Novel biochemical markers for non-invasive detection of pancreatic cancer

Pavel HRABÁK1, Jan ŠOUPAL2, Marta KALOUSOVÁ3, Tomáš KRECHLER1, Michal VOČKA4, Tomáš HANUŠ5, Luboš PETRUŽELKA1, Štěpán SVAČINA1, Aleš ŽÁK1, Tomáš ZIMA3

14th Department of Medicine-Department of Gastroenterology and Hepatology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; 3rd Department of Medicine-Department of Endocrinology and Metabolism, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; Department of Oncology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; Department of Urology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

*Correspondence: marta.kalousova@lf1.cuni.cz

Received July 30, 2021 / Accepted January 11, 2022

To identify non-invasive biomarkers of non-metastatic pancreatic cancer (PC), the blood from 186 patients (PC n=28; DM-diabetes mellitus n=60; ChP-chronic pancreatitis n=47; healthy controls n=51) was analyzed for 58 candidate biomarkers. Their effectiveness to identify PC was compared with CA19-9. Panel defined by Random-forest (RF) analysis (CA19-9, AAT, IGFBP2, albumin, ALP, Reg3A, HSP27) outperforms CA19-9 in discrimination of PC from DM (AUC 0.92 vs. 0.82). Panel (S100A11, CA72-4, AAT, CA19-9, CB, MMP-7, S100P-s, Reg3A) is better in discrimination PC from ChP than CA19-9 (AUC 0.90 vs. 0.75). Panel (MMP-7, Reg3A, sICAM1, OPG, CB, ferritin) is better in discrimination PC from healthy controls than CA19-9 (AUC 0.89 vs. 0.78). Panel (CA19-9, S100P-pl, AAT, albumin, adiponectin, IGF-1, MMP7, S100A11) identifies PC among other groups better than CA19-9 (AUC 0.91 vs. 0.80). Panel defined by logistic regression analysis (prealbumin, IGFBP-2, DJ-1, MIC-1, CA72-4) discriminates PC from DM worse than CA19-9 (AUC 0.80 vs. 0.82). Panel (IGF-1, S100A11, Reg1alfa) outperforms CA19-9 in discrimination PC from ChP (AUC 0.76 vs. 0.75). Panel (IGF-2, S100A11, Reg3A) outperforms CA19-9 in discrimination PC from healthy controls (AUC 0.95 vs. 0.78). Panel (albumin, AAT, S100P-serum, CRP, CA19-9, TFF1, MMP-7) outperforms CA19-9 in identification PC among other groups (AUC 0.89 vs. 0.8). The combination of biomarkers identifies PC better than CA19-9 in most cases. S100A11, Reg3A, DJ-1 were to our knowledge identified for the first time as possible serum biomarkers of PC.

Key words: serum biomarker; pancreatic cancer; non-invasive diagnosis; panel of biomarkers; S100 proteins family

Pancreatic cancer (PC) belongs to the most severe diseases of the gastrointestinal tract. It’s the 8th most common cause of death from malignancy [1]. Its incidence is growing in developed countries. The only curable method of PC is surgery, but just a minority of patients are diagnosed in the early stage of the disease. Identification of biomarkers, which could detect patients in the early stage of PC could be the key factor of successful treatment.

Many previous studies failed while looking for the solitary biomarker of early PC [2]. As the gold standard CA19-9 is still widely used, despite its false negativity in fucosyltrans- ferase-deficient individuals, false positivity in several situations (e.g., acute, or chronic pancreatitis, cholangitis, liver cirrhosis, and obstructive jaundice), and low sensitivity for detecting small (<3 cm) pancreatic tumors [3].

The purpose of this study was to identify a set of markers that can help us to differentiate patients with pancreatic cancer from the other groups (DM-type 2 diabetes mellitus, ChP-chronic pancreatitis, healthy controls). In this study, 58 markers were studied (Table 1). We studied not only „new biomarkers”, which were chosen as candidate markers based on relevant articles and results of published studies but also markers routinely measured in clinical praxis. We presume, that mechanism of carcinogenesis in PC is different in each risk group (DM, ChP) and thus represented by different biomarkers. Panels of biomarkers identified in this study
were compared with ‘the gold standard’ CA19-9 – a widely used but nonspecific biomarker.

Patients and methods

Study design. In this study, 4 groups were compared: patients with PC with the rest of probands and PC with other groups (newly/<3years/diagnosed DM, ChP, healthy controls) separately.

Patients with histologically confirmed ductal adenocarcinoma were included in the study after staging (inclusion criterion – stage I–III of American Joint Committee on Cancer classification (AJCC)). In the DM group, patients treated in the 3rd Department of Medicine of the General University Hospital in Prague with newly (<3 years) diagnosed type 2 diabetes mellitus were included. For the chronic pancreatitis group, patients were included in the study according to the Mannheim criteria [4].

