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,4 Beate Kamlage,1 Christian Pilarsky,4 Marius Dirstler,5 Tim Fahldus,6 Ansgar Chromik,7 Fritz Klein,8 Marcus Bahra, 9 Waldemar Uhl, 6 Robert Grützmann, 4 Ujjwal M Mahajan, 2, 9, 2 Frank U Weiss, 9, 9 Julia Mayerle, 9, 2 Markus M Lerch 9

ABSTRACT
Objective Chronic pancreatitis (CP) is a fibroinflammatory syndrome leading to organ dysfunction, chronic pain, an increased risk for pancreatic cancer and considerable morbidity. Due to a lack of specific biomarkers, diagnosis is based on symptoms and specific but insensitive imaging features, preventing early diagnosis and appropriate management.

Design We conducted a type 3 study for multivariable prediction for individual prognosis according to the TRIPOD guidelines. A signature to distinguish CP from controls (n=160) was identified using gas chromatography-mass spectrometry and liquid chromatography-tandem mass spectrometry on ethylidenaminetetraacetic acid (EDTA)-plasma and validated in independent cohorts.

Results A Naive Bayes algorithm identified eight metabolites of six ontology classes. After algorithm training and computation of optimal cut-offs, classification according to the metabolic signature detected CP with an area under the curve (AUC) of 0.85 (95% CI 0.79 to 0.91). External validation in two independent cohorts (total n=502) resulted in similar accuracy for detection of CP compared with non-pancreatic controls in EDTA-plasma (AUC 0.85 (95% CI 0.81 to 0.89)) and serum (AUC 0.87 (95% CI 0.81 to 0.95)).

Conclusions This is the first study that identifies and independently validates a metabolomic signature in plasma and serum for the diagnosis of CP in large, prospective cohorts. The results could provide the basis for the development of the first routine laboratory test for CP.

INTRODUCTION
Chronic pancreatitis (CP) is an inflammatory syndrome of the pancreas in which repetitive episodes result in fibrotic tissue replacement, organ dysfunction and chronic pain. Multiple aetiologies and risk factors lead to the development of CP in humans, which include moderate alcohol consumption, 1 tobacco smoke and variety of genetic predispositions. 4, 5, 6 Affected patients are at high risk for developing exocrine pancreatic insufficiency leading to malnutrition, as well as endocrine insufficiency, leading to diabetes mellitus type 3c and pancreatic cancer. 7 Chronic pain, malnutrition and brittle diabetes will often lead to a significant reduction in quality of life, increased healthcare utilisation and reduced life expectancy. 8, 9, 10, 11
In spite of multiple aetiologies and pathogeneses most CP patients will develop a similar clinical and histopathological phenotype. Following current guidelines, the diagnosis of CP is made when the following criteria are met: recurrent bouts of pancreatic pain with documented rise in amylase or lipase activity, and imaging evidence such as pancreatic calcifications, histological evidence of CP, unequivocal changes in pancreatic duct morphology or severely abnormal pancreatic function tests with malabsorption. Sensitivity and specificity of imaging in CP is variable and stage dependent. Non-invasive biomarkers for diagnosis of CP have not been established and remain an unmet clinical need. Moreover, it is presently impossible to identify patients at risk, or at an early stage of CP, due to a lack of reliable biomarkers. While numerous genomic studies in large cohorts have identified a growing number of genetic modifiers and risk factors, transcriptome or proteome-based approaches have failed to produce robust diagnostic tools for CP.

Little is known about the potential role of metabolomic signatures including lipidomics of body fluids as a diagnostic tool for CP. We have recently identified and validated a metabolomic signature to distinguish between pancreatic cancer and CP. This prompted the search for a biomarker signature that can discriminate between CP and controls including patients with non-pancreatic conditions.

**MATERIALS AND METHODS**

**Study design**

We conducted a type 3 study for multivariable prediction for individual prognosis according to the TRIPOD guidelines. A total of 670 patients and controls were prospectively enrolled in the three cohorts and the analysis was done retrospectively. The diagnosis ‘CP’ was made if one or more of the clinical and imaging criteria described by Mayerle et al were met and no other diagnosis was more likely. Similar distribution of age and gender between patients and controls was attempted. Figure 1 and table 1 comprise an overview over the identification and validation cohorts.

The identification study was performed in a case control cohort at a university referral centre and included 80 CP patients and 80 non-pancreatic disease controls who underwent small, non-pancreas-related surgical procedures under general anaesthesia (table 1 and online supplemental methods).

For the first validation study, 144 CP patients and 204 non-pancreatic controls were consecutively recruited from three different university referral centres.

For the second validation study, conducted to validate the robustness of the method using serum instead of plasma samples, CP, liver cirrhosis patients and healthy blood donors were enrolled at a fourth referral centre. In this cohort, some blood samples had low sample quality, identified by their extremely low glucose levels (<2800 μmol/L), and were consecutively excluded from analysis. The samples from 49 CP patients, 57 liver cirrhosis patients and 56 healthy blood donors were included in the final analysis.

Clinical information on gender, age and body mass index (BMI) were recorded. In addition, information on the disease aetiology, disease duration, calcifications, endocrine insufficiency and enzyme supplementation were recorded if available. All cohort characteristics can be found in table 1.

