Novel mRNA-based Immunological Biomarker for Early Detection of Resectable Pancreatic Cancer

Jong-chan Lee  
Seoul National University Bundang Hospital

Hong Sik Kim  
HuVet bio Inc

Eun-Jin Sim  
HuVet bio Inc

Hyunjung Cho  
Dongguk University

Yuna Youn  
Seoul National University Bundang Hospital

Jaihwan Kim  
Seoul National University Bundang Hospital

Hyoung-Hwa Jeong  
HuVet bio Inc

Jihie Kim  
Dongguk University

Jin-Hyeok Hwang (✉ woltoong@snu.ac.kr)  
Seoul National University Bundang Hospital

Research Article

Keywords: liquid biopsy, mRNA, buffy coat, pancreatic cancer, biomarker

Posted Date: April 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2658368/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: In this study, we aimed to develop and validate a novel mRNA-based immunological biomarker panel for the early detection of resectable pancreatic cancer (RPC).

Methods: Blood samples and clinical information of a cohort of 297 individuals (102 patients with PC and 195 non-disease controls) were obtained between 2015 and 2020. mRNA levels of candidate immunological biomarkers identified through an extensive literature review in buffy coat samples were measured using quantitative polymerase chain reaction. The final biomarker panel was confirmed using biological feature selection. Subsequently, the marker panel was trained and validated through artificial intelligence (AI)-based cross-validation.

Results: Among 55 candidate biomarkers, eight mRNAs (CCL2, CCL5, CXCR2, IFN-γ, IFN-λ1, PTGES2, SLC27A2, and TNF) were selected based on biological and statistical performance to develop a biomarker panel. The octet-biomarker panel successfully identified all stages of PC with an area under the curve (AUC) of 0.925 in the training set (AUC = 0.922 in the validation set). In patients with RPC, the AUC for the panel was 0.973, whereas that for CA19-9 level was 0.809. For the combination of the panel and CA19-9 level, the AUC was 0.990 in patients with RPC. Furthermore, even in patients with RPC and normal CA19-9 level, the AUC for the octet-biomarker panel was 0.969, whereas that for CA19-9 level alone was 0.787 and for the combination of the panel and CA19-9 level was 0.977.

Conclusion: Compared with CA19-9, the mRNA-based immunologic octet-biomarker panel improved diagnostic performance, especially in RPC with normal CA19-9 levels.

Background

Pancreatic cancer (PC) is among the most lethal cancers, with a 5-year survival rate of approximately 10–12% [1, 2], and is the fourth leading cause of cancer-related deaths [3]. Although surgical resection is the only curative strategy for PC, less than 20% of patients are diagnosed in the resectable stage, indicating that more than 80% of patients have an unresectable PC initially and have a low probability of being cured [4, 5]. Therefore, developing effective markers to detect early-stage PC is essential for improving overall survival [6, 7].

Currently, among the potential biomarkers for PC, the only US FDA-approved clinical biomarker for PC is serum carbohydrate antigen 19–9 (CA19-9) [8]. However, despite its good performance in reflecting treatment response of patients with advanced PC, the diagnostic role of CA19-9 in screening and detecting early PC is limited [9–13]. According to previous studies, the sensitivity and specificity of CA19-9 level for PC diagnosis are under 80% and 90%, respectively. Additionally, positive predictive value (PPV) constitutes only 0.9% of the asymptomatic population owing to the low incidence of PC [14, 15]. Furthermore, the CA19-9 level does not increase in Lewis antigen-negative patients, even in those with advanced PC [16–19]. Therefore, CA19-9 is not considered a suitable marker for the early diagnosis of PC.

Studies have reported various serum biomarkers such as proteins, microRNAs, cell-free DNAs (cfDNAs), and their combinations for PC diagnosis [20–23]. However, such studies have not reported meaningful diagnostic performance for early PC or PC with normal CA19-9 levels, and none have been approved by the FDA.

