

“IULIU HAȚIEGANU” UNIVERSITY OF MEDICINE AND PHARMACY CLUJ-NAPOCA

DOCTORAL SCHOOL

Summary of the PhD Thesis

Proteome-based plasma biomarker studies for pancreatic diseases

PhD student **Maria Ilieș**

Scientific supervisor **Prof. Felicia Loghin, PhD**

CLUJ-NAPOCA 2017



UMF
UNIVERSITATEA DE
MEDICINĂ ȘI FARMACIE
IULIU HAȚIEGANU
CLUJ-NAPOCA

TABLE OF CONTENTS

INTRODUCTION	15
REVIEW OF THE LITERATURE	
1. Pancreatic ductal adenocarcinoma	19
1.1. Global overview	19
1.2. Risk factors	20
1.3. Pathophysiology	21
1.4. Diagnosis	23
1.5. Clinical management	24
1.6. Future perspectives	26
2. Intraductal papillary mucinous neoplasm of the pancreas	29
2.1. Global overview	29
2.2. Risk factors	29
2.3. Pathophysiology	29
2.4. Diagnosis	31
2.5. Clinical management	32
2.6. Future perspectives	34
3. Proteome based biomarker studies	37
3.1. Tissue based proteomics studies	38
3.2. Blood based proteomics studies	40
3.3. Future perspectives	41
PERSONAL CONTRIBUTIONS	
1. Aims	45
2. General methodology	47
2.1. Study design and sampling	52
2.2. Determination of protein concentration	55
2.3. Protein tryptic digestion	55
2.4. nano LC UDMS ^E analysis	55
2.5. nano LC HRM analysis	56
2.6. Enzyme linked immunosorbent assay (ELISA)	58
3. Study 1 - Influence of the blood sample collection method on the highly abundant protein depletion efficiency	
3.1. Introduction	59
3.2. Work hypothesis	60
3.3. Materials and methods	60
3.4. Results	61
3.5. Discussions	62
3.6. Conclusions	63

4. Study 2 - Impact of blood sample collection methods on protein profiling studies	
4.1. Introduction	65
4.2. Work hypothesis	66
4.3. Materials and methods	66
4.4. Results	66
4.5. Discussions	77
4.6. Conclusions	79
5. Study 3 - Plasma protein absolute quantification by nano LC UDMS^E	
5.1. Introduction	81
5.2. Work hypothesis	82
5.3. Materials and methods	82
5.4. Results	83
5.5. Discussions	90
5.6. Conclusions	93
6. Study 4 - Plasma protein profiling of IPMN as a potential precursor lesion of the pancreatic cancer	
6.1. Introduction	95
6.2. Work hypothesis	96
6.3. Materials and methods	96
6.4. Results	97
6.5. Discussions	106
6.6. Conclusions	109
7. General discussions	111
8. General conclusions	113
9. Originality and innovative contributions of the thesis	115
REFERENCES	117

KEY WORDS: proteomics, ultra high resolution mass spectrometry, diagnostic biomarkers, plasma, pancreatic cancer

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC), the most common form of pancreatic cancer, is an extremely lethal disease. With a 5 year survival rate of solely 8%, PDAC is currently the fourth leading cause of cancer related death worldwide. Of concern, by 2020 PDAC is predicted to take the second place. Therefore, PDAC is recognized as a critical health problem and in urgent need of its overall clinical management improvement.

The major reason for the worse outcome of PDAC is its late clinical diagnosis. Lacking in specific syndromes or even asymptomatic, over 80% of the patients are diagnosed at an advanced stage of the disease, when surgery - the only curative approach - is not an option anymore. Thus, early detection of PDAC is considered the Holy Grail in the clinical management. Tremendous efforts of several multidisciplinary approaches headed towards early detection tools failed to provide a reliable option up until now. One segment of the PDAC clinical management remaining a promising option for the early detection is represented by one of the precursor lesions, namely intraductal papillary mucinous neoplasm (IPMN).

IPMNs are cystic lesions of the pancreas, which can be easily detected by imaging techniques. About 30% of the IPMNs are progressing into PDAC over a time frame of up to 6 years, which offers clinicians a great opportunity for the early detection of the fatal PDAC. However, recent studies have evaluated the accuracy of the existing diagnostic guidelines and showed that there is a high prevalence of incorrectly diagnosed cysts and consequently, a high prevalence of overtreated as well as undertreated patients. All these are main reasons for the worrisomely rates of IPMN morbidity and mortality after developing PDAC and are clearly witnessing the need for a better diagnostic tool, one reliable option being a biofluid based biomarker.

Currently, proteome studies are in the light of biomarker research. Several proteins showed promising results for IPMN in tissue based studies, but further verification and validation are still needed. The latest research trends are targeting blood protein multimarker panels by employing mass spectrometry based proteomics techniques. To our knowledge there is only one proteome based study conducted so far which proposed a 6 blood protein panel for the diagnosis of IPMN. However, the panel was not tested in a PDAC patients group and its utility as a potential early detection tool for PDAC remains unknown.

Therefore, the major goal of this work was to explore the blood proteome of IPMN patients in comparison to PDAC patients and to find potential biomarkers for the early detection of the patients at risk for PDAC.

By using the most advanced proteomics techniques, this work follows a stepwise strategy: first, three methodological issues were addressed, which were found to be of critical importance for the blood proteomics studies and then the most important findings were applied in the search for IPMN proteome-based plasma biomarkers.

PERSONAL CONTRIBUTION

Study 1. Influence of the blood sample collection method on the highly abundant protein depletion efficiency

1.1 Work hypothesis

Since the sample collection method can critically affect protein measurements in blood, it is tempting to evaluate the impact on the highly abundant proteins (HAP) abundance and moreover, also on their depletion efficiency, a frequent pre-processing step employed in a proteomic workflow. Because of the routinely use of multiaffinity removal of HAP to enrich lower abundant proteins, the aim of the present study was to investigate the influence of different blood sample collection tubes on HAP and their depletion efficiency by employing a data independent acquisition (DIA) nano LC UDMS^E approach.

