers in Greifswald, Dresden, Berlin, and Bochum, all in Germany.

For the identification study, EDTA plasma samples were collected within a case-control study from 80 patients with CP and 80 non-pancreatic control patients, who underwent small, non-pancreas-related surgical procedures under general anesthesia (see below).

For the first validation study, 144 chronic pancreatitis patients and 204 non-pancreatic control patients were recruited at three different centers. Because it was acknowledged that CP patients are relatively young and mostly male, during patient recruitment, special care was taken to also recruit younger and mostly male patients for the control group with the aim to achieve a similar average patient age and a similar ratio of male to female subjects in both groups.

In the second validation study, a different sample type was utilized: serum samples taken from 49 chronic pancreatitis patients, 56 controls, and 57 liver cirrhosis patients were analyzed. These samples were collected in a fourth independent center. Furthermore, the control group consisted of healthy blood donors instead of patients waiting to undergo a small surgery.

The key study dates for the three studies were as follows: accrual for the identification study was started on 2009-01-13, end of accrual was on 2013-08-01. Accrual for the first validation study was started on 2013-09-09 and ended on 2015-09-28. Accrual for the second validation study started on 2002-10-23 and ended on 2010-06-10.

The general inclusion criteria for all groups included written informed consent prior to any study procedures, age ≥ 18 to 85 years and eight hours fasting prior to blood draw.

The general exclusion criteria for all groups included type I diabetes, pregnancy or lactation phase, known viral infections like hepatitis B, hepatitis C, HIV, major surgery within the last 4 weeks before sample collection, acute anemia (Hb<9 g/dl or <5,58 mmol/l), malignant tumors within the last 5 years.

Chronic pancreatitis patients were included if one or more of the following criteria were met and no other diagnosis was more likely [2, 3, 4]: recurrent bouts of pancreatic pain with documented rise in amylase or lipase activity for a duration of more than one year plus radiological evidence supporting the diagnosis, pancreatic calcifications, histological proof

of chronic pancreatitis, unequivocal changes in pancreatic duct morphology, severely abnormal pancreatic function tests with maldigestion. Calcifications were identified on CT-scan, diabetes was diagnosed as suggested by the WHO definition and exocrine insufficiency was determined by either fecal elastase measurement or concurrent pancreatic enzyme supplementation. Pancreatitis patients were excluded if they had undergone pancreatitis surgery within 6 months before sample collection, bile duct stent placement or surgery, endoscopically assisted pancreatic aspiration <5 days before sample collection or had known liver cirrhosis.

Liver cirrhosis patients were included if preexisting liver cirrhosis had been diagnosed based on imaging and clinical chemistry. Liver cirrhosis patients were excluded if concomitant chronic pancreatitis was present.

Control patients were included if they were undergoing minor non-pancreatic surgery under general anesthesia. Control patients were excluded if they had chronic pancreatitis or liver cirrhosis or if a hernia was due to solid organ transplantation.

For the blood donors, only the standard blood donor inclusion criteria applied, i.e. the donors had to be in good general health, body weight at least 110 pounds. Participants with diabetes type II were excluded from the blood donor group because of the requirement of a fasting period of at least 8 hours which was not considered feasible for diabetics.

From the patients in the non-pancreatic control group in the identification study, 20 patients underwent vascular surgery, 18 patients received a hernia repair, 3 were resected for goiter and 39 received various other small surgical procedures under general anesthesia. From the patients in the non-pancreatic control group in the first validation study, 164 underwent hernia repair surgery and 40 were resected for thyroid goiter. None were operated in metabolically deranged state. Thus, the data in the first validation study differed from the identification study as it was a multicentric study, and in the composition of the non-pancreatic controls. The second validation study differed in the matrix used for analysis (serum instead of plasma), the center where the samples were obtained, the control group (healthy blood donors instead of non-pancreatic controls), and the inclusion of liver cirrhosis patients as an additional control group. Furthermore, 22.5% of the non-pancreatic controls in the identification study were diabetes type II patients, while 13.5% of the patients suffered from diabetes type II in the first validation study, and diabetes patients were excluded as control in the second validation study. As opposed to the identification study, the genesis of pancreatitis, calcifications, exocrine insufficiency, and enzyme supplementation were only partially available in the validation studies.

Sample Storage

Before freezing, the EDTA plasma samples and serum samples were aliquoted to avoid freeze-thaw cycles during the measurement period. Samples were stored at the respective center at -80°C until transport to the measurement location, which occurred on dry ice. Samples were stored at the measurement location at -80°C until measurement.