In the early phase of PC development, including acinar-to-ductal metaplasia (ADM) or low-grade pancreatic intraepithelial neoplasia (PanIN), epithelial-to-mesenchymal transition (EMT) occurs in precancerous pancreatic cells; hence, circulating tumor cells (CTCs) can be detected in blood vessels [24–29]. These CTCs could be precursors of metastasis, which can be found as CTC clusters composed of non-malignant cells such as mesenchymal stem cells and immune cells [30]. During the circulation of CTC clusters in the bloodstream, cancer-associated mesenchymal
stem cells (CA-MSCs) promote the proliferation, invasion, and metastasis of PC and induce immunosuppressive polarization in immune cells [31]. In particular, granulocyte macrophage colony-stimulating factor (GM-CSF) secreted by CA-MSCs induces the transformation of neutrophils into polymorphonuclear myeloid-derived suppressor cells (PMN-MDSCs) through GM-CSF/STAT5/SLC27A2 pathway-mediated arachidonic acid uptake, which ultimately leads to prostaglandin E2 (PGE2)-mediated CD8 + T cell suppression [32–35]. In this process of immune system reprogramming, various types of signaling pathways are involved (Supplementary Fig. 1). However, such immunological changes may have distinct characteristics between early and advanced PCs. In the present study, we aimed to identify novel immunologic markers for PC diagnosis with a focus on PC-specific early reprogramming of the immune system, which is substantially associated with tumorigenesis in resectable PC.

Methods

Study design and patient recruitment

We performed biomarker research with prospectively collected biobank specimens to evaluate the performance of blood biomarkers and develop a biomarker quantitative polymerase chain reaction (qPCR) diagnostic kit with improved diagnostic performance compared with that of CA19-9 alone.

Buffy coat samples of 297 individuals were used in the present study, with the following inclusion criteria: (1) pancreatic cancer group; individuals who were diagnosed with “histologically confirmed” pancreatic ductal adenocarcinoma (PDAC); (2) non-disease control group, healthy control group, or high-risk control groups, including patients with chronic pancreatitis, late-onset diabetes, and intraductal papillary mucinous neoplasm (IPMN) or mucinous cystic neoplasm (MCN) of the pancreas.

Samples were obtained from the Human Bioresource Center of Seoul National University Bundang Hospital between March 2015 and September 2020. The buffy coat samples comprised 102 PDAC and 195 non-disease control samples (177 healthy controls and 18 high-risk patients). This study was approved by the Institutional Review Board of Seoul National University of Bundang Hospital (SNUBH) (IRB approval number, X-2011-651-903). All methods were carried out in accordance with the Declaration of Helsinki.

Marker selection and additional validation

The marker-selection scheme is illustrated in Fig. 1. In the first step, through an extensive literature search and review, multiple markers (the number of markers is defined as $N_1$ in Fig. 1a) were selected as potential candidate markers. A PRISMA-based literature search was conducted in the PubMed and MEDLINE databases using the following query: (pancreatic cancer OR pancreatic ductal adenocarcinoma) AND (liquid biopsy OR early detection OR biomarker) AND (immune OR immunologic reprogramming). These candidate markers highly correlated with immunological pathways during the development of PDAC from PanIN. Considering the pathophysiological differences that occur during the early progression phase of PDAC, the mRNA expression levels of individual markers in immune cells are expected to vary (upregulated or downregulated). The mRNA expression levels of the potential candidate biomarkers were estimated using qPCR analysis, and the $\Delta Ct$ of each biomarker was calculated for statistical analysis.