1.2 Materials and methods

Six healthy controls (HC) (3 male, 3 female) aged 24-29 years without documented chronic disease or medications were enrolled in this study. Four venous blood samples were drawn from each subject. The different samples were collected into the following BD Vacutainer® tubes: (i) EDTA, (ii) sodium heparin, (iii) sodium citrate, and (iv) serum separator SST™ II Advance tubes. In total, 24 plasma and serum samples were obtained.

1.3 Results

All biological samples were analyzed before and after the removal of 6 HAP: serum albumin, IgG, IgA, serotransferrin, haptoglobin, and alpha-1-antitrypsin by using a MARS6 column. Protein average abundance was further compared among the 48 samples analyzed. In all samples the 6 proteins covered more than 50% of total protein abundance before depletion. Independent of the blood collection method used, percentages of HAP were very similar (fig.1):

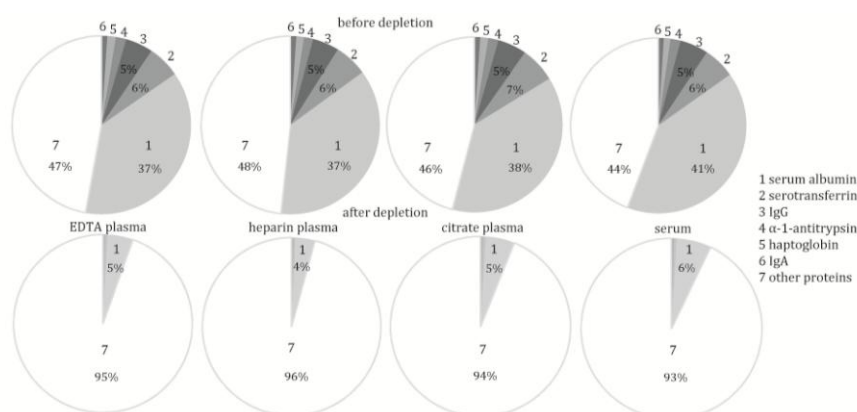


Fig. 1. Highly abundant proteins abundance distribution before and after depletion

1.4 Conclusions

The four different sample collection methods analyzed did not influence the abundance of the 6 highly abundant blood proteins. Also, they did not influence the depletion efficiency of the immunoaffinity MARS6 column. As such, this depletion method can be successfully employed in future blood proteomic studies.

Study 2. Impact of blood sample collection methods on protein profiling studies

2.1 Work hypothesis

Blood profiling studies using proteomics methods for biomarker investigations can be extended to a wider range of possible sample collection tubes if their impact on low and medium protein measurements in blood as important pre-analytical factor will be clarified. The aim of the present study was to investigate the impact of the four different blood sample collection tubes on the medium and low abundant blood protein profile.

2.2 Materials and methods

The materials and the methods used in this study are the same as for Study 1. Here, as a consequent step, a comprehensive analysis only on the low and medium abundant proteins of the depleted samples was addressed. In addition, system suitability of the nano LC UDMS^E measurements was tested.

2.3 Results

Among all sample collection methods, a total of 510 proteins could be quantified. Comparing the intensities of the 311/313/294/304 proteins identified in the EDTA-/heparin-/citrate plasma and serum samples, the mean biological variance for each test tube system was estimated at 0.28/0.20/0.23 and 0.20, respectively. The distribution of the 510 proteins over all sample collection methods unraveled: 187 proteins defined a core blood protein set independent of the sample collection method and 61/46/37 and 67 proteins represented unique signatures of the EDTA-/heparin-/citrate plasma and serum samples (fig. 2). On the basis of the median protein abundance, 97% of the core set was represented by classical plasma proteins predominantly secreted by the liver. Unique signatures were represented only by very small portions of the total protein.

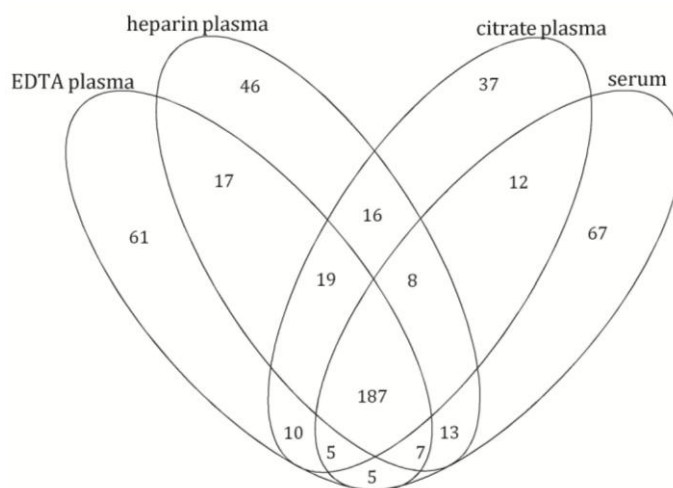


Fig. 2. Core and unique blood protein sets

To allow a comparison of serum and plasma samples, normalization was carried out using the protein with the lowest CV across all sample collection methods, complement factor I. Pair wise comparisons and Pearson's correlation coefficients evidenced strong correlation among all pairs, the highest correlation was found between heparin- and citrate plasma and EDTA- and citrate plasma core set proteins. A lower correlation was found in the different plasma and the serum core set pairs, but also between heparin- and EDTA plasma proteins.

2.4 Conclusions

This study unravelled decisive aspects of the sample collection method onto the blood protein pattern. Sample collection method did not severely influence the number of quantifiable proteins. Therefore, even if several specimen obtained by different sample collection methods are included in a study, a large number of proteins can be reliably quantified. Though, protein variation was influenced by the sample collection method, in our study heparin plasma and serum samples having the lowest CV at protein level. 187 proteins defined a very stable blood protein core set. In addition, complement factor I showed to be a potential candidate for normalization between different sampling methods. Moreover, unique proteins that were specific to the sample collection method could be uncovered. As expected, sample collection method needs to be considered essential pre-analytical player in blood profiling studies for biomarker screening.