186 patients were in total included in the study. 60 patients with DM, 47 patients with ChP, 28 with PC in stage 1–3 according to AJCC (non-metastatic), and 51 healthy controls. The basic characteristics of experimental groups are summarized in Table 2.

The study was approved by the local Institutional Ethical Committee (No.: 336/11 S–IV) and was conducted in accordance with the Declaration of Helsinki. All subjects have given informed consent with participation in the study.

Laboratory analysis. Blood samples were obtained in the General University Hospital in Prague. Blood from each individual was collected through puncture of the cubital vein into tubes containing ethylene diamine tetraacetic acid (EDTA) as an anticoagulant (to obtain plasma) and into tubes without an anticoagulant agent (for obtaining serum). Samples were centrifuged for 10 min at 1450×g and serum and plasma aliquots were stored according to the study protocol at –80 °C until analysis was performed.

Laboratory analysis was performed by using the following methods (serum analysis, if not written otherwise):

- CEA, CA19-9: chemiluminescence assay (CLIA), Architekt, Abbott, USA; CA72-4, insulin, C-peptide: electrochemiluminescence assay (ECLIA), Modular, Roche, Germany; AFP: CLIA, Centaur, Siemens, Germany; AAT: nephelometry, Immage, Beckman Coulter, USA; Protein S100P (serum and plasma): ELISA (enzyme linked immunosorbent assay), MBL International, USA; MMP 7: ELISA, R&D Systems, USA; DJ-1, ALCAM, CEACAM-1: ELISA, Abnova, Taiwan; IGF-1, IGF-2, IGFBP-1, IGFBP-2, IGFBP-3: ELISA Medigagnost, Germany; Osteopontin, TIMP-1, MCP-1, HSP-27, sICAM-1, MMP-9 (plasma): ELISA, eBioscience, USA; HSP60 (plasma), HSP70: ELISA, StressMarq, Canada; Adiponectin, sRAGE, MIC-1, Reg1α, Midkine, Leptin, OPG, TFF1, S100A6, REG3A, S100A11, PANDER: ELISA, Biovendor Laboratorní medicína, s.r.o., Czech Republic; HSP60 (plasma), HSP70: ELISA, StressMarq, Canada; Adiponectin, sRAGE, MIC-1, Reg1α, Midkine, Leptin, OPG, TFF1, S100A6, REG3A, S100A11, PANDER: ELISA, Biovendor Laboratorní medicína, s.r.o., Czech Republic.

Routine biochemical parameters were analyzed by standard methods by automatic analyzers.

Statistical analysis. Standard descriptive statistics measures were used for concentrations of all studied parameters (mean, standard deviation, min, max, median). To compare biomarker levels in all groups, ANOVA analysis, followed by Tukey’s test was performed (p-values obtained for these analyses are summarized in Table 1).

The logistic regression model has been built as a basic model for analysis, the biomarkers were included in the model by using forward and backward stepwise regression analysis, where p-value (p<0.05) from ANOV A analysis was the inclusion criteria.

Table 1. Concentrations of analyzed biomarkers in the studied group (for each subgroup separately) and p - values for ANOVA and post-hoc test (Tukey’s test).

<table>
<thead>
<tr>
<th></th>
<th>Pancreatic cancer group</th>
<th>Diabetes Mellitus group</th>
<th>Chronic pancreatitis group</th>
<th>Healthy controls</th>
<th>p-values for ANOVA+Tukey’s test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean±SD</td>
<td>med</td>
<td>mean±SD</td>
<td>med</td>
<td>mean±SD</td>
</tr>
<tr>
<td>AAT (g/l)</td>
<td>2±0.5</td>
<td>2.0</td>
<td>1.5±0.3</td>
<td>1.5</td>
<td>1.7±0.4</td>
</tr>
<tr>
<td>Adiponectin (ng/ml)</td>
<td>23602.5±14223.1</td>
<td>21195.0</td>
<td>18033.8±7509.3</td>
<td>18630.0</td>
<td>21072.3±12193.2</td>
</tr>
<tr>
<td>Albumin (g/l)</td>
<td>40.1±4.2</td>
<td>40.6</td>
<td>45.1±3.7</td>
<td>45.3</td>
<td>45.1±4.6</td>
</tr>
<tr>
<td>ALCAM (pg/ml)</td>
<td>214.6±65.8</td>
<td>200.5</td>
<td>199.6±86.5</td>
<td>186.0</td>
<td>196.2±76</td>
</tr>
<tr>
<td>ALP (ukat/l)</td>
<td>2.5±3.6</td>
<td>1.6</td>
<td>1±0.3</td>
<td>0.9</td>
<td>1.8±2</td>
</tr>
<tr>
<td>ALT (ukat/l)</td>
<td>0.9±0.8</td>
<td>0.5</td>
<td>0.6±0.3</td>
<td>0.4</td>
<td>0.6±0.5</td>
</tr>
<tr>
<td>AMS-P (ukat/l)</td>
<td>1.1±3</td>
<td>0.4</td>
<td>0.4±0.2</td>
<td>0.4</td>
<td>0.7±0.8</td>
</tr>
<tr>
<td>ApoB (g/l)</td>
<td>1±0.2</td>
<td>1.0</td>
<td>1±0.3</td>
<td>0.9</td>
<td>1±0.3</td>
</tr>
<tr>
<td>AST (ukat/l)</td>
<td>0.8±0.7</td>
<td>0.5</td>
<td>0.5±0.2</td>
<td>0.4</td>
<td>0.5±0.3</td>
</tr>
</tbody>
</table>
Table 1. Continued ...