Our own work has shown that there were no significant differences in the metabolome of plasma samples due to storage when stored at -80°C for up to 7 years [5]. Even though the sample collection for one of the studies employed here started in 2003 already, all samples were measured within 7 years after sample collection. The longest sample storage time before measurement was 1637 days (about 4.5 years). Thus, a marked influence of freeze-thaw cycles or sample age on the results of this study can be excluded.

Metabolite profiling details

Briefly, proteins were removed from the samples by precipitation, using three volumes of acetonitrile. Polar and nonpolar fractions were separated by adding water and a mixture of ethanol and dichloromethane (2:1, v/v). For GC-MS analysis, the nonpolar fraction was treated with methanol under acidic conditions to yield the fatty acid methyl esters derived from both free fatty acids and hydrolyzed complex lipids. The polar and nonpolar fractions were further derivatized with O-methyl-hydroxylamine hydrochloride to convert oxo-groups to O-methyl-oximes, and subsequently with N-methyl-N-(trimethylsilyl)trifluoroacetamide prior to analysis. For LC-MS/MS analysis, both fractions were dried and reconstituted in appropriate solvent mixtures. High-performance liquid chromatography was performed by gradient elution using methanol/water/formic acid on reversed phase separation columns. Mass spectrometric detection technology was applied as described in patent WO2003073464 [6] which allows targeted and high-sensitivity multiple reaction monitoring (MRM) profiling in parallel to a full screen analysis. In brief, mass spectrometric detection was performed with repetitive cycles of MRM transitions for pre-selected metabolites followed by a full scan from a mass-to-charge ratio of 100 to 1000. The instrument was operated in positive ionization mode for metabolites in the nonpolar fraction, and in negative ionization mode for metabolites in the polar fraction. Metabolite identification was done by comparing sample data to authentic standards where applicable, as outlined previously [7].

MxP® Lipids covered profiling of sphingolipids (ceramides, sphingomyelins, and sphingobases). Total lipids were extracted from the sample by liquid/liquid extraction using chloroform/methanol. The lipid extracts were subsequently fractionated by normal phase liquid chromatography (NPLC) into different lipid groups according to [8, 9]. The fractions were analyzed by LC-MS/MS using electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) with detection of specific MRM transitions for preselected sphingolipids.

Metabolite profiling generated semi-quantitative data of metabolite concentrations calculated by determining metabolite levels in each study sample relative to metabolite concentrations in reference pool samples that were created from aliquots of all study samples. The normalization to reference pool samples compensates for inter- and intra-instrumental variation, i.e. variability that occurs when different analytical sequences were analyzed by different devices. To allow comparison of data sets between the different studies, the semi-quantitative data were further normalized to the median of MxPool™ samples representing a pool of commercial human EDTA plasma containing more than 2,000 different metabolites of known concentrations. A one-point calibration was used to calculate quantitative absolute concentrations for those metabolites present in the MxPool. Both types of pooled reference samples were run in parallel through the entire process.

Biomarker selection

The metabolites for the biomarker panel were nominated based on biomedical expertise. In a first step, features that markedly differentiate CP patients from controls that could have an influence on the metabolome were considered. CP patients frequently suffer from lipid malabsorption and gut microbiome changes due to reduced bile acid secretion, reduced endocrine pancreatic function, pancreatic tissue fibrosis, and pancreatic inflammation. In a second step, metabolite groups that were expected to be different between CP patients and controls based on these physiological differences were collected: nutritional lipids that would be affected from malabsorption, microbiome-derived metabolites that could be affected by gut microbiome changes, carbohydrate metabolites that that would be affected by the reduced endocrine function, metabolites that would be altered in response to fibrosis, and metabolites that would be altered in response to inflammatory processes. In a third step, single representative metabolites from these groups were chosen for the signature panel based on methodical experience (the metabolites needed to allow for robust measurements above the limit of detection), available literature, and experience from previous experiments with CP patients and controls.

Prediction model

One prediction model was employed for all three study cohorts, i.e. the beta coefficients obtained from the first cohort were then applied to the individuals from the other 2 cohorts. Our prediction model, consisting of the biomarker signature, the corresponding algorithm, and the established cut-off, predicts whether a patient suffers from chronic pancreatitis. The biomarker enables a clinical diagnosis, supporting the standard diagnostic means for

diagnosis of chronic pancreatitis (see above). The biomarker is not designed to be applied for screening of the general population.