In the second step, non-parametric tests (Mann–Whitney U test and Kolmogorov–Smirnov test) were performed to select primary biomarker candidates (the number of markers is defined as $N_2$ in Fig. 1b). Subsequently, feature selection was performed to estimate the performance of each biomarker in distinguishing PDAC from normal tissues (the number of markers is defined as $N_3$ in Fig. 1b).
In the third step, to determine the final combination of biomarkers (the number of markers is defined as $N_4$ in Fig. 1c), a prediction model was developed using logistic regression (LR) with 10-fold cross-validation. PDAC and control samples were randomly divided among the total buffy coat samples to configure a training set and validation set. The training and validation sets were randomly split at a ratio of 7:3. The classification model was generated based on the training dataset using binary classification techniques (LR algorithm), and a 10-fold CV procedure was applied to stabilize each classification model. The area under the curve (AUC) of the receiver operating characteristic (ROC) was used as a criterion to evaluate the performance of the classification models. We evaluated the diagnostic performance in distinguishing between non-disease controls and PC using a biomarker panel and compared the results of the biomarker panel against the diagnostic performance of CA19-9 alone. The optimal cut-off value of the biomarker panel was determined using the Youden index.

Finally, additional validation tests with subjects whose CA19-9 levels were within the normal range (< 37.0 U/mL) and high-risk groups were developed to verify the performance of the biomarker panel.

**Biological sample collection, storage, and processing**

Normal samples were defined as samples from non-disease controls without any malignancy or pancreatic disease. Whole blood samples were collected in ethylenediaminetetraacetic acid tubes. The samples were centrifuged at 1800 $\times$ g for 10 min, and 250 µL of the buffy coat layer was collected from the separated blood. These buffy coat samples were stored at −80°C immediately after separation. Total RNA was isolated from buffy coat samples using the NucleoSpin RNA Blood kit (MACHEREY-NAGEL, Düren, Germany) following the manufacturer's instructions. For cDNA synthesis, 1 µg of total RNA was subjected to reverse transcription using the GoScript Reverse Transcription System (Promega, Madison, WI, USA). The cDNA produced was stored at −80°C until further use as a template for quantitative polymerase chain reaction (qPCR).

**Quantitative reverse transcription PCR assay**

Quantitative reverse transcription PCR (RT-qPCR) was performed in a probe-based multiplex reaction. Each probe was labeled at the 5’ end with a reporter dye, FAM, or HEX, and at the 3’ end with ZEN/Iowa Black FQ double-quencher dye. Primers and probes were obtained from Integrated DNA Technologies, Inc. (IDT, Coralville, IA, USA), as shown in Supplementary Table 1. The mRNA expression levels were measured using the GoTaq Probe qPCR Master Mix (Promega, Madison, WI, USA) following the manufacturer's instructions, in a reaction mixture of final volume 20 µL. RT-qPCRs were performed using the QuantStudio 3 and QuantStudio 5 Real-Time PCR systems (Applied Biosystems, Foster City, CA, USA), with standard cycling conditions consisting of one cycle at 50°C for 2 min and 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and 60°C for 1 min. Relative gene expression levels were normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) level; the results are shown as ΔCt values.

**Statistical analysis**

Non-parametric statistical tests (Mann–Whitney U test or Kolmogorov–Smirnov test) were carried out using GraphPad Prism™ (version 9.3.1; San Diego, CA, USA) to compare the expression levels of different mRNAs. A feature selection analysis was performed with the package “Boruta” in R (version 4.1.3) to estimate the performance of each biomarker in non-disease control and PC samples (RPC and all stages of PC). A logistic regression analysis was conducted to estimate the diagnostic performance of a biomarker panel by comparing normal and PDAC samples. The diagnostic performance of the biomarker panel for the discrimination of normal and PDAC samples was assessed using ROC curve analysis, particularly the AUC, and DeLong’s test was used to evaluate the AUC values. Sensitivity and specificity were assessed using ROC curves represented by the corresponding AUC values with 95% confidence intervals (CIs). The p-values indicated the following: *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001.
Results