<table>
<thead>
<tr>
<th></th>
<th>Pancreatic cancer group</th>
<th>Diabetes Mellitus group</th>
<th>Chronic pancreatitis group</th>
<th>Healthy controls</th>
<th>p-values for ANOVA+Tukey's test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean±SD</td>
<td>med</td>
<td>mean±SD</td>
<td>med</td>
<td>mean±SD</td>
</tr>
</tbody>
</table>
| CA19-9 (KU/l) | 1108.6±2979.4 | 215.2 | 14.2±10.4 | 12.4 | 22.8±57.8 | 10.5 | 9.4±6.1 | 8.4 | ***,$$$, &$$, %$$%
| CA72-4 (KU/l) | 6.7±14.9 | 2.1 | 2.2±2.2 | 1.4 | 2.1±2.8 | 1.3 | 1.9±1.9 | 1.3 | **,$$, &$$, %$$%
| CEA (ug/l)    | 3.3±3.6 | 2.4 | 1.7±1.2 | 1.5 | 2±1.5 | 1.7 | 1±0.6 | 0.7 | ***,$, &$$&$$,%$$%
| CEACAM (ng/ml)| 89.3±95.3 | 61.1 | 47.2±48.1 | 38.8 | 75.6±75.8 | 51.4 | 46.9±11.9 | 44.6 | **,$&%$$%
| Chol (mmol/l) | 4.8±1.1 | 4.8 | 4.7±1.1 | 4.4 | 4.8±1.2 | 4.7 | 5.7±1.1 | 5.7 | ***,$$$, %$$%
| Cpeptid (mmol/l) | 0.8±0.5 | 0.7 | 1.1±0.4 | 1.0 | 0.7±0.3 | 0.6 | 0.8±0.2 | 0.8 | ***,$,$%
| CRP (mg/ml)   | 22.2±44.5 | 5.3 | 4.3±5.2 | 2.7 | 9.4±19.5 | 1.5 | 3±4 | 1.7 | ***,$$&$$&$$,%$$%
| DJ-1 (ng/ml)  | 52.3±85.1 | 31.4 | 18.1±11.1 | 15.0 | 38.4±35.6 | 30.5 | 38.2±39.9 | 24.7 | **,$&$$%
| Ferritin (ug/l) | 358.7±371.1 | 266.7 | 211.4±229.9 | 117.6 | 207.6±197.5 | 138.0 | 124.5±119.6 | 85.6 | ***,$$$, %$$%
| GGT (ikat/l)  | 4.4±10.5 | 1.0 | 0.7±0.7 | 0.5 | 2.6±8.6 | 0.6 | 0.5±0.3 | 0.4 | *,$&%$$%
| Glyk,HbA1c (mmol/mol) | 45.5±11.6 | 45.0 | 49.5±29.3 | 50.0 | 42.6±12.9 | 39.5 | 34.8±5.5 | 35.0 | **%
| Glyk.prot (unol/l) | 247.1±44.6 | 237.5 | 261.4±59.2 | 247.0 | 245±37.9 | 238.0 | 229±17 | 230.0 | **%
| Glykemie (mmol/l) | 6.8±2.4 | 6.3 | 7.5±2.3 | 6.8 | 6.1±2.8 | 5.0 | 4.9±0.6 | 4.9 | ***,$$$%
| HDL-chol (mmol/l) | 1.2±0.4 | 1.3 | 1.3±0.3 | 1.2 | 1.4±0.6 | 1.3 | 1.5±0.4 | 1.6 | ***,$%
| HSP27 (pg/ml) | 2485.6±1400.4 | 2106.1 | 1499.4±834.4 | 1280.2 | 1712.6±634.5 | 1687.7 | 1838.3±1059.5 | 1443.2 | ***,$$$, %$$&$$%
| HSP60 plasma (mg/ml) | 4963.7±1383.1 | 5017.5 | 4226.7±1426.6 | 3985.5 | 4639.2±1440.1 | 4432.0 | 5157±1749.1 | 5053.0 | *
| HSP70 (ng/ml) | 32.8±33.2 | 21.9 | 19.3±30.4 | 14.6 | 27.6±25.8 | 22.8 | 24.7±26.6 | 18.6 | **%
| IGFBP-1 (ng/ml) | 124.6±51.1 | 122.5 | 162.4±56.7 | 153.1 | 170.2±82.5 | 154.8 | 184.7±46.1 | 181.5 | ***,$$$, &$$, %$$%
| IGFBP-2 (ng/ml) | 535.8±183.6 | 563.0 | 743.1±236.5 | 701.3 | 646.6±200.7 | 649.2 | 744.7±127.6 | 721.2 | ***,$$$&$$, %$$%
| IGFBP-3 (ng/ml) | 11.8±8.1 | 8.5 | 5.5±4.6 | 4.1 | 12.9±11.9 | 10.9 | 6.7±6.6 | 4.8 | ***,$&$$, %$$%
| IGFBP-2 (ng/ml) | 597±295.9 | 553.0 | 299.6±198.9 | 255.7 | 575.6±272.2 | 511.4 | 330.7±143 | 315.3 | ***,$$$&$$, %$$%
| IGFBP-3 (ng/ml) | 2427.3±552.8 | 2370.6 | 2622.6±66.3 | 2654.3 | 2483.8±587.5 | 2483.4 | 2695.8±411.3 | 2723.3 | ***,$$$&$$, %$$%
| INS (U/ml) | 9.2±7.5 | 7.9 | 14.4±8.5 | 12.4 | 8.6±8 | 6.4 | 8.7±3.8 | 7.7 | ***,$%
| Leptin (ng/ml) | 12.3±11.3 | 9.6 | 19.4±14.3 | 16.2 | 6.4±8.3 | 3.7 | 13.5±12.4 | 9.4 | ***%
| MCP-1 (pg/ml) | 897.6±785.2 | 703.4 | 690.3±537.7 | 543.8 | 786.4±595.1 | 565.8 | 812.9±823.4 | 580.7 | ***,%%%
| MIC-1 (pg/ml) | 3715.9±2336.1 | 2896.5 | 3351.8±2022.2 | 2646.0 | 2789±1953.2 | 2149.0 | 1257.7±733.7 | 1099.0 | ***,%%%
| Midkine (ng/ml) | 0±0 | 0.0 | 0±0.1 | 0.0 | 0.4±2.3 | 0.0 | 0±0.1 | 0.0 | ***,%%%
| MMP-7 (ng/ml) | 2.3±1.1 | 2.0 | 1.5±0.7 | 1.4 | 1.5±1 | 1.4 | 0.9±0.5 | 0.9 | ***,$$$, %$$&$$, %$$%