To avoid any bias when analyzing the concentrations of the metabolites present in the biomarker signature, the diagnosis was blinded to the scientists measuring the samples using mass spectrometry. The concentration values in the plasma samples of the 8 metabolites present in the biomarker signature are the only predictors used in the prediction model. Furthermore, the calculation of the biomarker score by the algorithm and selection of the cut-off was done fully automated, without human interference. After the initial calculation based on the identification study results, there were no subsequent interventions like patient exclusions, cut-off optimization, or re-training of the algorithm. Vice versa, the clinical diagnosis was established in the participating clinical centers according to the criteria mentioned above before the plasma samples were taken and analyzed in this study. Thus, the outcome obtained with the prediction model did not have any effect on the clinical diagnosis.

Statistical analysis details

Power analysis was performed to estimate an adequate sample size using representative metabolite profiling standard deviations that were determined by metanomics GmbH in earlier studies. Primary goal of the study was to determine a 20% metabolic difference on a 5% significance level with approximately 72-99% power on the basis of the patient samples. Metabolic difference was defined as absolute or relative difference in concentrations of individual metabolites. Power estimates were based on t-test statistics.

Missing data were handled differently depending on the analysis. For the Naive Bayes algorithm and the principal component analysis, missing values were imputed with the NIPALS (Non-linear Iterative PArtial Least Squares) algorithm [10]. In the second validation study, BMI values were not available for all participants. For the inclusion of the BMI as a confounding factor in the MANOVA, the missing BMI values were imputed using K-means clustering [11] for this purpose.

All R packages used, sorted by analysis step and including the utilized functions, are listed in supplemental **Table S1**.

Suppl. Results

Metabolomic analyses details

The metabolomics data underwent a strict quality control after which 505 known and 115 unknown metabolites remained for statistical analysis in the datasets based on plasma samples. Most of these metabolites could also be detected in the study conducted with serum samples. In this dataset, 498 known and 118 unknown metabolites remained for statistical analysis that met the quality control criteria.

Concentration data were missing for beta-carotin from 2 samples in the second validation study, for cryptoxanthin in 7 samples from the second validation study, for N-acetylcytidine in 2 samples from the first and 1 sample from the second validation study, for behenic acid in 1 sample from the identification study, 6 samples from the first validation study, and 1 sample in the second validation study, for mannose in 16 samples from the first validation study and 5 samples from the second validation study, for indole-3-acetic acid for 1 sample in the identification study, for 28 samples in the first validation study, and for 7 samples in the second validation study, for hippuric acid for 1 sample in the second validation study, and for ceramide (d18:1,C24:1) for 1 sample in the first validation study and 17 samples in the second validation study. Normalized to the number of samples in each study, this means that a maximum of 1% of the values for a given metabolite were missing in the identification study, maximally 8% in the first validation study and maximally 10% in the second validation study. There were no cases where the outcome (diagnosis) was unknown or missing.

In total, 60 metabolites were not significantly different ($p > 0.05$ or $q > 0.2$) between CP and control groups in any of the three studies. 516 metabolites were significantly different in some of the studies and 39 metabolites were significantly ($p < 0.05$ and $q < 0.2$) different in all of the three studies. 6 of the 8 metabolites from the signature metabolite panel had a p value below the significance threshold ($p < 0.05$) in all three studies (see **Table 4**). The other 2 of the 8 metabolites from the panel (behenic acid and indole-3-acetic acid) were significantly different between the groups in the identification and the first validation study, but not in the second validation study that had serum as a sample matrix.

The panel metabolites were not the best 8 discriminators between CP patients and controls. Beta-carotene and cryptoxanthin were among the top 3 discriminators in the plasma-based studies (identification and first validation study), with lycopene being the best discriminator in the plasma studies. Looking at all three studies together, beta-carotene, cryptoxanthin, and mannose were among the top 5 discriminators, with 3-hydroxybutyrate being the best discriminator.

The distribution of age and BMI over the biomarker signature score is shown in supplemental **Figure S1**. The age gap between CP patients and non-pancreatic controls is markedly higher in the identification study than in the validation. In the first validation study, the age of the patients follows an even Gaussian distribution for both CP patients and non-

pancreatic controls. In both studies, the score is markedly higher for CP patients compared to non-pancreatic controls irrespective of the age. As the BMI was calculated with a decimal, there are a lot of potential sublevels, which results in more data being needed for Gaussian curves. Nevertheless, the BMI of non-pancreatic controls in the validation study also follows a Gaussian distribution, while the BMI of CP patient is clearly skewed due to the increased frequency of patients with low BMI. This is an inherent feature of the disease concomitant with the malnutrition caused by CP. These trends can also be observed in the identification study. Despite the uneven BMI distribution, the graphs show that the biomarker score is markedly higher for CP patients compared to non-pancreatic controls irrespective of the BMI.