All participants gave their written informed consent prior to inclusion. Further information on study details and sample processing can be found in online supplemental material and methods.

**Metabolite profiling**

All samples were analysed with MxP Global Profiling and MxP Lipids. MxP Global Profiling was performed employing (1) gas chromatography-mass spectrometry (GC-MS) using an Agilent 6890 gas chromatograph coupled to an Agilent 5973 mass-selective detector and (2) liquid chromatography-tandem mass spectrometry (LC-MS/MS) using an Agilent 1100 high-performance liquid chromatography system coupled to an Applied Biosystems API 4000 triple quadrupole mass spectrometer, as has been described in detail before.

Up to 1449 metabolites were detected within the studies depending on the sample matrix and the analytical technique. The metabolites originated from 10 different ontology classes and comprised 838 known metabolites and 611 unknown spectral features. Only those metabolites that met specific quality criteria as described in were included in further statistical analyses. Furthermore, quality assessment of plasma samples was performed using the MxP Biofluids Quality Control assay (see our patent application WO2015145387A1).

**Statistics**

All metabolite profiling data were log10-transformed before further analysis to achieve an approximate normal distribution. R V3.3.4 was used for data analyses, see online supplemental methods for a list of R packages used (online supplemental table S1).

For an exploratory multivariate analysis (principal component analysis, PCA), the log10-transformed data of the identification and first validation study were centred and scaled to unit

---

**Figure 1** Study design. The biomarker signature was identified on the metabolomic data from the identification study, comparing chronic pancreatitis patients (CP) with control patients. These data were used as a training set for the algorithm. Participants of the first validation study were recruited independently and their sample data served as a test set. For the second validation study, participants were recruited independently as well. In this study, liver cirrhosis patients (LC) were included as an additional control group.
The log10-transformed, scaled and imputed ratios from the second validation study were also used for a PCA, which was visualised separately because of the different sample matrix. TIBCO Spotfire V7.12.0 was used to visualise the PCs.

To differentiate between CP patients and controls depending on their metabolic profiles, a Naive Bayes model was fitted using the log10-transformed, median-imputed, centred and scaled data from the identification study. Based on biomedical expertise a panel was nominated (see online supplemental methods). An algorithm was trained with the data on the selected panel metabolites. The fitted model was evaluated with 10-fold cross-validation. Optimal coefficients were determined, and an optimal cut-off based on the criteria of a sensitivity of 0.8 was calculated in order to classify the patients.

To validate the generated model for patient classification, the algorithm was applied to log10-transformed, centred and scaled data from the first validation set. For scaling of the first validation dataset, the mean and SD of the identification dataset were applied. A prediction score was calculated for each patient. For a successful validation, we expected the confidence intervals for the performance (AUC) to overlap between the training and the test set.

### Advancement criteria
The goal of our study was to design a biomarker that can discriminate between CP patients and controls with an AUC of at least 0.8. In addition, the specificity needed to be higher than the sensitivity because guidelines emphasise that diagnosis of CP comes with great burden and may induce stigmatisation for the patient. For a successful validation, we expected the confidence intervals for the performance (AUC) to overlap between the training and the test set.

### RESULTS
#### Study cohorts and patient characteristics
Clinical characteristics of all cohorts are shown in table 1. All studies were unbalanced between male and female participants.

### Variance and missing values
Variance, and missing values were imputed before the analysis (see online supplemental methods).

The log10-transformed, scaled and imputed ratios from the validation study were used for a PCA, which was visualised separately because of the different sample matrix. TIBCO Spotfire V7.12.0 was used to visualise the PCs.

First Validation Study (plasma)

<table>
<thead>
<tr>
<th>Male</th>
<th>Age, years (mean±SD, range)</th>
<th>51 (±11.5, 22–79)</th>
<th>52 (±14.8, 24–90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI, kg/m² (mean±SD, range)</td>
<td>23.6 (±3.70, 15.7–34.8)</td>
<td>27.3 (±5.16, 16.4–48.9)</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>28%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>% alcoholic CP: 43%</td>
<td>% alcoholic CP: 68% (unknown in 29)</td>
<td>% alcoholic CP: 60% (unknown in 17)</td>
<td></td>
</tr>
<tr>
<td>% PERT: 60% (unknown in 17)</td>
<td>Mean disease duration, years: 7.3±8.55 (unknown in 36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean disease duration, years: 13±6.16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Second Validation Study (serum)

<table>
<thead>
<tr>
<th>Male</th>
<th>Age, years (mean±SD, range)</th>
<th>47 (±11.0, 25–71)</th>
<th>54.6 (±11.3, 25–68)</th>
<th>59.4 (±11.3, 28–82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI, kg/m² (mean±SD, range)</td>
<td>24.7 (±4.1, 17–34, unknown in 37)</td>
<td>27.6 (±4.1, 20–36)</td>
<td>28.6 (±4.1, 19–38, unknown in 32)</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>12%</td>
<td>0%</td>
<td>11%</td>
<td></td>
</tr>
<tr>
<td>% alcoholic CP: 100%</td>
<td>% alcoholic CP: 60% (unknown in 34)</td>
<td>% alcoholic CP: 18% (unknown in 28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% PERT: 16% (unknown in 32)</td>
<td>Mean disease duration, years: 1.2±1.85 (unknown in 35)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 1 Cohort characteristics