et al.
<table>
<thead>
<tr>
<th></th>
<th>Pancreatic cancer group</th>
<th>Diabetes Mellitus group</th>
<th>Chronic pancreatitis group</th>
<th>Healthy controls</th>
<th>p-values for ANOVA+Tukey’s test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean±SD</td>
<td>med</td>
<td>mean±SD</td>
<td>med</td>
<td>mean±SD</td>
</tr>
<tr>
<td>MMP-9 (ng/ml)</td>
<td>202±179.7</td>
<td>137.6</td>
<td>197±180.5</td>
<td>156.6</td>
<td>197.7±180.5</td>
</tr>
<tr>
<td>OPG (pmol/l)</td>
<td>8±3</td>
<td>7.2</td>
<td>6.4±2</td>
<td>6.2</td>
<td>7.2±5</td>
</tr>
<tr>
<td>PANDER (ng/ml)</td>
<td>9.3±20.8</td>
<td>1.5</td>
<td>10.5±22.7</td>
<td>0.4</td>
<td>4.6±15.3</td>
</tr>
<tr>
<td>PCT (ug/l)</td>
<td>0.2±0.1</td>
<td>0.2</td>
<td>0.2±0.1</td>
<td>0.3</td>
<td>0.2±0.1</td>
</tr>
<tr>
<td>Prealb (g/l)</td>
<td>0.2±0.1</td>
<td>0.2</td>
<td>0.3±0.1</td>
<td>0.3</td>
<td>0.3±0.1</td>
</tr>
<tr>
<td>Reg1A (ng/ml)</td>
<td>264.1±203.3</td>
<td>213.1</td>
<td>171.8±107.4</td>
<td>134.1</td>
<td>175.6±116.9</td>
</tr>
<tr>
<td>REG3A (U/ml)</td>
<td>685.4±555.8</td>
<td>505.5</td>
<td>315±287.4</td>
<td>221.0</td>
<td>469.9±513.8</td>
</tr>
<tr>
<td>S100A11 (ng/ml)</td>
<td>7.2±6.1</td>
<td>4.6</td>
<td>5±6.6</td>
<td>3.3</td>
<td>3.7±2.1</td>
</tr>
<tr>
<td>S100A6 (ng/ml)</td>
<td>99.2±105.9</td>
<td>58.8</td>
<td>46±25.1</td>
<td>40.0</td>
<td>75.7±34.6</td>
</tr>
<tr>
<td>S100P-plasma (ng/ml)</td>
<td>13.6±17.8</td>
<td>4.4</td>
<td>5.5±4.9</td>
<td>3.7</td>
<td>4.3±3.6</td>
</tr>
<tr>
<td>S100P-serum (ng/ml)</td>
<td>17±13.4</td>
<td>14.2</td>
<td>8.9±4.3</td>
<td>8.7</td>
<td>14.2±6.7</td>
</tr>
<tr>
<td>CASA (mg/ml)</td>
<td>28.8±58.2</td>
<td>8.2</td>
<td>10.7±26.2</td>
<td>4.9</td>
<td>26.4±62.6</td>
</tr>
<tr>
<td>TFF1 (ng/ml)</td>
<td>24±12.3</td>
<td>22.4</td>
<td>26.3±10.7</td>
<td>24.5</td>
<td>32.4±11.7</td>
</tr>
<tr>
<td>sCAM-1 (ng/ml)</td>
<td>577±209.9</td>
<td>531.3</td>
<td>405.2±104</td>
<td>383.4</td>
<td>510.5±202.7</td>
</tr>
<tr>
<td>sRAGE (pg/ml)</td>
<td>550.5±213</td>
<td>496.0</td>
<td>530±175.1</td>
<td>517.0</td>
<td>614.8±234.4</td>
</tr>
<tr>
<td>TAG (mmol/l)</td>
<td>1.4±0.7</td>
<td>1.3</td>
<td>1.8±1</td>
<td>1.5</td>
<td>1.4±1.1</td>
</tr>
<tr>
<td>TFF1 (ng/ml)</td>
<td>3.6±3</td>
<td>2.3</td>
<td>3.5±3.9</td>
<td>2.1</td>
<td>3±3</td>
</tr>
<tr>
<td>TIMP-1 (ng/ml)</td>
<td>655.9±290</td>
<td>529.6</td>
<td>619.1±748.6</td>
<td>478.6</td>
<td>556.7±252.2</td>
</tr>
<tr>
<td>TP (g/l)</td>
<td>66±7.7</td>
<td>67.7</td>
<td>70.8±5.1</td>
<td>70.5</td>
<td>71.4±6.1</td>
</tr>
<tr>
<td>Transferrin (g/l)</td>
<td>2.4±0.4</td>
<td>2.3</td>
<td>2.9±0.6</td>
<td>2.8</td>
<td>2.7±0.5</td>
</tr>
<tr>
<td>Vitamin B12 (pmol/l)</td>
<td>472.9±81</td>
<td>422.0</td>
<td>409.4±101</td>
<td>389.0</td>
<td>456.2±373</td>
</tr>
</tbody>
</table>