Patient characteristics

In this study, 297 buffy-coat samples of 297 patients were divided into two groups: pancreatic cancer (n = 102, 34%) and non-disease control (n = 195, 66%) groups. In the all pancreatic cancer (all PC) group, patients were divided into three groups according to clinical staging: a group of RPC (n = 35, 12%), group of borderline resectable and locally advanced PC (BRPC/LAPC, n = 31, 10%), and group of metastatic PC (MPC, n = 36, 12%). The non-disease control group was divided into a high-risk group (n = 18, 6%), comprising patients with chronic pancreatitis, intraductal papillary mucinous neoplasm (IPMN), and mucinous cystic neoplasm (MCN), and healthy controls (n = 177, 60%). The average of the evaluated CA19-9 levels was significantly different between PC and non-disease controls (PDAC; 300.0 U/mL vs. control; 7.4 U/mL). The demographic data are summarized in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Pancreatic cancer</th>
<th>Non-disease control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RPC</td>
<td>BRPC/LAPC</td>
<td>MPC</td>
</tr>
<tr>
<td>Number of patients</td>
<td>102 (34%)</td>
<td>35 (12%)</td>
<td>31 (10%)</td>
</tr>
<tr>
<td>Age</td>
<td>66 (57–72)</td>
<td>66 (60–75)</td>
<td>67 (58–70)</td>
</tr>
<tr>
<td>Sex</td>
<td>male</td>
<td>49</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>53</td>
<td>25</td>
</tr>
<tr>
<td>Baseline tumor marker</td>
<td>CEA</td>
<td>3.6</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>(1.9–6.5)</td>
<td>(1.5–4.5)</td>
<td>(2.1–19.5)</td>
</tr>
<tr>
<td></td>
<td>CA19-9</td>
<td>300.0</td>
<td>101.0</td>
</tr>
<tr>
<td></td>
<td>(61.3–1883.3)</td>
<td>(27.8–255.0)</td>
<td>(69.5–1231.0)</td>
</tr>
</tbody>
</table>

*High risk group includes patients with chronic pancreatitis, intraductal papillary mucinous neoplasm (IPMN), and mucinous cystic neoplasm (MCN). Data are presented as median (interquartile range) or n (%). PDAC, pancreatic ductal adenocarcinoma; RPC, resectable pancreatic cancer; BRPC, borderline resectable pancreatic cancer; LAPC, locally advanced pancreatic cancer; MPC, metastatic pancreatic cancer; CA 19−9, carbohydrate antigen 19−9; CEA, carcinoembryonic antigen.

Candidate marker screening (Step 1)
Using the PRISMA-based literature search for immunologically relevant markers, we identified 1755 articles and screened 55 (N7) biomarkers, including ANGPT2, ARG1, CCL2, CLEC6A, FGF2, IFNB1, IL10, IL22, LGALS3, OLR1, SLC27A2, TNF, and VEGFA. Information on these biomarkers and their relevant signaling pathways associated with immune system reprogramming is summarized in Supplementary Table 2.

**Marker selection from potential candidate biomarkers (Step 2)**

Statistical analysis was performed using non-parametric tests (Mann–Whitney U test or Kolmogorov–Smirnov test) of the 55 candidate immunological markers to determine the performance of each marker in distinguishing between the non-disease controls and patients with PDAC. A p-value < 0.05 was considered to meet the significance criterion. Analysis of the statistical results indicated that 17 markers (N2) were significantly different between the non-disease controls and patients with PDAC (Fig. 2). A list of the 17 markers is shown in Supplementary Table 3.

Feature selection was conducted among the 17 biomarkers in the experimental group of non-disease controls and patients with PDAC to verify their performance and prioritize individual markers. We set the standard of performance of each marker at the shadow maximum point and identified 15 markers that exceeded this point (Fig. 3a). Additional feature selection was performed to estimate marker performance in the experimental group of non-disease controls and patients with RPC. Using the Boruta algorithm, we set the standard performance for each marker using the shadowmax point. We identified 11 markers (N3) that exceeded the shadowmax point when comparing patients and non-disease controls (Fig. 3b). To construct a biomarker panel with optimal predictive performance in comparing patients and non-disease controls, we added markers consecutively from the high-ranking 11 markers based on the feature selection results and eliminated markers that degraded the performance of the panel.