<table>
<thead>
<tr>
<th>Identification study (plasma)</th>
<th>Chronic pancreatitis</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>76%</td>
<td>52%</td>
</tr>
<tr>
<td>Age, years (mean±SD, range)</td>
<td>51 (±9.9, 20–73)</td>
<td>64 (±14.5, 22–88)</td>
</tr>
<tr>
<td>BMI, kg/m² (mean±SD, range)</td>
<td>23.1 (±4.82, 14.5–50.0)</td>
<td>26.8 (±4.80, 17.6–42.3)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>44% (unknown in six patients)</td>
<td>23%</td>
</tr>
<tr>
<td>% alcoholic CP: 75%</td>
<td>% alcoholic CP: 68% (unknown in 29)</td>
<td>% alcoholic CP: 60% (unknown in 17)</td>
</tr>
<tr>
<td>% PERT: 60% (unknown in 17)</td>
<td>Mean disease duration, years: 13±16.91</td>
<td>Mean disease duration, years: 13±6.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First Validation Study (plasma)</th>
<th>Blood donors</th>
<th>Liver cirrhosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>83%</td>
<td>66%</td>
</tr>
<tr>
<td>Age, years (mean±SD, range)</td>
<td>51 (±11.5, 22–79)</td>
<td>52 (±14.8, 24–90)</td>
</tr>
<tr>
<td>BMI, kg/m² (mean±SD, range)</td>
<td>23.6 (±3.70, 15.7–34.8)</td>
<td>27.3 (±5.16, 16.4–48.9)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>28%</td>
<td>13%</td>
</tr>
<tr>
<td>% alcoholic CP: 43%</td>
<td>% alcoholic CP: 68% (unknown in 29)</td>
<td>% alcoholic CP: 60% (unknown in 17)</td>
</tr>
<tr>
<td>% PERT: 60% (unknown in 17)</td>
<td>Mean disease duration, years: 7.3±8.55 (unknown in 36)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second Validation Study (serum)</th>
<th>Blood donors</th>
<th>Liver cirrhosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>86%</td>
<td>68%</td>
</tr>
<tr>
<td>Age, years (mean±SD, range)</td>
<td>47 (±11.0, 25–71)</td>
<td>54.6 (±11.3, 25–68)</td>
</tr>
<tr>
<td>BMI, kg/m² (mean±SD, range)</td>
<td>24.7 (±4.1, 17–34, unknown in 37)</td>
<td>27.6 (±4.1, 20–36)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>12%</td>
<td>0%</td>
</tr>
<tr>
<td>% alcoholic CP: 100%</td>
<td>% alcoholic CP: 60% (unknown in 34)</td>
<td>% alcoholic CP: 18% (unknown in 28)</td>
</tr>
<tr>
<td>% PERT: 16% (unknown in 32)</td>
<td>Mean disease duration, years: 1.2±1.85 (unknown in 35)</td>
<td></td>
</tr>
</tbody>
</table>

**BMI**, body mass index; CP, chronic pancreatitis; LC, liver cirrhosis; NA, not applicable; PEL, pancreatic enzyme insufficiency; PERT, pancreatic enzyme replacement therapy.

**RESULTS**

**Study cohorts and patient characteristics**

Clinical characteristics of all cohorts are shown in table 1. All studies were unbalanced between male and female participants.

**First Validation Study**

- Male: 83%
- Age, years (mean±SD, range): 51 (±11.5, 22–79) vs. 52 (±14.8, 24–90)
- BMI, kg/m² (mean±SD, range): 23.6 (±3.70, 15.7–34.8) vs. 27.3 (±5.16, 16.4–48.9)
- Diabetes: 28% vs. 13%

**Second Validation Study**

- Male: 86%
- Age, years (mean±SD, range): 47 (±11.0, 25–71) vs. 54.6 (±11.3, 25–68) vs. 59.4 (±11.3, 28–82)
- BMI, kg/m² (mean±SD, range): 24.7 (±4.1, 17–34, unknown in 37) vs. 27.6 (±4.1, 20–36) vs. 28.6 (±4.1, 19–38, unknown in 32)
- Diabetes: 12% vs. 0% vs. 11%

**Cohort characteristics**

Table 1. Cohort characteristics

<table>
<thead>
<tr>
<th>Identification study (plasma)</th>
<th>Chronic pancreatitis</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>76%</td>
<td>52%</td>
</tr>
<tr>
<td>Age, years (mean±SD, range)</td>
<td>51 (±9.9, 20–73)</td>
<td>64 (±14.5, 22–88)</td>
</tr>
<tr>
<td>BMI, kg/m² (mean±SD, range)</td>
<td>23.1 (±4.82, 14.5–50.0)</td>
<td>26.8 (±4.80, 17.6–42.3)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>44% (unknown in six patients)</td>
<td>23%</td>
</tr>
<tr>
<td>% alcoholic CP: 75%</td>
<td>% alcoholic CP: 68% (unknown in 29)</td>
<td>% alcoholic CP: 60% (unknown in 17)</td>
</tr>
<tr>
<td>% PEL: 60% (unknown in 17)</td>
<td>Mean disease duration, years: 13±16.91</td>
<td>Mean disease duration, years: 13±6.16</td>
</tr>
</tbody>
</table>