Notes: *p<0.05 ANOVA; **p<0.01 ANOVA; ***p<0.001 ANOVA; $p<0.05 PC vs. ChP; $$p<0.01 PC vs. ChP; $$$p<0.001 PC vs. ChP; &p<0.05 PC vs. DM; &&p<0.01 PC vs. DM; &&&p<0.001 PC vs. DM; %p<0.05 PC vs. healthy controls; %%p<0.01 PC vs. healthy controls; %%%p<0.001 PC vs. healthy controls; Abbreviations: PC-pancreatic cancer; DM-T2 diabetes mellitus; ChP-Chronic pancreatitis; RF-random forest analysis; LR-logistic regression; AAT – alpha-1-antitrypsin; ALCAM-activated leukocyte cell adhesion molecule; ALP-alkaline phosphatase; ALT-alanine aminotransferase; AMS-P-pancreatic amylase; ApoB-apolipoprotein B; AST-aspartate aminotransferase; CA19-9-carbohydrate antigen 19-9; CA125-carbohydrate antigen 72-4; TP-total protein; CEA-carcinoembryonic antigen; CEACAM-carcinoembryonic antigen-related adhesion molecules; CRP-C-reactive protein; DJ-1-(ng/ml) protein deglycase DJ-1; GGT-Gama glutamyl transferase; Glyk.HbA1c-glycated hemoglobin A1c; Glyc.prot.-glycated protein (fructosamine); HDL-chol: high density lipoprotein; HSP-heat shock protein; IGF-insulin like growth factor; IGFBP-insulin like growth factor binding protein; INS-insulin; MCP-1-monocyte chemotactrant protein -1; MIC-1-macrophage inhibitory cytokine -4; MMP-matrix metalloproteinase; OPG-osteoprotegerin; PCT-procalcitonin; Prealb-prealbumin; Reg-regenerating protein; SAA-serum amyloid A; sICAM-soluble Intercellular Adhesion Molecule; sRAGE-soluble receptor for advanced glycation end-products; TAG-triacylglycerol; TFF1-trefoil factor 1; TIMP-1-(ng/ml) tissue inhibitor of metalloproteinases 1; PANDER-serum pancreatic derived factor.
For deeper analysis statistical method 'random forest' has been used because not all requirements and assumptions have been met in terms of normality, multicollinearity, etc.

Random forest (RF) is an ensemble learning method for classification and regression belonging to supervised machine learning algorithms. RF consists of many individual decision trees that operate as an ensemble. Each individual tree in the RF splits out a class prediction and the class with the most votes (modus) becomes the model's prediction.