Study 3. Plasma protein absolute quantification by nano LC UDMS^E

3.1 Work hypothesis

The aim of our study was to test a simple and rapid absolute quantification method for the verification of plasma protein sets which might be used as reliable biomarker panels in the clinical practice. The method uses synthetic *Escherichia coli* (*E. coli*) standard peptides for the absolute quantification as described earlier by Silva *et al.* and was adapted for the analyses of blood plasma using a high throughput DIA nano LC UDMS^E proteomics approach.

3.2 Materials and methods

In this study 6 undepleted EDTA plasma samples from the 6 HC enrolled for the studies 1 and 2 were analyzed. After lyophilization, *E. coli* standard peptides (protein disaggregation chaperone ClpB, Waters, 186006012) were spiked in each sample (1.65 fmol/100 ng) for the absolute quantification. Samples were subjected to nano LC UDMS^E analysis, whereas the spectral library was extended by the amino acid sequence of the three peptides with highest signal intensities of the spike-in standard (LPQVEGTGGDVQPSQDLVRNPNVLIGEPGVGKVTDAEIAEVLAR) and absolute protein quantification was performed on the spiked in amount of the *E. coli* standard peptide.

3.3 Results

Using nano LC UDMS^E 242 non-redundant plasma proteins were absolutely quantified in every one of the 6 individual samples analyzed. Based on the average concentration, 86% of proteins were represented by classical proteins, known also as secreted proteins, whereas a 7% fraction was represented by leakage proteins. The plasma protein concentration dynamic range covered was 10⁵ with a maximum average concentration of 839.79 μmol/L (serum albumin) and a minimum of 0.01 μmol/L (serum paraoxonase/lactonase 3). Among these, 63 proteins showed a CV ≤ 0.25 and were subsequently considered as the most stable protein set. The plasma concentration distribution of the most stable protein set in comparison to the complete set is illustrated in figure 3.

Further, correlation analysis of the absolute protein concentration with the reported data was employed. A set of highly and medium abundant proteins was reviewed by Hortin *et al.* (122), with 94 proteins being common to our protein set (Pearson's correlation $r=0.7664$). Our results were also compared to the protein panel quantified by Percy *et al.* (123) with 88 proteins in common to our study (Pearson's correlation $r=0.6710$). As a further proof of concept, plasma concentrations of hemoglobin subunit alpha (HBA) and hemoglobin subunit beta (HBB) were analyzed in relation to their sex specific plasma concentration. Our findings were found to be in accordance with reported data.

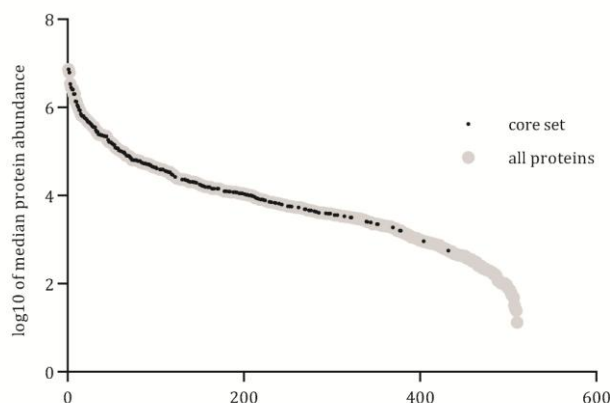


Fig. 3. Distribution of the most stable proteins among all quantified proteins

3.4 Conclusions

The study describes an absolute plasma protein quantification method using a state-of-the-art DIA nano LC UDMS^E proteomics approach. A high number of proteins in blood plasma could be analyzed in a single run, covering a wide dynamic range and showing high stability at protein level. The method has proven to be a reliable tool for the proteomics community offering a simple and fast method for the absolute quantification of blood proteins in a single run analysis. Without requiring additional pre-fractionation steps, which are time consuming and costly, and being entirely in accordance with other methods currently employed for biomarker research, this method is worth further validation over a larger number of samples and application for biomarker verification in the clinical practice.

Study 4. Plasma protein profiling of IPMN as a potential precursor lesion of the pancreatic cancer

4.1 Work hypothesis

Efforts for the early diagnosis of the PDAC have recently been directed to one of the precursor lesions, namely intraductal papillary mucinous neoplasm of the pancreas IPMN. Only a few studies have focused on IPMN molecular biology and its overall progression to cancer. IPMN lacks of comprehensive characterization which makes its clinical management controversial. Therefore, the aim of the present study was to explore plasma protein profiles of IPMN patients with small cyst size in comparison to HC, benign and malignant pancreatic disease patients by a high throughput MS DIA based approach, namely HRM.

4.2 Materials and methods

Four groups of participants were included in this study: 16 HC (aged 44-63 years), 17 patients with IPMN (aged 34-80 years), 15 chronic pancreatitis (CP) (aged 35-85 years), and 15 PDAC (aged 48-89 years). In addition, findings of this study were validated by ELISA based quantification. Plasma levels of human tissue inhibitor of TIMP-1, protein S100A9, and human periostin were determined using ELISA technique following manufacturer's instructions.

4.3 Results

To find proteins which might support overall IPMN diagnosis, proteins with specifically different abundance levels from the HC group (49 proteins, $p < 0.05$) were evaluated. Among proteins with fold

change (fc) greater than 2 were: carbonic anhydrases 1, 2, and 3 (CAH1 fc=5.5, CAH2 fc=15.9, CAH3 fc=3.9), and hemoglobin subunits alpha and beta, (HBA fc=3.9, HBB fc=5.1) (fig. 5A).