**First Validation Study (plasma)**

- Male: 83%
- Age, years (mean±SD, range): 51 (±11.5, 22–79) vs. 52 (±14.8, 24–90)
- BMI, kg/m² (mean±SD, range): 23.6 (±3.70, 15.7–34.8) vs. 27.3 (±5.16, 16.4–48.9)
- Diabetes: 28% vs. 13%

**Second Validation Study (serum)**

- Male: 86%
- Age, years (mean±SD, range): 47 (±11.0, 25–71) vs. 54.6 (±11.3, 25–68) vs. 59.4 (±11.3, 28–82)
- BMI, kg/m² (mean±SD, range): 24.7 (±4.1, 17–34, unknown in 37) vs. 27.6 (±4.1, 20–36) vs. 28.6 (±4.1, 19–38, unknown in 32)
- Diabetes: 12% vs. 0% vs. 11%
Pancreas

due to the nature of CP, which is much more frequent in men than in women.27

Metabolomic analyses in samples from CP patients and non-pancreatic controls
The metabolomics data underwent a strict quality control after which 505 known and 115 unknown metabolites from plasma samples, and 498 known and 118 unknown metabolites from serum samples remained for statistical analysis. Most of the metabolites could be detected in both plasma and serum samples. Their distribution over the ontology classes is shown in figure 2.

Discrimination of CP and control patients by multivariate statistics
We investigated whether metabolic profiles of CP and control patients could generally be separated in an unsupervised multivariate statistical approach.

A PCA was performed on all plasma sample data. The best separation of groups was obtained in the principal components (PC) 1 and 2, which accounted for 12% and 7% of the whole variance of the dataset, respectively (figure 3A). The two groups showed a major overlap but samples from CP patients had a tendency towards lower scores in PC1, which was remarkable for a heterogeneous cohort with high interindividual variability due to diverse lifestyles, medications and comorbidities.

In the PCA obtained in the second validation study, liver cirrhosis patients were added as a third diagnosis group in addition to CP and controls (figure 3B). The best separation between the groups was again observed in PC1 and PC2 (21% and 13% of the observed variance). Remarkably, an almost complete separation of the control group from the other two could be observed. The CP patients tended to have higher scores in the PC2 than the liver cirrhosis patients, resulting in a visible separation between these groups.

While it is not common to perform a PCA in the validation cohorts, it was done in this case to show that the metabolomic profiles and the distribution of the CP versus control group patients in the identification study and the first validation study are very similar, proving that the two studies, although conducted independently, are actually comparable. The PCA of the second validation study yields a good overview on how the metabolic profile of the liver cirrhosis patients relate to the metabolic profiles of CP patients and controls.

Biomarker discovery and training
The observed separation tendencies in our multivariate approach indicated the possibility to compile a biomarker signature that allowed differentiation between CP and control patients.

As result of the Naive Bayes algorithm and biomedical expertise applied on the identification study a panel of eight metabolites (table 2) was nominated. Using the optimal calculated cut-off of 0.479 of the prediction score, the biomarker signature detected CP in comparison to control patients with an AUC of 0.85 (95% CI 0.79 to 0.91) as explained in online supplemental...
The specificity was 0.86 and the sensitivity 0.71 (table 3); ROC curves see figure 4A.

Table 2: List of metabolites selected for the biomarker signature

<table>
<thead>
<tr>
<th>Metabolite name</th>
<th>Ontology class</th>
<th>Biological background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-carotene</td>
<td>Vitamins, cofactors and related</td>
<td>Exogenous compound, lipid malabsorption</td>
</tr>
<tr>
<td>Cryptoxanthin</td>
<td>Miscellaneous (Subclass: diet related)</td>
<td>Exogenous compound, lipid malabsorption</td>
</tr>
<tr>
<td>Behenic acid (C22:0)</td>
<td>Complex lipids, fatty acids and related (Subclass: fatty acids)</td>
<td>Poorly absorbable, de novo synthesis, ceramide related</td>
</tr>
<tr>
<td>Indole-3-acetic acid</td>
<td>Amino acids and related</td>
<td>Altered bioavailability due to microbiome changes</td>
</tr>
<tr>
<td>Hippuric acid</td>
<td>Miscellaneous (Subclass: microbiome related)</td>
<td>Altered blood glucose levels due to impaired endocrine function inhibits mannose conversion to glucose.</td>
</tr>
<tr>
<td>Mannose</td>
<td>Carbohydrates and related</td>
<td>Altered blood glucose levels due to impaired endocrine function inhibits mannose conversion to glucose.</td>
</tr>
<tr>
<td>Ceramide (d18:1, C24:1)</td>
<td>Complex lipids, fatty acids and related (Subclass: sphingolipids)</td>
<td>Increased in response to cellular stress (ie, fibrosis and calcifications)</td>
</tr>
<tr>
<td>N-acetylcystidine</td>
<td>Nucleobases and related</td>
<td>Increased in chronic inflammation</td>
</tr>
</tbody>
</table>