The goal of this method is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

By target variable is meant one of four groups of patients (DM; PC; ChP; Healthy controls) and by features are meant biomarkers present in the dataset.

Decision rules are inferred from the data in the training dataset that is created as a subset from the original dataset. Then, the validity of these rules is tested on a testing dataset, which consists of the remaining data in the original dataset. Technically, in our work, each analyzed group of patients (PC vs. DM; PC vs. ChP; PC vs. Healthy controls; PC vs. DM+ChP+Healthy controls) was divided into two subsets in the ratio 80% : 20%. To minimize the risk of correlation among parameters, which can complicate interpretability of RF results (mainly features importance is biased) – the number of features was reduced by using a dendrogram created by Hierarchical Ward-linkage clustering based on the Spearman correlation. Because of the difference in the number of patients included in each subgroup, and the possibility that the dataset could not be well balanced, we ran RF modeling with the appropriate settings – weights of each class were automatically adjusted to be inversely proportional to the class frequencies in the input data. In the case of LR analysis, we balanced the dataset by using the down-sampling method. This method takes random n-values from a bigger dataset, where n is the size of a smaller dataset.

Models obtained by both LR and RF were based on analysis of a larger subgroup (training dataset).

The precision of the predictions tested on a smaller subset of each group was evaluated by metrics like specificity, sensitivity, recall, accuracy, and AUC. Results obtained by the logistic regression and by random forest were compared and evaluated.

All models and analysis were performed in the Python programming language using Spyder software (licensed under MIT, freeware, https://www.spyder-ide.org/).

Results

Results for logistic regression analysis. A panel of markers (prealbumin, IGFBP2-2, DJ-1, MIC-1, Ca72-4) obtained by LR analysis of a group of 70 patients was tested on a set of 18 patients. This panel shows good efficiency in discriminating patients with PC from DM (sensitivity 0.64, specificity 0.81, AUC 0.80) but seems to be less effective when compared with random forest model based only on CA19-9 (sensitivity: 0.64 vs. 0.71, specificity: 0.81 vs. 0.93, AUC 0.80 vs. 0.82).

A panel of markers (IGF-1, S100A11, Reg1alfa) obtained by LR analysis of 60 patients was tested on a set of 15 patients to evaluate its effectiveness in discriminating CP from ChP. This panel has a similar AUC (0.76 vs. 0.75), better sensitivity (0.86 vs. 0.57), and worse specificity (0.38 vs. 0.94) when compared with the random forest model based only on CA19-9.

A panel of markers (IGF-2, S100A11, Reg3A) obtained by LR analysis of 63 patients was tested on a set of 16 individuals. This model is more effective than CA19-9-based random forest model – (sensitivity: 0.86 vs. 0.71, specificity: 0.76 vs. 0.84, AUC 0.95 vs. 0.78) in discriminating PC from Healthy controls.

A panel of markers (albumin, AAT, S100P-serum, CRP, CA19-9, TFF1, and MMP-7) used for discriminating PC from other groups (ChP, DM, Healthy controls) was obtained by LR analysis of 148 patients and tested on 38 individuals. The panel has similar sensitivity and specificity, and a better AUC than RF model based only on CA19-9 (sensitivity: 0.71 vs. 0.71; specificity 0.9 vs. 0.89; AUC 0.89 vs. 0.8).

All mentioned logistic regression models and their parameters meet the criteria of p<0.05. Unfortunately, we were not able to build a statistically correct logistic regression model only with CA19-9 and after exclusion of CA19-9 as well, therefore for creating a panel based only on CA19-9 and for creating a comparative panel of biomarkers without CA19-9 only results for RF analysis are presented.

Result for RF analysis. By using RF analysis, sets of markers that outperform solitary CA19-9 in identifying PC among other groups were identified. All panels are sorted according to the importance of used markers.

A panel (CA19-9, AAT, IGFBP2, albumin, ALP, Reg3A, HSP27) obtained by analysis of 70 patients was tested on a set of 18 individuals and is better in discrimination PC from the DM group than CA19-9 (sensitivity 0.89 vs. 0.71; specificity 0.89 vs. 0.83; AUC 0.92 vs. 0.82). CA19-9 has the highest predictive value.

A panel (S100A11, CA72-4, AAT, CA19-9, CB, MMP-7, S100P-s, Reg3A) obtained by RF analysis of a set of 60 patients, was tested on 15 individuals and is better in discrimination...
Table 3. Discrimination of pancreatic cancer from other diagnoses using the specific groups of biomarkers.