Proteins which might help to stratify patients with malignant disease status from those with benign disease or healthy patients should present significantly lower plasma abundance in the IPMN group as well as in HC when compared to PDAC patients (fig. 5 B). Proteins with higher abundance in IPMN and in HC plasma than in PDAC plasma samples included 3 proteins involved in cholesterol efflux and transport and might reflect an imbalance of lipid transport in PDAC patients (fig. 5 C). Proteins that display significantly lower plasma abundance in the IPMN group when compared to CP and PDAC might help in the characterization of the low risk IPMN patients. Ten proteins with significant difference in level were identified (fig. 5 D). Periostin and TSP1 were the only reported proteins associated with IPMN malignancy that could be identified in our plasma samples. Based on our MS results, IPMN plasma levels of TSP1 were significantly different from the HC group (fc=-1.78), from the CP (fc=-2.0) and the PDAC group (fc=-2.3). Periostin plasma levels did not show any significant difference and were very close to the detection limit, but were estimated to be highest in the IPMN and the PDAC groups (fig. 25 E).

We finally verified our MS based approach by ELISA assay on TIMP-1, protein S100A9 and periostin. In general, both MS and ELISA measurements showed the same plasma protein trend in the IPMN and the PDAC groups.

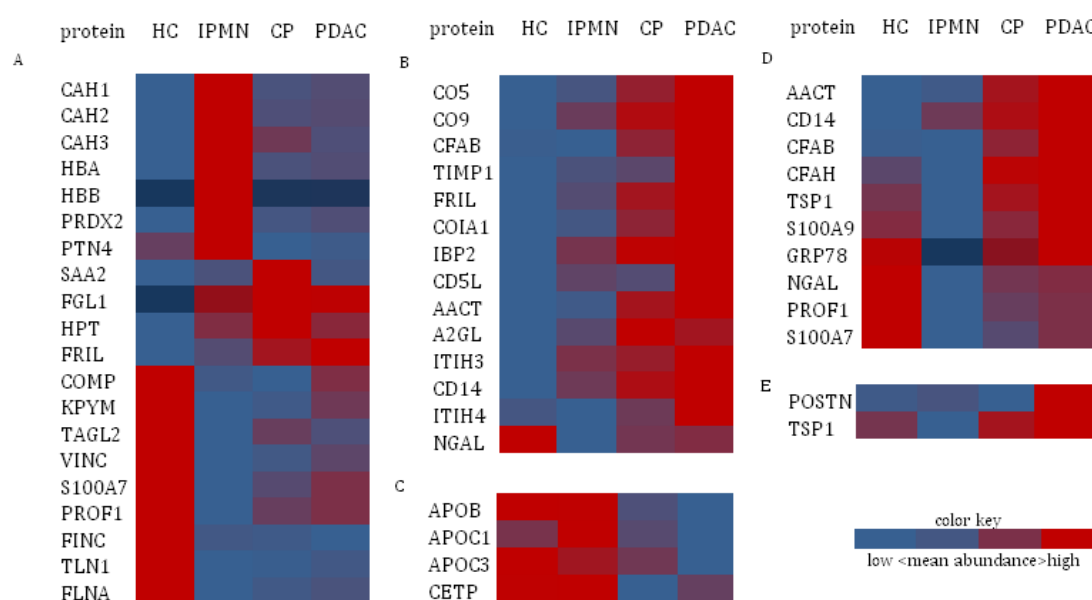


Fig. 4. Specific IPMN plasma proteins clusters displayed as heatmaps: A) Proteins with higher abundance in IPMN plasma in comparison with HC. B) Proteins with lower abundance in IPMN plasma when compared to PDAC. C) Proteins with higher abundance in IPMN and in HC plasma than in PDAC plasma D) Protein set with lower level in IPMN plasma when compared to the CP and PDAC. E) Abundance distribution of proteins that were previously reported as malignancy indicators for IPMN.

4.4 Conclusions

This study is the first to characterize plasma from IPMN patients by global protein profiling and to screen for protein sets displaying distinctively different abundances in comparison to healthy controls, as well as chronic pancreatitis and pancreatic ductal adenocarcinoma patients. Profiles of plasma proteins of IPMN are distinct from those of HC, and indicate disease related alterations detectable in plasma. Proteins identified as being specific for IPMN in our sample set have to be verified in a larger cohort and tested for their value in the risk stratification of IPMN. Proteins found to be increased in IPMN and PDAC in comparison with HC and CP might be reliable indicators for malignancy. Data on disease progression were not available for the IPMN patients of the current study population but will be in the focus of future investigations. Nevertheless, the plasma protein profiling data we provide can be used as a source of potential biomarkers for the precursor lesions of pancreatic cancer which so often ends fatally.

General conclusions

PDAC is a highly aggressive malignant disease, being the fourth most common cause of cancer related deaths. Even though tremendous efforts have already been conducted for its early detection, PDAC lacks in biomarkers of current clinical use. This has prompted current research attention to one of the PDAC precursor lesions, the IPMN. The general aim of these studies was to investigate the blood proteome as a potential source of biomarkers for the pancreatic diseases.

In the first two studies, the impact of four commonly used blood sample collection methods on blood protein profiling studies by using a label free data independent acquisition nano LC UDMS^E approach was addressed. Decisive aspects of the sample collection method onto the blood protein pattern were unraveled. The following findings can assist the planning of future investigations for blood protein profiling. The sample collection methods analyzed did not influence the 6 HAP: serum albumin, serotransferrin, IgG, haptoglobin, IgA, α -1-antitrypsin, did not influence the depletion of the 6 HAP by using the immunoaffinity MARS6 column, did not severely influence the number of quantifiable proteins, influenced protein variation, in our study heparin and serum samples had lowest CV at protein level. As such, the depletion method employed can be successfully used with very good depletion efficiency.

Independent on the sampling method, 187 proteins defined a very stable blood protein core set independent on the blood collection method, but also unique proteins that were specific to the sample collection method could be uncovered. As expected, the blood sample collection tubes need to be considered as essential pre-analytical players in blood profiling for biomarker screening.