Multivariate statistical analysis of the biomarker panel in the three study cohorts

Separate analysis of the three datasets revealed that the eight chosen metabolites were all significantly altered in CP patients versus controls in both plasma-based studies (p<0.05; FDR<0.2), and 6 of them also in the serum-based study. The variance analysis results (fold-changes) for the panel metabolites are shown in table 4. FDR-corrected p values (q values) are provided in online supplemental table S3. Of note, the fold-changes were in the same range across all studies. A striking feature was the very small p values for the panel metabolites with small shifts in concentration levels between the groups. In addition, the PCAs show that the groups in the second validation study were less homogenous than in the first validation study, leading to higher p values in the MANOVA.

The calculation of prediction scores revealed the relative importance of the panel metabolites, with beta-carotene and cryptoxanthin having the highest impact (figure 4A). Boxplots indicate the inter-individual variability of the panel metabolite levels (figure 5A–H) in the different studies. They illustrate why the carotenoids were chosen as the most important metabolites by the algorithm, as they discriminated best between groups. Nevertheless, single metabolites were unable to discriminate between CP and controls. Only the computation of the biomarker panel yielded a sufficiently accurate diagnosis. The distribution of age and BMI over the biomarker signature score is shown in online supplemental figure S1.

Effect of pancreatic insufficiency

Because the identification of beta-carotene and cryptoxanthin suggested a pathophysiological mechanism of malabsorption/malabsorption, we analysed whether pancreatic exocrine insufficiency (PEI) and enzyme supplementation (pancreatic enzyme replacement therapy, PERT) had an effect on plasma carotenoid levels (online supplemental figure S2). No significant increase of carotenoid plasma levels in PEI with PERT was observed. However, a significant increase of the biomarker signature score values was seen when comparing CP patient from the identification and first validation cohorts with and without pancreatic insufficiency, indicating correlation of this metabolic biomarker signature with disease stage (online supplemental figure S3).

DISCUSSION

In our proof-of-concept biomarker study following the TRIPOD guidelines,20 we show for the first time that a signature comprised of 6 metabolites of six different ontology classes can successfully differentiate between CP and controls with acceptable accuracy (AUC >0.8) in serum and EDTA-plasma samples.

There is no recommended blood-based biomarker for diagnosis of CP in medical guidelines.5 Some proteins or miRNAs have been proposed to have a potential to take up this role, but validation studies in larger cohorts are still lacking.28 29 Those biomarkers analytically validated for diagnosis using mass spectrometry are reviewed by Chou et al, but diagnostic accuracy is mostly unknown.30 Studies in rodents employing spontaneous and pharmacologically induced models of CP suggested significant alterations to the pancreatic metabolome, including energy production, anabolism, lipid synthesis and ROS detoxification pathways.31 32 Small and due to their heterogeneity inconclusive nuclear MR-spectroscopy-based human studies identified changes in citrate and adenosine levels in urine and

Table 3: Performance characteristics for the biomarker signature

<table>
<thead>
<tr>
<th>Dataset (matrix)</th>
<th>Chronic pancreatitis versus control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identification study (plasma) First validation study (plasma) Second validation study (serum)</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.85 (0.79 to 0.91) 0.85 (0.81 to 0.89) 0.88 (0.81 to 0.95)</td>
</tr>
<tr>
<td>Sensitivity (LL – UL)</td>
<td>0.71 (0.60–0.80) 0.84 (0.77–0.89) 0.78 (0.64–0.87)</td>
</tr>
<tr>
<td>Specificity (LL – UL)</td>
<td>0.86 (0.77–0.92) 0.66 (0.50–0.73) 0.89 (0.78–0.95)</td>
</tr>
<tr>
<td>Cut-off</td>
<td>0.479 0.479 0.479</td>
</tr>
</tbody>
</table>

AUC, area under the curve; LL, lower limit; UL, upper limit.
We could not discover an association between PEI or PERT and transcriptional markers.43 44 This is the first study to describe a decrease in beta-caro tene when compared with controls.35–40 Whether the deficiency in beta-carotene and cryptoxanthin (exogenous compounds) is secondary to malabsorption remains unknown.41 42 We could not discover an association between PEI or PERT and beta-carotene levels in the identification cohort. However, these data need to be interpreted with caution as previous studies have been inconsistent regarding the relationship of PEI and nutritional markers.43–46 We did not find a further increase after repeated episodes of acute pancreatitis, it underscores the role of the pancreas in sugar metabolism during health and disease.