The full prediction model can be used universally. The weighting of the metabolites as shown in supplemental **Table S2** can be used as coefficients to be multiplied with the respective concentrations of the eight metabolites in the biomarker signature (in µmol / L) to calculate the biomarker signature score. Whether the score is above or below the cut-off value of 0.479 determines whether the patient is evaluated as positive or negative for the diagnosis “chronic pancreatitis”.

Effect of exocrine insufficiency and enzyme supplementation on carotenoid levels

Because the identification of beta-carotene and cryptoxanthin suggested a pathophysiological mechanism of malabsorption, it was analyzed whether pancreatic exocrine insufficiency and enzyme supplementation had an effect on plasma carotenoid levels. This analysis was limited to the identification study because the full information was available for this cohort only. Almost all patients with exocrine insufficiency also received enzyme supplementation so that a separate comparison of the effect of exocrine insufficiency alone was not possible. As obvious from **Figure S2**, there was no significant increase of carotenoid levels in plasma of chronic pancreatitis patients supplemented with enzymes to treat exocrine insufficiency.

Biomarker signature score increases with disease severity

In clinical daily routine, patients that will are tested for chronic pancreatitis are not always as healthy as the control groups used in this study. We therefore wanted to investigate whether the biomarker signature score was more accurate in patients with advanced disease than in less severe cases. Because other clinical data regarding disease severity was elusive, we used the information whether the patients suffered from pancreatic endocrine or exocrine insufficiency, which is a good surrogate marker for severity and time

since disease onset. 199 chronic pancreatitis patients from the identification and the first validation study for whom this information was available together were categorized in three groups: those without insufficiencies, those with either endo- or exocrine insufficiency, and those with both endo- and exocrine insufficiencies. The distribution of biomarker signature score values in the three groups is shown in **Figure S3**. The average biomarker signature score was 0.68 in patients without insufficiencies, 0.78 in patients with either endo-or exocrine insufficiency, and 0.90 in patients with both endo- and exocrine insufficiency. An ANOVA was employed to test whether the differences in the biomarker signature score were significant. While the group with one pancreatic insufficiency did not have a significantly different score compared to the other groups, the scores of the groups without pancreatic insufficiencies and with both endo- and exocrine insufficiencies were significantly different ($p = 0.0018$). This indicates that the biomarker signature score is higher in patients with more severe pancreatic disease.

Using the established cut-off of 0.479, we concluded for each patient in the three groups whether the diagnosis based on the signature score was correct or a false negative. A chi-squared test was employed to investigate whether the diagnosis and the severity were co-dependent. The result of $p = 0.056$ shows a clear trend towards a higher fraction of correct diagnoses in more severe cases, although there was no significant dependency between the group affiliation and the diagnosis.

References

1 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015;**102**:148-58.

2 Mayerle J, Hoffmeister A, Werner J, Witt H, Lerch MM, Mossner J. Chronic pancreatitis--definition, etiology, investigation and treatment. Dtsch Arztebl Int 2013;**110**:387-93.

3 Ito T, Ishiguro H, Ohara H, Kamisawa T, Sakagami J, Sata N, *et al.* Evidence-based clinical practice guidelines for chronic pancreatitis 2015. J Gastroenterol 2016;**51**:85-92.

4 Whitcomb DC, Shimosegawa T, Chari ST, Forsmark CE, Frulloni L, Garg P, *et al.* International consensus statements on early chronic Pancreatitis. Recommendations from the working group for the international consensus guidelines for chronic pancreatitis in collaboration with The International Association of Pancreatology, American Pancreatic Association, Japan Pancreas Society, PancreasFest Working Group and European Pancreatic Club. Pancreatology 2018.

5 Wagner-Golbs A, Neuber S, Kamlage B, Christiansen N, Bethan B, Rennefahrt U, *et al.* Effects of Long-Term Storage at -80 °C on the Human Plasma Metabolome. Metabolites 2019;**9**.

6 Walk TB, Dostler M. Mass Spectrometry Method for Analysing Mixtures of Substances. 2003:PCT/EP2003/001274.

7 Kamlage B, Neuber S, Bethan B, González Maldonado S, Wagner-Golbs A, Peter E, *et al.* Impact of Prolonged Blood Incubation and Extended Serum Storage at Room Temperature on the Human Serum Metabolome. Metabolites 2018;**8**.