The third study is a subsequent step of blood profiling studies, namely the absolute quantification of the proteins. Here, a simple and rapid absolute quantification method was described towards biomarker panel verification. In a single run analysis, without a pre-fractionation step such as depletion, the method enables a high number of plasma proteins quantified, covers a wide dynamic range covered and displays high stability at protein level. Thus, the method tested demonstrated applicability for blood protein panel verification and after validation it could be implemented towards biomarker studies.

Study four presents an in depth blood proteome characterization of the IPMN, a potential precursor lesion of the PDAC. Also, based on the results of the previous studies the blood sample used was heparin plasma. IPMN plasma proteome was characterized by global protein profiling and distinctive protein sets in comparison to HC, CP and PDAC were highlighted: carbonic anhydrases and hemoglobin subunits were among the protein sets that showed to be potential diagnostic markers for IPMN, complement components and ferritin light chain were among the proteins which showed to be potential differentiators between benign and malign diseases of the pancreas, and TSP1 was found to be potential malignancy indicator of a subsequent PDAC development in the IPMN group. Proteins identified as being specific for IPMN in our sample set have to be verified in a larger cohort and tested for their value in the risk stratification of IPMN.

All together, the four studies present decisive aspects regarding blood protein profiling studies and their application to investigate potential biomarkers for pancreatic diseases. The findings provided are reliable starting points for future studies that aim to investigate the blood proteome as a source of biomarkers towards clinical use. Moreover, the new markers described for the selected pancreatic diseases could be tested in larger cohorts and the proteins could be absolutely quantified towards biomarker verification using the method described.

UNIVERSITATEA DE MEDICINĂ ȘI FARMACIE "IULIU HAȚIEGANU" CLUJ-NAPOCA

ȘCOALA DOCTORALĂ

Rezumatul tezei de doctorat

Studiul biomarkerilor proteici în afecțiuni ale pancreasului

Doctorand **Maria Ilieș**

Coordonator științific **Prof. Dr. Felicia Loghin**

CLUJ-NAPOCA 2017



UMF
UNIVERSITATEA DE
MEDICINĂ ȘI FARMACIE
IULIU HAȚIEGANU
CLUJ-NAPOCA

CUPRINS

INTRODUCERE	15
STADIUL ACTUAL AL CUNOAȘTERII	
1. Adenocarcinomul ductal de pancreas	19
1.1. Aspecte generale	19
1.2. Factori de risc	20
1.3. Fiziopatologie	21
1.4. Diagnostic	23
1.5. Management clinic	24
1.6. Perspective	26
2. Neoplasmul papilar mucinos intraductal de pancreas	29
2.1. Aspecte generale	29
2.2. Factori de risc	29
2.3. Fiziopatologie	29
2.4. Diagnostic	31
2.5. Management clinic	32
2.6. Perspective	34
3. Studii ale biomarkerilor proteici	37
3.1. Biomarkeri proteici din țesut	38
3.2. Biomarkeri proteici din sânge	40
3.3. Perspective	41
CONTRIBUȚII PERSONALE	
1. Obiective	45
2. Metodologie generală	47
2.1. Designul studiului	52
2.2. Determinarea conținutului proteic	55
2.3. Digestia triptică a proteinelor	55
2.4. nano LC UDMS ^E	55
2.5. nano LC HRM	56
2.6. Tehnica imunoenzimatică ELISA	58
3. Studiul 1 – Influența modului de colectare a sângelui asupra eficienței depleției proteinelor abundente	
3.1. Introducere	59
3.2. Ipoteza de lucru	60
3.3. Materiale și metode	60
3.4. Rezultate	61
3.5. Discuții	62
3.6. Concluzii	63

4. Studiul 2 – Impactul modului de colectare a sângelui asupra studiului biomarkerilor proteici

4.1. Introducere	65
4.2. Ipoteza de lucru	66
4.3. Materiale și metode	66
4.4. Rezultate	66
4.5. Discuții	77
4.6. Concluzii	79

5. Studiul 3 – Cuantificarea absolută a proteinelor plasmatice prin nano LC UDMS^E

5.1. Introducere	81
5.2. Ipoteza de lucru	82
5.3. Materiale și metode	82
5.4. Rezultate	83
5.5. Discuții	90
5.6. Concluzii	93

6. Studiul 4 - Profiling-ul proteinelor plasmatice al pacienților diagnosticați cu neoplasm papilar mucinos intraductal de pancreas ca leziune precursoră a cancerului pancreatic

6.1. Introducere	95
6.2. Ipoteza de lucru	96
6.3. Materiale și metode	96
6.4. Rezultate	97
6.5. Discuții	106
6.6. Concluzii	109

7. Discuții generale

8. Concluzii generale

9. Originalitate și contribuții inovatoare ale tezei

BIBLIOGRAFIE

Cuvinte cheie: proteomică, spectrometrie de masă de înaltă rezoluție, biomarkeri diagnostic, plasmă, cancer de pancreas

INTRODUCERE

Adenocarcinomul ductal de pancreas (PDAC), cea mai comună formă de cancer al pancreasului, este o boală extrem de agresivă. Având o rată de supraviețuire de numai 8%, PDAC este a patra cauză de deces la nivel global prin cancer. Îngrijorător, PDAC va deveni în anul 2020 a doua cauză de deces din cauza unei afecțiuni maligne. De aceea, managementul PDAC reprezintă un aspect critic în sistemul de sănătate și necesită perfecționare.

Motivul major pentru prognosticul alarmant al pacienților cu PDAC este diagnosticul tardiv al bolii. Întrucât PDAC este adesea asimptomatic sau cu simptome nespecifice, peste 80% dintre pacienți sunt diagnosticați într-un stadiu avansat, când intervenția chirurgicală – în prezent singura metodă curativă – nu mai poate fi aplicată. Prin urmare, un diagnostic timpuriu al PDAC este extrem de necesar.