Table 4 Statistical analysis results (linear model) of the signature metabolites (CP vs control)

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Identification study (plasma)</th>
<th>First validation study (plasma)</th>
<th>Second validation study (serum)</th>
<th>Importance of rank in the algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fold change (p value)</td>
<td>Fold change (p value)</td>
<td>Fold change (p value)</td>
<td></td>
</tr>
<tr>
<td>Beta-carotene</td>
<td>0.37 (6.8x10⁻⁶)</td>
<td>0.38 (2.5x10⁻⁷)</td>
<td>0.47 (0.00017)</td>
<td>1</td>
</tr>
<tr>
<td>Cryptoxanthin</td>
<td>0.40 (8.8x10⁻⁷)</td>
<td>0.38 (8.8x10⁻⁷)</td>
<td>0.48 (0.00043)</td>
<td>2</td>
</tr>
<tr>
<td>Mannose</td>
<td>1.47 (5.9x10⁻⁶)</td>
<td>1.52 (3.9x10⁻⁹)</td>
<td>2.80 (7.3x10⁻⁴)</td>
<td>3</td>
</tr>
<tr>
<td>Behenic acid (C22:0)</td>
<td>0.72 (0.00024)</td>
<td>0.67 (4.9x10⁻¹⁰)</td>
<td>1.14 (0.10586)</td>
<td>4</td>
</tr>
<tr>
<td>Ceramide (d18:1, C24:1)</td>
<td>1.27 (0.00016)</td>
<td>1.28 (1.3x10⁻⁸)</td>
<td>1.96 (6.7x10⁻⁸)</td>
<td>5</td>
</tr>
<tr>
<td>Indole-3-acetic acid</td>
<td>0.63 (0.00086)</td>
<td>0.62 (4.2x10⁻¹¹)</td>
<td>1.15 (0.33726)</td>
<td>6</td>
</tr>
<tr>
<td>Hippuric acid</td>
<td>0.68 (0.039)</td>
<td>0.49 (2.2x10⁻⁹)</td>
<td>0.51 (0.00173)</td>
<td>7</td>
</tr>
<tr>
<td>N-acetylcycteine</td>
<td>1.22 (0.0096)</td>
<td>1.21 (2.3x10⁻¹⁰)</td>
<td>2.36 (3.3x10⁻¹⁰)</td>
<td>8</td>
</tr>
</tbody>
</table>

CP, chronic pancreatitis.
components of the algorithm. This is further supported by the finding, that the biomarker score increases with presence of pancreatic insufficiency, a surrogate for disease stage.

A recent study by an US American consortium used a 62-plex luminex assay to identify potential biomarkers for diagnosis of CP. In blood samples of 41 controls, 20 recurrent acute pancreatitis (RAP) and 40 CP patients they found that GM-CSF, IFNβ, Leptin, PDGFB and Resistin could distinguish between CP and control (AUC 0.86) and resistin, SCF, MIP-1α and IL-17F between RAP and CP (AUC 0.77). Although the results are not comparable due to methodological differences it appears clear that only a combination of markers rather than a single protein allows for adequate discrimination. Independent validation of these data are pending.

A potential weakness of our study is the fact that CP patients and controls were unmatched for age, gender and BMI due to disease heterogeneity and consecutive recruitment. CP is diagnosed predominantly in middle-aged males at risk for malnutrition. Non-pancreatic controls (eg, day-surgery patients or blood donors) comprise by default a different cohort. Consequently, gender, BMI and age were not included in the prediction model because we aimed to avoid the pitfall that a shift in age alone could be sufficient to change a positive to a negative classification or vice versa.

It cannot be ruled out that some of the differences in the metabolic profiles between the groups are due to these features. Nevertheless, the MANOVA statistics were corrected for age, gender and BMI, and still show a significant difference between the groups for all chosen metabolites. This supports our hypothesis that this metabolic signature is able to identify CP patients irrespective of gender, age and BMI.

Figure 5  Boxplots of individual metabolite levels from the biomarker signature within the three studies, separated by group. CP, chronic pancreatitis.
We found the metabolic signature developed to distinguish CP from controls to be less discriminative when applied to liver cirrhosis samples. The similarity was interpreted as being likely due to activation of fibrosis and alcohol abuse, a common feature of both disorders. Fortunately, having to discriminate between liver cirrhosis and CP is uncommon and diseases rarely overlap. In the second validation study however, an almost complete separation of the control group from CP was seen. In this cohort, the control group was comprised healthy blood donors instead of non-pancreatic patients, which explains the significant improvement.

A potential clinical use of this metabolic signature is the identification of CP patients early in the disease course (early CP), of patients with unexplained abdominal symptoms and a history of pancreatic disease, but (yet) no definitive morphological signs of CP (probable CP), or of patients with RAP at risk for developing CP. These groups so far are only vaguely defined by international consensus diagnostic criteria. We, therefore, recruited patients with definitive CP for the sake of biomarker development. Whether the presented metabolic signature is sensitive enough under the above circumstances needs further testing in trials with long-term follow-up, ideally in a design that includes disease staging via chronic pancreatitis diagnosis score.

Whether a metabolic biomarker can distinguish between CP and cirrhosis which share a common aetiology (alcohol), also needs further prospective studies.

In conclusion, we have identified and validated an LC-MS/MS-based human blood-metabolome signature, which successfully discriminates between healthy individuals and patients with CP. Whether this biomarker has clinical value for diagnosing early stages of CP or can be used to monitor disease progression needs further prospective studies.