8 Christie WW. Rapid separation and quantification of lipid classes by high performance liquid chromatography and mass (light-scattering) detection. J Lipid Res 1985;**26**:507-12.

9 Schmidt H, Schmidt R, Geisslinger G. LC-MS/MS-analysis of sphingosine-1-phosphate and related compounds in plasma samples. Prostaglandins Other Lipid Mediat 2006;**81**:162-70.

328 10 Noonan R, Wold H. NIPALS Path Modelling with Latent Variables. Scandinavian Journal of Educational
329 Research 1977;**21**:33-61.
330 11 Gajawada S, Toshniwal D. Missing Value Imputation Method Based on Clustering and Nearest
331 Neighbours. International Journal of Future Computer and Communication 2012;**1**:pp. 206-8.
332

Suppl. Figures and Tables:

Figure S1. Distribution of age (A,C) and BMI (B,D) over biomarker signature score in the identification study (A,B) and the first validation study (C,D). Demographics are shown on the x axes, the biomarker signature score on the y axes. Stacked columns for age and side-by-side columns for BMI are colored according to outcome (diagnosis). Horizontal lines / functions represent Gaussian approximation of the data.

Figure S2. Boxplots of effect of enzyme supplementation on plasma carotenoid levels. Shown are data for chronic pancreatitis patients of the identification study (plasma). There was no significant increase of carotenoid levels in plasma of patients supplemented with enzymes to treat exocrine insufficiency.

Figure S3. Boxplots of biomarker signature scores in chronic pancreatitis patients from the identification study and the first validation study, which were categorized depending on whether they suffer from endo- and/or exocrine insufficiencies as a measure of disease stage. The average biomarker signature score increased with disease severity and this increase was significant comparing patients without insufficiency and patients with both endo- and exocrine insufficiency.

352 **Table S1**: R packages and functions used

Analysis	Packages	Functions
PCA	Stats (included in R core pkge)	prcomp()
Biomarker training and test	E1071 ROCR ROCR pROC	naivebayes() performance() prediction() roc()
ANOVA/MANOVA	slme stats (included in R core pkge)	lme() p.adjust()
Visualizations	ggplot2	

353

354

355

356 **Table S2**: Weightings of all metabolites in the biomarker signature

Metabolite	**Weight**	**Mean**	**SD ±**
Beta-carotene	1.1749422	-0.9281	0.4641
Cryptoxanthin	1.2061438	-1.1714	0.4558
Mannose	0.7865473	1.6548	0.1678
Behenic acid (C22:0)	0.6728898	1.5537	0.2095
Ceramide (d18:1/ C24:1)	0.6066595	0.1294	0.1427
Indole-3-acetic acid	0.5846360	-0.0213	0.3273
Hippuric acid	0.5009786	-0.3020	0.4408
N-Acetylcytidine	0.1068720	0.0807	0.1897

357 SD: standard deviation

358 **Table S3**. Statistical analysis results (linear model) of the signature metabolites (CP vs.
359 control) with FDR corrected p-values (q values).

Dataset (matrix)	Identification study (plasma)	First validation study (plasma)	Second validation study (serum)
Metabolite	**Fold change (q-value)**	**Fold change (q-value)**	**Fold change (q-value)**
Beta-carotene	0.37 (9.8×10^{-6})	0.38 (1.8×10^{-19})	0.47 (0.00039)
Cryptoxanthin	0.40 (9.9×10^{-6})	0.38 (4.5×10^{-18})	0.48 (0.00090)
Mannose	1.47 (9.9×10^{-6})	1.52 (6.7×10^{-27})	2.80 (8.4×10^{-13})
Behenic acid (C22:0)	0.72 (0.0080)	0.67 (1.0×10^{-15})	1.14 (0.14)
Ceramide (d18:1, C24:1)	1.27 (0.0059)	1.28 (1.0×10^{-10})	1.96 (5.9×10^{-12})
Indole-3-acetic acid	0.63 (0.013)	0.62 (2.4×10^{-8})	1.15 (0.40)
Hippuric acid	0.68 (0.15)	0.49 (1.2×10^{-7})	0.51 (0.0034)
N-Acetylcytidine	1.22 (0.058)	1.21 (8.1×10^{-5})	2.36 (3.2×10^{-12})

360 CP: Chronic pancreatitis

361