Neoplasmul intraductal papilar mucinos de pancreas (IPMN) este una dintre leziunile precursore ale PDAC și reprezintă o opțiune promițătoare în vederea unui diagnostic timpuriu al PDAC. IPMN este o leziune chistică a pancreasului, ușor detectabilă prin tehnici imagistice. Aproximativ 30% dintre pacienții diagnosticați cu IPMN dezvoltă PDAC într-un interval de timp de aproximativ 6 ani, oferind posibilitatea detectării timpurii a pacienților cu risc crescut. Studii recente care au evaluat acuratețea ghidurilor de diagnostic al IPMN, au arătat că există o prevalență crescută în diagnosticarea greșită a chisturilor, și prin urmare există pacienți al căror management clinic poate conduce la erori. De aceea, este absolut necesară o metodă de diagnostic mai precisă, o soluție în acest sens fiind utilizarea biomarkerilor proteici.

Studiul biomarkerilor proteici cu rol diagnostic este un domeniu de cercetare de mare actualitate, studiile anterioare generând rezultate încurajatoare, dar este necesară validarea proteinelor cu potențial de biomarker pe grupuri mai mari de pacienți. Metodologia analitică curentă vizează identificarea de paneluri proteice de tip multimarker prin aplicarea tehnicilor proteomice bazate pe spectrometria de masă de înaltă rezoluție. La momentul actual, am identificat un singur studiu care propune un set de 6 proteine plasmatică în vederea diagnosticului pacienților cu IPMN. Acest set de proteine nu a fost testat la pacienți cu PDAC, astfel încât utilitatea lui ca metodă de diagnostic timpuriu al pacienților cu risc de dezvoltare a PDAC rămâne a fi demonstrată.

Astfel, obiectivul major al acestei lucrări a fost explorarea proteomului plasmatic al pacienților cu IPMN în comparație cu cel al pacienților diagnosticați cu PDAC în vederea identificării unor potențiali biomarkeri proteici.

Utilizând cele mai avansate tehnici analitice folosite în proteomică, această lucrare a urmat o strategie progresivă: în prima parte au fost abordate aspecte metodologice utile studiului proteinelor din sânge ca potențiali biomarkeri, iar mai apoi rezultatele au fost aplicate în studiul proteinelor plasmatică cu potențial rol de biomarker de diagnostic precoce al PDAC.

CONTRIBUȚII PERSONALE

Studiul 1. Influența modului de colectare a sângelui asupra eficienței depleției proteinelor abundente

1.1 Ipoteza de lucru

Întrucât modul de colectare a probelor de sânge poate afecta într-un mod critic studiul proteinelor, evaluarea impactului acestuia asupra concentrației proteinelor cu abundență plasmatică ridicată (high abundant proteins - HAP) și a depleției acestora – o metodă de preprocesare frecvent utilizată în proteomice – este foarte importantă. Întrucât sistemul de depleție bazat pe multiafinitate este frecvent utilizat pentru depleția HAP în vederea evidențierii proteinelor mai puțin abundente, frecvent raportate ca potențiali biomarkeri, obiectivul acestui studiu a fost investigarea influenței sistemelor utilizate la colectarea sângelui asupra HAP și a eficienței depleției acestora utilizând sistemul MARS6, aplicând tehnica nano LC UDMS^E de achiziție independentă a datelor (DIA).

1.2 Materiale și metode

Șase donatori sănătoși (HC) (3 bărbați, 3 femei) cu vârsta cuprinsă între 24-29 ani, fără boli cronice și fără tratament medicamentos au fost incluși în studiu. Au fost colectate în tuburi BD Vacutainer® conținând: (i) EDTA, (ii) heparină sodică, (iii) citrat sodic, și (iv) separator de ser SST™ II câte 4 probe de sânge. În total au fost obținute 24 probe de plasmă și ser.

1.3 Rezultate

Toate probele au fost analizate înainte și după depleția a 6 dintre cele mai abundente proteine: albumina serică, IgG, IgA, serotransferina, haptoglobina, și alpha-1-antitripsina folosind coloana de depleție MARS6. În toate probele, abundența celor 6 HAP a reprezentat peste 50% din totalul de proteine. Independent de sistemul de colectare, procentele individuale ale celor 6 HAP au fost similare (fig.1):

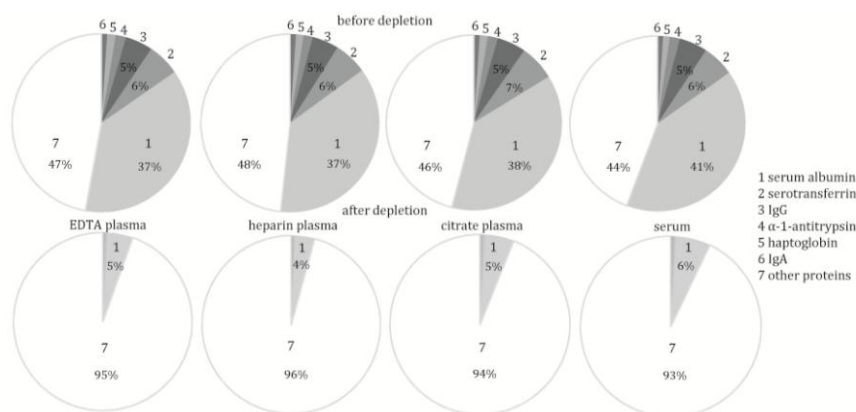


Fig. 1. Abundența celor 6 HAP înainte și după depleția acestora

1.4 Concluzii

Cele patru tipuri de anticoagulate utilizate la colectarea sângelui testate nu au avut o influență majoră asupra concentrației celor 6 proteine cu abundență plasmatică ridicată luate în studiu, eficiența depleției fiind similară. Astfel, coloana MARS6 poate fi utilizată cu succes pentru depleție în studii ale proteomului, independent de tipul de sistem de colectare a sângelui folosit.