Author affiliations
1 Metanomics Health GmbH, Berlin, Berlin, Germany
2 Department of Medicine II, Ludwig-Maximilians-Universität Munich, Munich, Bayern, Germany
3 Transil Medizin GmbH, Ludwigshafen am Rhein, Rheinland-Pfalz, Germany
4 Department of Surgery, Erlangen University Hospital, Erlangen, Bayern, Germany
5 Clinic and Outpatient Clinic for Visceral-, Thorax- and Vascular Surgery, Dresden University Hospital, Dresden, Sachsen, Germany
6 St. Josef Hospital, Department of Surgery, Ruhr University Bochum, Bochum, Nordrhein-Westfalen, Germany
7 Asklepios Clinic Harburg, Department for General and Visceral Surgery, Asklepios Hospital Group, Hamburg, Hamburg, Germany
8 Department of Surgery, Charité Universitätsmedizin Berlin Campus Charité Mitte, Berlin, Berlin, Germany
9 Department of Medicine A, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, Germany

Correction notice This article has been corrected since it published Online First. The author, Tim Fahlbusch’s, name has been corrected.

Acknowledgements We are grateful to technical staff in the analytical department of Metanomics Health for their excellent technical support and Susanne Wiche, and Doris Jordan for technical assistance.

Contributors Study design: JM, BK and MML; patient and biomaterial recruitment: CP, TF, MD, FK, FUV, MB, WU, RG, GB, UMM, JM and MML; data acquisition and statistical analysis: MGA, BK, NC; writing committee: MGA, GB, JM, BK and MML; manuscript revision and approval: BK, NC, CP, MD, TF, FK, FUV, MB, WU, RG, JM and MML.

Funding Supported by the Deutsche Krebshilfe/Dr. Mildred-Scheel-Stiftung (109102), the Deutsche Forschungsgemeinschaft (DFG MA 4115/1-2/3, SFB 1321: Project-ID 329628492, BE 6395/1-1), the Federal Ministry of Education and Research (BMBF G07-167533A) and BMBF (01KI1528) and EFRE-State Ministry of Economics MV (V-630-S1502-132/133), ESF/14-BM-AS5-0045/16 PePPP and ESF/14-BM-AS5-0010/18 EnE(Gie).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval All studies were designed and conducted in adherence to the Declaration of Helsinki and approved by the local ethics review boards of all four participating centres.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data of de-identified patient data, which are available from the corresponding author after appropriate application for reuse under research purpose only.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and derive various versions, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Ujjwal M Mahajan http://orcid.org/0000-0002-7238-8775
Frank U Weiss http://orcid.org/0000-0003-3639-6369
Julia Mayerle http://orcid.org/0000-0002-3666-6459

REFERENCES
15 Whitcomb DC, Shimosegawa T, Chari ST, 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...


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#, Georg Beyer#, Nicole Christiansen1,4, Beate Kamlage1, Christian Pilarsky5, Marius Distler6, Tim Falbusch7, Ansgar Chromik8, Fritz Klein9, Marcus Bahra9, Waldemar Uhl7, Robert Grützmann5, F. Ulrich Weiss10, 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 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

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


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.
### Table S1: R packages and functions used

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Packages</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Stats (included in R core pkg)</td>
<td>prcomp()</td>
</tr>
<tr>
<td>Biomarker training and test</td>
<td>E1071</td>
<td>naivebayes()</td>
</tr>
<tr>
<td></td>
<td>ROCR</td>
<td>performance()</td>
</tr>
<tr>
<td></td>
<td>pROC</td>
<td>prediction()</td>
</tr>
<tr>
<td></td>
<td></td>
<td>roc()</td>
</tr>
<tr>
<td>ANOVA/MANOVA</td>
<td>slme</td>
<td>lme()</td>
</tr>
<tr>
<td></td>
<td>stats (included in R core pkg)</td>
<td>p.adjust()</td>
</tr>
<tr>
<td>Visualizations</td>
<td>ggplot2</td>
<td></td>
</tr>
</tbody>
</table>
Table S2: Weightings of all metabolites in the biomarker signature

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

SD: standard deviation

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

<table>
<thead>
<tr>
<th>Dataset (matrix)</th>
<th>Identification study (plasma)</th>
<th>First validation study (plasma)</th>
<th>Second validation study (serum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolite</td>
<td>Fold change (q-value)</td>
<td>Fold change (q-value)</td>
<td>Fold change (q-value)</td>
</tr>
<tr>
<td>Beta-carotene</td>
<td>0.37 (9.8 x 10^-6)</td>
<td>0.38 (1.8 x 10^-19)</td>
<td>0.47 (0.00039)</td>
</tr>
<tr>
<td>Cryptoxanthin</td>
<td>0.40 (9.9 x 10^-6)</td>
<td>0.38 (4.5 x 10^-18)</td>
<td>0.48 (0.00090)</td>
</tr>
<tr>
<td>Mannose</td>
<td>1.47 (9.9 x 10^-6)</td>
<td>1.52 (6.7 x 10^-27)</td>
<td>2.80 (8.4 x 10^-13)</td>
</tr>
<tr>
<td>Behenic acid (C22:0)</td>
<td>0.72 (0.0080)</td>
<td>0.67 (1.0 x 10^-15)</td>
<td>1.14 (0.14)</td>
</tr>
<tr>
<td>Ceramide (d18:1, C24:1)</td>
<td>1.27 (0.0059)</td>
<td>1.28 (1.0 x 10^-15)</td>
<td>1.96 (5.9 x 10^-12)</td>
</tr>
<tr>
<td>Indole-3-acetic acid</td>
<td>0.63 (0.013)</td>
<td>0.62 (2.4 x 10^-8)</td>
<td>1.15 (0.40)</td>
</tr>
<tr>
<td>Hippuric acid</td>
<td>0.68 (0.15)</td>
<td>0.49 (1.2 x 10^-7)</td>
<td>0.51 (0.0034)</td>
</tr>
<tr>
<td>N-Acetylcytidine</td>
<td>1.22 (0.058)</td>
<td>1.21 (8.1 x 10^-9)</td>
<td>2.36 (3.2 x 10^-12)</td>
</tr>
</tbody>
</table>