**Biomarker panel analysis (Step 3)**

A logistic regression analysis was conducted along with the construction of an octet-biomarker panel consisting of eight final markers (N4) (Table 2). In the comparison of all patients with PDAC and non-disease controls in the training set, the ROC analysis revealed an AUC of 0.925, sensitivity of 93.0%, and specificity of 83.9% for the biomarker panel; whereas for CA19-9 only, the AUC was 0.923, sensitivity was 84.5%, and specificity was 100%. In the ROC analysis of the validation set, the AUC was 0.922, sensitivity was 83.9%, and specificity was 92.5% (Fig. 4a, 4b, and Table 3).
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Reference sequence</th>
<th>Function</th>
<th>Mapping*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCL2</td>
<td>C-C Motif Chemokine Ligand 2</td>
<td>NM_002982</td>
<td>A ligand for C-C chemokine receptor CCR2.</td>
<td>A, C</td>
</tr>
<tr>
<td>CCL5</td>
<td>C-C Motif Chemokine Ligand 5</td>
<td>NM_002985</td>
<td>Chemoattractant for blood monocytes, memory T-helper cells and eosinophils.</td>
<td>B, D</td>
</tr>
<tr>
<td>CXCR2</td>
<td>C-X-C Motif Chemokine Receptor 2</td>
<td>NM_001557</td>
<td>Receptor for interleukin-8 which is a powerful neutrophil chemotactic factor.</td>
<td>B</td>
</tr>
<tr>
<td>IFNG</td>
<td>Interferon Gamma</td>
<td>NM_000619</td>
<td>Type II interferon produced by immune cells such as T-cells and NK cells that plays crucial roles in antimicrobial, antiviral, and antitumor responses by activating effector immune cells and enhancing antigen presentation.</td>
<td>A</td>
</tr>
<tr>
<td>IFNL1</td>
<td>Interferon Lambda 1</td>
<td>NM_172140</td>
<td>Cytokine with antiviral, antitumor and immunomodulatory activities. Plays a critical role in the antiviral host defense, predominantly in the epithelial tissues.</td>
<td>B</td>
</tr>
<tr>
<td>PTGES2</td>
<td>Prostaglandin E synthase</td>
<td>NM_025072</td>
<td>Isomerase that catalyzes the conversion of PGH2 into the more stable prostaglandin E2 (PGE2).</td>
<td>A</td>
</tr>
<tr>
<td>SLC27A2</td>
<td>Solute carrier family 27 member 2</td>
<td>NM_003645</td>
<td>Acyl CoA synthetase that activates long-chain and very long-chain fatty acids (VLCFAs) by catalyzing the formation of fatty acyl-CoA.</td>
<td>A</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor Necrosis Factor</td>
<td>NM_000594</td>
<td>The regulation of a wide spectrum of biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation</td>
<td>A, D</td>
</tr>
</tbody>
</table>

* Mapping in the immune system reprogramming: A, immune suppression; B, pro-tumoral activity; c, M2 polarization; D, angiogenesis. (Refer to the Supplementary Fig. 1)
Table 3
Diagnostic performance of the marker panel

<table>
<thead>
<tr>
<th>Marker panel</th>
<th>Training set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>CA19-9 only</td>
<td>84.5</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All PC 8 markers</td>
<td>93.0</td>
<td>83.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 markers + CA19-9</td>
<td>98.6</td>
<td>92.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA19-9 only</td>
<td>62.5</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPC 8 markers</td>
<td>95.8</td>
<td>92.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 markers + CA19-9</td>
<td>100.0</td>
<td>93.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUC, area under curve; All PC, all pancreatic cancer patients; RPC, resectable pancreatic cancer; CA 19–9, carbohydrate antigen 19–9.