Studiul 2. Impactul modului de colectare a sângelui asupra studiului biomarkerilor proteici

2.1 Ipoteza de lucru

Studiul biomarkerilor proteici prin metode proteomice poate fi extins semnificativ dacă se cunoaște impactul sistemelor utilizate ca anticoagulante la colectarea sângelui asupra proteinelor mai puțin abundente care sunt cel mai frecvent raportate ca potențiali biomarkeri. Astfel, obiectivul acestui studiu a fost investigarea impactului a patru sisteme utilizate la colectarea sângelui, asupra profilului proteinelor plasmatiche mai puțin abundente.

2.2 Materiale și metode

Materialele și metodele utilizate pentru acest studiu sunt identice celor utilizate în studiul 1. În acest studiu au fost analizate doar aspecte legate de proteinele mai puțin abundente.

2.3 Rezultate

În total au fost identificate și cuantificate un număr de 510 de proteine. Comparând intensitățile celor 311/313/294/304 proteine identificate în probele recoltate în tuburile cu EDTA-/heparină-/citrat/ gel separator de ser, variația medie a proteinelor a fost estimată la 0.28/0.20/0.23 și 0.20 coeficient de variație (CV). Distribuția proteinelor în cele patru tipuri de tuburi a cuprins: un set unic de 187 proteine comune celor 4 tuburi și 61/46/37 și 67 proteine reprezentând semnături ale probelor provenite prin colectarea sângelui pe EDTA-/heparină-/citrat/ gel separator de ser (fig. 2). 97% din setul de 187 proteine este reprezentat de proteine clasice predominant secretate de ficat.

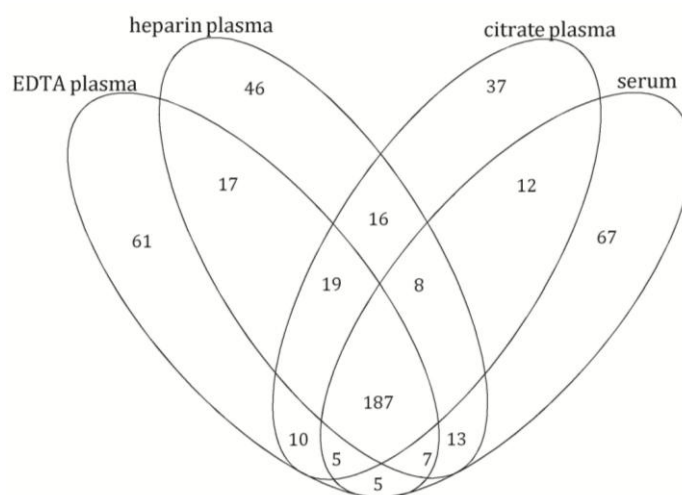


Fig. 2. Seturi de proteine comune și semnături

Pentru a permite compararea abundenței proteinelor din probele de ser și plasmă, datele au fost normalizate cu ajutorul celei mai stabile proteine din cele 4 seturi, și anume factorul I al complementului. Compararea a câte două tipuri de probe și analiza Pearson's a evidențiat o corelație puternică între toate tipurile de probe, cea mai puternică fiind între probele de plasmă recoltate pe heparină și citrat și între probele de plasmă recoltate pe EDTA și citrat. O corelație mai slabă a fost găsită între probele de plasmă și ser, dar și între proteinele din probele de plasmă colectate pe heparină și EDTA.

2.4 Concluzii

Acest studiu a identificat aspecte decisive ale diferitelor sisteme de colectare a sângelui și impactul acestora asupra studiului proteomului. Modul de colectare a sângelui nu a avut o influență majoră asupra numărului total de proteine cuantificate, astfel încât probe obținute prin diferite metode de colectare pot fi incluse în același studiu. În schimb, variabilitatea proteinelor a fost afectată de modul de colectare a probelor, studiul nostru evidențând faptul că plasma colectată pe heparină și serul sunt probele în care proteinele identificate au avut cea mai mică variabilitate, fiind cele mai reproductibile. 187 de proteine au alcătuit un set comun de proteine foarte stabil. Factorul I al complementului a fost evidențiat ca potențial candidat de normalizare a abundenței proteinelor provenind din diferite tuburi de colectare. Mai mult, au fost evidențiate semnăturile proteice în cazul celor 4 moduri de colectare a probelor luate în studiu.

Studiul 3. Cuantificarea absolută a proteinelor plasmatice prin nano LC UDMS^E

3.1 Ipoteza de lucru

Obiectivul acestui studiu a fost elaborarea și testarea unei metode simple și rapide de cuantificare absolută a seturilor de proteine cu potențial de biomarkeri de diagnostic utili în practica clinică. Metoda utilizează pentru cuantificarea absolută un standard peptidic intern provenind de la *Escherichia coli* (*E. coli*) și se bazează pe metoda descrisă de Silva *et al.* adaptată pentru analiza proteinelor plasmatice prin tehnica DIA nanoLC UDMS^E.

3.2 Materiale și metode

În acest studiu au fost analizate 6 probe de plasmă colectate pe EDTA obținute de la cei 6 voluntari incluși în studiile 1 și 2 și fără a aplica depleția proteinelor cu abundență ridicată. Standardul intern peptidic de *E. coli* (protein disaggregation chaperone ClpB, Waters, 186006012) a fost adăugat în fiecare probă (1.65 fmol/100 ng). Probele au fost analizate prin nano LC UDMS^E și proteinele au fost cuantificate pe baza concentrației standardului intern.

3.3 Rezultate

Utilizând tehnica nano LC UDMS^E, în toate cele 6 probe analizate au fost cuantificate absolut 242 proteine plasmatice. 86% din proteine sunt reprezentate de proteinele secretate. Intervalul de concentrații determinate a acoperit un domeniu de cinci ordine de mărime (domeniu dinamic 10⁵) cu o concentrație maximă medie de 839.79 μmol/L (albumina serică) și minimă de 0.01 μmol/L (paraoxonaza serică/lactonasa 3). Între acestea, 63 proteine au avut un CV ≤ 0.25 și au fost considerate ca fiind cel mai stabil set de proteine. Distribuția concentrațiilor proteinelor din acest set în raport cu toate proteinele cuantificate este ilustrată în figura 3.