CP: Chronic pancreatitis
Biomarker signature score

<table>
<thead>
<tr>
<th></th>
<th>No insufficiencies</th>
<th>Endo- or exocrine insufficiency</th>
<th>Endo- and exocrine insufficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>57</td>
<td>85</td>
<td>57</td>
</tr>
</tbody>
</table>
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 Adam¹,², Georg Beyer³, Nicole Christiansen¹,⁴, Beate Kamlage¹, Christian Pilarsky⁵, Marius Distler⁶, Tim Falbusch⁷, Ansgar Chromik⁸, Fritz Klein⁹, Marcus Bahra⁶, Waldemar Uhl⁷, Robert Grützmann⁵, F. Ulrich Weiss¹⁰, Julia Mayerle³,¹⁰, and Markus M. Lerch¹⁰

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


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.
Table S1: R packages and functions used

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Packages</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Stats (included in R core pkg)</td>
<td>prcomp()</td>
</tr>
<tr>
<td>Biomarker training and test</td>
<td>E1071</td>
<td>naivebayes()</td>
</tr>
<tr>
<td></td>
<td>ROCR</td>
<td>performance()</td>
</tr>
<tr>
<td></td>
<td>pROC</td>
<td>prediction()</td>
</tr>
<tr>
<td></td>
<td></td>
<td>roc()</td>
</tr>
<tr>
<td>ANOVA/MANOVA</td>
<td>slme</td>
<td>lme()</td>
</tr>
<tr>
<td></td>
<td>stats (included in R core pkg)</td>
<td>p.adjust()</td>
</tr>
<tr>
<td>Visualizations</td>
<td>ggplot2</td>
<td></td>
</tr>
</tbody>
</table>
Table S2: Weightings of all metabolites in the biomarker signature

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

SD: standard deviation

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

<table>
<thead>
<tr>
<th>Dataset (matrix)</th>
<th>Identification study (plasma)</th>
<th>First validation study (plasma)</th>
<th>Second validation study (serum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolite</td>
<td>Fold change (q-value)</td>
<td>Fold change (q-value)</td>
<td>Fold change (q-value)</td>
</tr>
<tr>
<td>Beta-carotene</td>
<td>0.37 (9.8 x 10^{-6})</td>
<td>0.38 (1.8 x 10^{-19})</td>
<td>0.47 (0.00039)</td>
</tr>
<tr>
<td>Cryptoxanthin</td>
<td>0.40 (9.9 x 10^{-6})</td>
<td>0.38 (4.5 x 10^{-18})</td>
<td>0.48 (0.00090)</td>
</tr>
<tr>
<td>Mannose</td>
<td>1.47 (9.9 x 10^{-6})</td>
<td>1.52 (6.7 x 10^{-23})</td>
<td>2.80 (8.4 x 10^{-13})</td>
</tr>
<tr>
<td>Behenic acid (C22:0)</td>
<td>0.72 (0.0080)</td>
<td>0.67 (1.0 x 10^{-15})</td>
<td>1.14 (0.14)</td>
</tr>
<tr>
<td>Ceramide (d18:1/C24:1)</td>
<td>1.27 (0.0059)</td>
<td>1.28 (1.0 x 10^{-15})</td>
<td>1.96 (5.9 x 10^{-12})</td>
</tr>
<tr>
<td>Indole-3-acetic acid</td>
<td>0.63 (0.013)</td>
<td>0.62 (2.4 x 10^{-8})</td>
<td>1.15 (0.40)</td>
</tr>
<tr>
<td>Hippuric acid</td>
<td>0.68 (0.15)</td>
<td>0.49 (1.2 x 10^{-7})</td>
<td>0.51 (0.0034)</td>
</tr>
<tr>
<td>N-Acetylcytidine</td>
<td>1.22 (0.058)</td>
<td>1.21 (8.1 x 10^{-6})</td>
<td>2.36 (3.2 x 10^{-12})</td>
</tr>
</tbody>
</table>

CP: Chronic pancreatitis
**Biomarker signature score**

<table>
<thead>
<tr>
<th></th>
<th>No insufficiencies</th>
<th>Endo- or exocrine insufficiency</th>
<th>Endo- and exocrine insufficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Count</strong></td>
<td>57</td>
<td>85</td>
<td>57</td>
</tr>
</tbody>
</table>

* Cut-off