For patients with RPC in the training set, the AUC for the biomarker panel was 0.969 (95% CI = 0.943–0.994), which was significantly greater than that of CA19-9 alone (AUC = 0.813, 95% CI = 0.714–0.911). The sensitivity of the biomarker panel was 95.8%, which was higher than that of CA19-9 alone (62.5% with criteria of < 37.0 U/mL), and specificity was 92.7%, which was slightly lower than that of CA19-9 alone (100.0%) (Fig. 4c, 4d, and Table 3). For the combination of the panel and CA19-9, the AUC was 0.990 and sensitivity was 100%. In the validation set, the biomarker panel showed greater performance with an AUC of 0.973 (95% CI = 0.938–1.000), which significantly higher than that of CA19-9 alone (AUC = 0.809, 95% CI = 0.659–0.959). The sensitivity of the biomarker panel was 100.0% and the specificity was 88.7%, indicating an excellent improvement in the sensitivity of the biomarker panel compared with that of CA19-9 alone (63.6%), while retaining slightly low specificity (98.1%). For the combination of the panel and CA19-9, the AUC was 0.995 and sensitivity was 94.3%.

### Subgroup validation of the octet-biomarker panel

The biomarker panel was then tested on subjects with the normal range of CA19-9 level (< 37.0 U/mL) to verify its performance. For RPC patients with normal CA19-9 levels, the AUC value of CA19-9 alone was 0.787 (95% CI = 0.631–0.944), with a sensitivity of 61.5% and specificity of 96.6%. In contrast, the biomarker panel consistently showed
superior diagnostic performance, with an AUC value of 0.969 (95% CI = 0.940–0.997), sensitivity of 92.3%, and specificity of 93.3%, which demonstrated improved sensitivity over CA19-9 alone from 61.5–92.3% (Fig. 5a and Table 4). Additionally, the diagnostic performance of the biomarker panel was tested for its applicability in discriminating between non-disease controls and high-risk groups. In terms of diagnostic performance, the biomarker panel had a higher AUC of 0.865 (95% CI = 0.779–0.950) than CA19-9 alone (0.581, 95% CI = 0.443–0.720), along with improved sensitivity and specificity from 55.6–77.8% and 50.8–81.9%, respectively (Fig. 5b and Table 4).

### Table 4
Additional analysis of diagnostic performance

<table>
<thead>
<tr>
<th>Marker panel</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA19-9 only</td>
<td>61.5</td>
<td>96.6</td>
<td>0.787</td>
<td>–</td>
</tr>
<tr>
<td>8 markers + CA19-9</td>
<td>92.3</td>
<td>94.3</td>
<td>0.969</td>
<td>0.01873</td>
</tr>
<tr>
<td>CA19-9 only</td>
<td>55.6</td>
<td>50.8</td>
<td>0.581</td>
<td>–</td>
</tr>
<tr>
<td>8 markers + CA19-9</td>
<td>77.2</td>
<td>89.9</td>
<td>0.875</td>
<td>0.00002</td>
</tr>
<tr>
<td>AUC, area under curve; RPC, resectable pancreatic cancer; CA 19 – 9, carbohydrate antigen 19 – 9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Discussion

We developed an effective mRNA-based, immunologic, and octet-biomarker panel to detect resectable stages of PC using RT-qPCR. We identified an octet-biomarker panel consisting of CCL2, CCL5, CXCR2, IFNG, IFNL1, PTGES2, SLC27A2, and TNF, which were selected from 55 candidate biomarkers using a series of non-parametric tests, feature selection, and LR algorithm analyses. From these results, we constructed an optimal biomarker panel that could detect immunological changes in early PDAC development (Supplementary Fig. 1). Although the biomarker panel showed slightly better diagnostic performance than CA19-9 alone in “all PC” group, the octet-biomarker panel demonstrated remarkable improvements over CA19-9 alone in discriminating patients with RPC and non-disease controls, even patients with RPC with normal CA19-9 levels.