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RF</td>
<td>LR</td>
</tr>
<tr>
<td>PC vs. DM</td>
<td>0.92</td>
<td>0.80</td>
</tr>
<tr>
<td>PC vs. ChP</td>
<td>0.90</td>
<td>0.76</td>
</tr>
<tr>
<td>PC vs. Healthy controls</td>
<td>0.89</td>
<td>0.95</td>
</tr>
<tr>
<td>PC vs. DM, ChP, Healthy controls</td>
<td>0.91</td>
<td>0.89</td>
</tr>
</tbody>
</table>

**Summary of results.** The results confirm that CA19-9 has a strong predictive role for the presence of PC but adding other markers to the model improved its efficiency in most
cases, especially in detecting patients with PC among the healthy population. On the contrary, after the exclusion of CA19-9 from the analyzed biomarkers, it was in some cases necessary to increase the number of biomarkers used in the panels to maintain sufficient effectiveness. Logistic regression analysis seems to have worse outcomes than random forest analysis. The results obtained by logistic regression could be improved by extension of the set of probands, while logistic regression usually performs better under this condition.

**Discussion**

Pancreatic cancer is a disease with an extremely poor prognosis especially when diagnosed in the advanced stages. The period of progression from the early to the advanced stage seems to be rapid [5]. Patients after appropriate surgical therapy in the early stage of the disease have better outcomes. A reliable non-invasive biomarker that can be used for identifying early pancreatic cancer in common and/or risk populations still does not exist.

The purpose of this study was to identify biomarkers that could improve the detection of early pancreatic cancer. In the current study, we measured levels of 58 biomarkers in healthy individuals and patients with PC, DM2, and ChP. Besides a standard statistical method, machine learning analysis was used to identify possible markers of early PC. To increase the performance of selected biomarkers, several panels were defined. Their effectiveness was compared to CA19-9.

Our results show that it is advantageous to use different panels of biomarkers for each risk group (e.g., DM, ChP). This suggests that processes leading to PC development can presumably differ among different patient groups. In ChP it is supposed that inflammatory response to various extrinsic or intrinsic factors leads to an excessive and prolonged exposition of pancreatic tissue to chemokines and cytokines. This results in pancreatic cells destruction, proliferation,
and epithelial-mesenchymal transition (EMT) [6]. In DM2, insulin resistance, increased production of growth factors, mild chronic systemic inflammation, and prolonged hyperglycemia contribute to developing PC [7, 8]. These differences in tumor development and differences in patient’s phenotypes (represented by e.g., significantly higher BMI in DM vs. ChP probands in this study as depicted in Table 2) are presumably reflected in different panels of suitable biomarkers identified in this work. Whether the identified biomarkers directly reflect different pathways of tumorigenesis in each group or not is to be elucidated.

A broad spectrum of measured biomarkers can be divided into several groups: biomarkers that directly contribute to carcinogenesis (D)-1, HSP27, IGF 1-2, IGFBP, MIC1, MMP7, Reg3A, S100A11, S100P, sICAM1, TFF1), biomarkers which correspond mainly with nutritional status of the individual (e.g., ferritin, albumin, prealbumin, TP), or reflect the presence of systemic inflammation (e.g., CRP, adiponectin, AAT, OPG, SAA).

Interestingly, some routinely measured parameters increased the effectiveness of PC diagnosing (TP, albumin, prealbumin, HDL, AST, ALT, TAG). Differences in nutritional parameters (albumin, prealbumin) presumably correspond with impaired nutritional status and systemic inflammatory changes in PC patients. Similar findings were presented by Ferri et al. [9]. Albumin and ferritin were identified as markers of worse prognosis in PC [10].

In recent years many studies proved a key role of pancreatic stellate cells (PSCs) in PC. Disruption of MMPs and TIMPs (tissue inhibitors of metalloproteinases) homeostasis due to PSCs activation is a key factor of excessive extracellular matrix production in PC tissue. In this work, we have confirmed that MMP-7, a member of the matrix metalloproteinase family, can improve diagnostic accuracy in discriminating PC from healthy controls or ChP. Several studies examined MMP-7 as a potential diagnostic or prognostic marker of PC [11]. We did not confirm that levels of TIMP play a role in the identification of PC.

Growth and proliferation of PSCs are stimulated besides other molecules by TTF1 (Trefoil factor 1). Expression of TTF1 was observed in most PC cell lines and its role in the development of ChP was well described [12]. As a candidate biomarker of PC, TTF1 was assessed in several studies [13, 14]. In our work, it was confirmed as a potential biomarker of PC.

ICAM-1, also produced by activated PSCs and pancreatic acinar cells expressing KRAS mutation, serves as an important signaling molecule for TAM (tumor-associated macrophage) [15, 16]. ICAM-1 has been assessed as a potential marker in several studies, but its ability to detect the early PC has not been proven yet [17]. In our work sICAM helped to distinguish between PC and healthy controls.

PSCs are also involved in the regulation of the IGF/IGFBP ratio. The increase of IGF or decrease of IGFBP leads to higher levels of free IGF. IGF stimulates the growth and invasion of PC cells. IGFBP I–VI, which have a strong affinity to IGF, are highly expressed in the PC cell membrane and contribute to the regulation of angiogenesis, growth, and invasion of tumor cells. A higher level of IGFBP-2 was confirmed in serum of patients with both PC and ChP [18, 19] as well as in its precursor lesions (Pan-IN) [20]. IGF/IGFBP-2 axis plays a role in PSCs proliferation and migration [21]. In our work IGFBP-2, IGF–1, and 2 were identified as potential markers of early PC.