In the last two decades, several studies have focused on the combination of multiple proteins, microRNAs, and cfDNAs because it is difficult to achieve acceptable diagnostic performance with a single biomarker considering the
heterogeneity of cancer during development [20–23, 36–38]. Klein et al. demonstrated a reasonable performance with an AUC of 0.910 for various cancer types [39]. However, the sensitivity of PC was between 60% and 80%, and even lower for stage I cancer, with a sensitivity of 43% [41]. Lee et al. reported a sensitivity of 92.5% in detecting resectable PC, but the sensitivity was also lowered to 64.3% in the normal CA19-9 group using triple protein marker panels, including LRG1, TTR, and CA19-9 [42, 43]. Recently, Nakamura et al. showed excellent diagnostic performance with their transcriptomic signature, with an AUC of 0.930 and sensitivity in the early stages of PC (stages I and II) [44]. However, the transcriptomic signature combining exosome-based miRNAs and cell-free miRNAs requires various extraction kits to obtain transcriptomic samples and presents low-stability exosome-based miRNAs, thus making the signatures less attractive from the commercial perspective of a PC diagnostic kit [45]. Hence, there is still a need for clinically feasible biomarkers to detect early stage PC.

In this context, our octet-biomarker panel showed several strengths compared to the biomarkers of previous studies. First, the octet-biomarker panel maintained diagnostic superiority over CA19-9 in patients with RPC (Fig. 4c, 4d) and even in patients with normal CA19-9 levels (Fig. 5a). Second, we observed changes in immune system reprogramming in the early phase of PC-specific carcinogenesis (Supplementary Fig. 1) [30–35]. As all these courses of immune system reprogramming could be reflected in changes in the mRNA of immune cells, researchers or clinicians only need to obtain the buffy coat of patient blood samples, and this superior accessibility and feasibility are our third strength.

Our study has some limitations. The major limitations are the small sample size and single-center nature of the study compared to other large-scale multicenter studies. To overcome these limitations, we performed a cross-validation analysis. Another limitation is that some molecular biological mechanisms have not yet been precisely elucidated. However, a structured representation method, including a comprehensive literature review and multifaceted comparison of experimental data, was implemented. Nevertheless, further clinical validation through large-scale, multi-center-based research is necessary to verify the feasibility of the mRNA-based immunologic biomarker panel for diagnosing PDAC, especially for early stage PDAC detection.

Conclusions

In conclusion, this study is the first to discover the potential of overcoming the limitations of CA19-9 in early stage PC detection using mRNA-based immunologic biomarkers. Our analysis provides reliable evidence of the relationship between the eight biomarkers (CCL2, CCL5, CXCR2, IFNG, IFNL1, PTGES2, SLC27A2, and TNF) and PDAC detection, particularly for early stage PDAC detection in non-disease controls, indicating that the biomarker panel can improve the diagnostic performance of RPC, even in patients with normal ranges of CA19-9 level. However, further validation of our findings using a larger cohort is required.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of Seoul National University of Bundang Hospital (SNUBH) (IRB approval number, X-2011-651-903). All methods were carried out in accordance with the Declaration of Helsinki. The informed consents were obtained from all the participants in this study.

Consent for publication

Not Applicable

Availability of data and materials
All data generated or analysed during this study are included in this published article and its supplementary information files (file name: 09 raw data for statistical analysis.xlsx). The data that support the findings of this study are available from Seoul National University Bundang Hospital but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Seoul National University Bundang Hospital. (Contact to the corresponding author, Jin-Hyeok Hwang woltoong@snu.ac.kr)

**Competing interests**

Jin-Hyeok Hwang and Jihie Kim declare conflict of interest including advisory committee of HuVet bio Inc. Hyoung-Hwa Jeong is a CEO of HuVet bio Inc. Other authors have no COI to declare.

**Funding**

This work was supported by HuVet Bio Inc.

**Authors’ contributions**

- Jong-chan Lee: Conceptualization, Analysis, Writing of original draft
- Hong Sik Kim: Laboratory works, Analysis, Writing of original draft
- Eun-Jin Sim: Laboratory works, Data handling, Validation
- Hyunjun Cho: Statistical analysis, Software, Visualization
- Yuna Youna: Laboratory works, Data handling
- Jaihwan Kim: Supervision, Validation
- Hyoung-Hwa Jeong: Conceptualization
- Jihie Kim: Supervision, Software, Visualization
- Jin-Hyeok Hwang: Conceptualization, Supervision, Corresponding author

** All the authors reviewed and participated the writing of manuscript.

**Acknowledgements**

We thank Editage for English language editing. This study was conducted using bioresources distributed by the Human Bioresource Center of Seoul National University Bundang Hospital (biobank lot number, DT-2019-007, DT-2020-001-01, and DT-2020-011-01) and was supported by HuVet Bio, Inc.

**Authors’ information**

**References**


Figures

Fig 1. Scheme of study

---

Figure 1
Schematic of the study. (a) In the process of comprehensive literature search, potential candidate markers were screened ($N_1$, number of candidate markers in step 1). (b) Two steps of marker selection were conducted: a non-parametric comparison of patients with pancreatic cancer and non-disease controls ($N_2$, reduced number of selected markers in this step) and feature selection using Boruta package ($N_3$, more reduced number of markers in this step). (c) Final analysis of diagnostic performance of markers using logistic regression and cross-validation ($N_4$, final number of the biomarker panel).

Figure 2. Seventeen candidate markers ($N_2 = 17$) with non-parametric MW-U test

**Figure 2**

Seventeen candidate markers ($N_2 = 17$). Mann–Whitney U test of the candidates showed significant differences between he pancreatic cancer and non-disease control groups.
Figure 3. Eleven candidate markers (N3=11) with biological feature selection test.

Figure 3

Eleven candidate markers (N3 = 11) with biological feature selection test. (3a, all pancreatic cancer vs. non-disease control; 3b, resectable pancreatic cancer vs. non-disease control) The green boxes depict the significant difference between the pancreatic cancer and non-disease control groups.
Figure 4. Diagnostic performance of the octet-biomarker panel vs. CA19-9 only vs. combination of octet-biomarker panel and CA19-9.

(a) Training set of comparison of all patients with PC vs. non-disease controls. (b) Validation set of comparison in all patients with PC vs. non-disease control. (c) Training set of comparison of patients with RPC vs. non-disease control. (c) Validation set of comparison of patients with RPC vs. non-disease control.
Figure 5. Subgroup validations for octet-biomarker panel

A Validation set: RPC with normal CA19-9 level vs. non-disease controls.

<table>
<thead>
<tr>
<th>ROC Curve</th>
<th>Description</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA19-9 Only</td>
<td></td>
<td>0.79</td>
</tr>
<tr>
<td>8 markers</td>
<td></td>
<td>0.97</td>
</tr>
<tr>
<td>8 markers + CA19-9</td>
<td></td>
<td>0.98</td>
</tr>
</tbody>
</table>

B Validation set: high risk group vs. healthy group.

<table>
<thead>
<tr>
<th>ROC Curve</th>
<th>Description</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA19-9 Only</td>
<td></td>
<td>0.58</td>
</tr>
<tr>
<td>8 markers</td>
<td></td>
<td>0.86</td>
</tr>
<tr>
<td>8 markers + CA19-9</td>
<td></td>
<td>0.88</td>
</tr>
</tbody>
</table>

Figure 5

Subgroup validations for a octet-biomarker panel. (a) Comparison of RPC patients with normal CA19-9 level vs. non-disease controls. (b) Comparison of high risk group vs. healthy group.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 07supptable.xlsx
- 08supppfigure.pptx
- 09rawdataforstatisticalanalysis.xlsx