Osteoprotegerin (OPG) is a receptor of TRAIL (TNF-related-apoptosis-inducing-ligand). Inhibition of TRAIL by binding to OPG leads to resistance of tumor cells to apoptotic signals [22]. OPG is also synthesized by activated PSCs [23]. Our study proved the ability of OPG (as well as Brand et al. [24]) to contribute to the discrimination of PC from healthy controls.

Intracellular S100 proteins take part in the regulation of transcription, protein phosphorylation, proliferation, and differentiation of cells. The relationship between the group of S100 proteins and pancreatic cancer was recently reviewed by Wu et al. [25].

Some of S100 proteins are ligands for RAGE (receptor for advanced glycation products). Activation of RAGE, which is highly expressed in PC tissue, induces chronic inflammation, activates KRAS, and increases tumor progression. This is similar to the situation of chronic hyperglycemia in DM2, where concentrations of advanced glycation products (AGEs), other ligands of RAGE, are increased [26].

In our work, we show that S100P, a ligand of RAGE, has a role as a potential circulating biomarker of the early PC. S100A11 produced by PC cells stimulates through RAGE the growth of stromal fibroblasts. Higher levels of S100A11 seem to be associated with increased mobility of PC cells [27]. In our study, S100A11 helps to discriminate PC from healthy controls and ChP significantly better than solitary CA19-9.

Higher levels of regenerating islet-derived protein 3A (Reg3A) were reported in PC and Pan-IN [28]. In our study, the panels of biomarkers containing Reg3A have shown to be more effective in distinguishing PC among other groups (ChP, DM, healthy controls) than solitary CA19-9. It is to our best knowledge for the first time, when the effectivity of plasmatic levels of Reg3A as a biomarker of non-metastatic PC was evaluated.

Midkine (neurite promoting growth factor 2) belongs to growth factors taking part in angiogenesis, fibrogenesis, cell migration, and proliferation in the PC environment [29, 30]. Grupp et al. did not prove that midkine levels have a relationship either to metastasizing or prognosis of PC [17]. In our work, which is to our best knowledge the first one to assess midkine as a possible diagnostic biomarker of PC, we confirmed the above-mentioned observation.

In conclusion, our finding supports existing evidence of the crucial role of PSCs/PC crosstalk in PC development and progression while some of the identified biomarkers are directly linked to PSCs activity.
Signaling pathways involved in PC development and progression are an interesting therapeutic target. Thus, it is crucial to describe key signaling cascades in PC. Biomarkers identified in this study are involved in several signaling cascades. DJ-1, S100A11, HSP-27, IGFBP-2, and MIC-1 are involved in the PI3K/AKT cascade [31–35], IGFBP 2, IGF1, and IGF-2 are involved in the IGF-related pathways. OPG plays a role in the TRAIL/TRAIL-receptor cascade [36]. These findings suggest an important role of these pathways in the early stages of PC.

This work confirmed CA19-9 as the most reliable marker in the detection of PC with known limitations (low sensitivity and specificity – 0.71 and 0.89 in our work). None of the analyzed markers outperformed CA19-9 in diagnostic performance and excluding CA19-9 from analysis led to a significant decrease in the effectiveness of diagnostic panels in most cases. On the other hand, combinations of the markers improved the effectiveness of CA19-9 and in some cases, panels without CA19-9 were even more effective than CA19-9 alone or in combination. The significant decrease of effectiveness of panels of biomarkers after exclusion of CA19-9 confirms the rationale of using CA19-9 as a routine marker of PC.

The main advantage of our study is a broad panel of assessed biomarkers. Midkine was to our knowledge, assessed for the first time as a possible biomarker but we did not prove its ability to detect the early PC. Reg3A and DJ-1 were to our knowledge identified for the first time, as potential blood biomarkers of the early PC. Another advantage is the use of machine learning methods for analysis. Results of our study should be confirmed on larger and more precisely defined patient cohorts.

Acknowledgments: The authors are thankful to nurses from clinical departments, to laboratory staff for technical assistance, and to data scientist Jan Hrabák for the statistical analysis. This study was supported by the project TIP ČR FR-T13/666 given to First Faculty of Medicine, Charles University, and the company Biovendor, and by research projects Progres Q25, DRO VFN64165, SVV 260524-2021, and BBMRI-CZ LM2018125.

Conflict of interest disclosure: All above-mentioned authors declare no financial/personal conflict of interest. The company Biovendor, manufacturer of some ELISA tests, participated in the project TIP ČR FR-T13/666 including some laboratory analyses of coded samples but had no influence on the evaluation and interpretation of the data in a relationship with clinical characteristics of the patients and writing of the manuscript.

References