Concentrațiilor absolute ale proteinelor cuantificate au fost evaluate comparativ cu datele raportate de către Hortin *et al.* (94 proteine în comun) (coeficient Pearson's $r=0.7664$) și de către Percy *et al.* (88 proteine comune) (coeficient Pearson's $r=0.6710$). Mai mult, concentrațiile plasmatice ale subunităților alfa și beta ale hemoglobinei (HBA, HBB) au fost comparate în raport cu sexul voluntarilor incluși în studiu. În general, datele studiului sunt în concordanță cu datele raportate în literatura de specialitate.

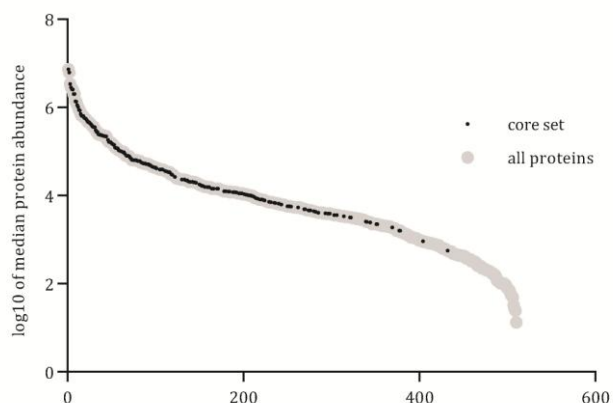


Fig. 3. Distributia concentrațiilor proteinelor din setul cel mai stabil în raport cu toate proteinele cuantificate

3.4 Concluzii

Studiul descrie o metodă de cuantificare absolută a proteinelor plasmatice utilizând tehnica DIA nano LC UDMS^E. A putut fi analizat un număr mare de proteine într-o singură analiză pe un interval dinamic, larg de concentrații și cu o stabilitate înaltă la nivelul proteinelor. Metoda utilizată este un instrument util comunității proteomice fiind simplă și rapidă, fără a necesita un pas de prefracționare, care de cele mai multe ori implică o durată mare de timp și costuri mari. Dată fiind concordanța rezultatelor obținute cu cele raportate în literatură, considerăm că aceasta metodă merită a fi validată și aplicată în practica clinică pentru studii ale biomarkerilor proteici.

Studiul 4. Profiling-ul proteinelor plasmatice al pacienților diagnosticați cu neoplasm papilar mucinos intraductal de pancreas ca leziune precursoră a cancerului pancreatic

4.1 Ipoteza de lucru

Eforturi pentru diagnosticul precoce al PDAC au fost recent îndreptate către unul dintre precursorii săi, IPMN. Câteva studii au evidențiat aspecte de biologie moleculară ale IPMN și evoluția acestuia la PDAC. Astfel, lipsa unei caracterizări complete a IPMN face ca managementul clinic al acestuia să fie expus unor controverse. De aceea, scopul acestui studiu a fost explorarea profilului proteic plasmatic al pacienților cu IPMN în comparație cu donatori, pacienți diagnosticați cu boli benigne și maligne ale pancreasului printr-o tehnică DIA MS/MS, și anume HRM.

4.2 Materiale și metode

Patru grupuri de pacienți au fost incluse în studiu: 16 HC (44-63 ani), 17 pacienți cu IPMN (34-80 ani), 15 pacienți cu pancreatită cronică (CP) (35-85 ani), și 15 pacienți cu PDAC (48-89 ani). Totodată, rezultatele au fost validate prin tehnica ELISA pentru proteinele TIMP-1, proteina S100A9 și periostin.

4.3 Rezultate

Au fost evaluate proteine cu potențial rol de diagnostic al pacienților cu IPMN în comparație cu HC (semnificativ statistic diferite și cu ordin de mărime - fold change (fc) ≥ 2). Setul propus pentru diagnostic a cuprins proteine: anhidrazele carbonice 1, 2, și 3 (CAH1 fc=5.5, CAH2 fc=15.9, CAH3 fc=3.9), și subunitățile alfa și beta ale hemoglobinei (HBA fc=3.9, HBB fc=5.1) (fig. 5A).

Proteine care ar putea ajuta la stratificarea pacienților cu boli benigne și maligne de cei sănătoși au fost reprezentate de proteine al căror nivel plasmatic este semnificativ statistic mai mic în grupul

pacienților cu IPMN dar și al HC în comparație cu PDAC (fig. 5 B) și de proteine cu abundență crescută semnificativ statistic care au inclus proteine implicate în metabolismul colesterolului și transportului de lipide (fig. 5 C). Proteine care sunt semnificativ mai puțin abundente în plasma pacienților cu IPMN comparativ cu CP și PDAC pot fi utilizate pentru identificarea pacienților cu risc scăzut de a dezvolta PDAC (fig. 5 D). Proteinele periostin și TSP1 au fost singurele proteine raportate ca fiind asociate cu malignitatea IPMN și care au fost identificate și în studiul nostru. Nivelele plasmatice ale TSP1 în IPMN au fost diferite semnificativ statistic față de HC ($fc=-1.78$), față de CP ($fc=-2.0$) și de PDAC ($fc=-2.3$). Nivelele plasmatice ale periostin nu au fost diferite semnificativ statistic, dar au fost identificate preponderent în grupurile de pacienți diagnosticați cu IPMN și PDAC (fig. 25 E).

Datele obținute prin analiza MS au fost validate prin tehnica ELISA pentru TIMP-1, proteina S100A9 și periostin. Rezultate au arătat aceeași tendință în ceea ce privește abundența proteinelor determinate în grupurile de pacienți diagnosticați cu IPMN și PDAC.

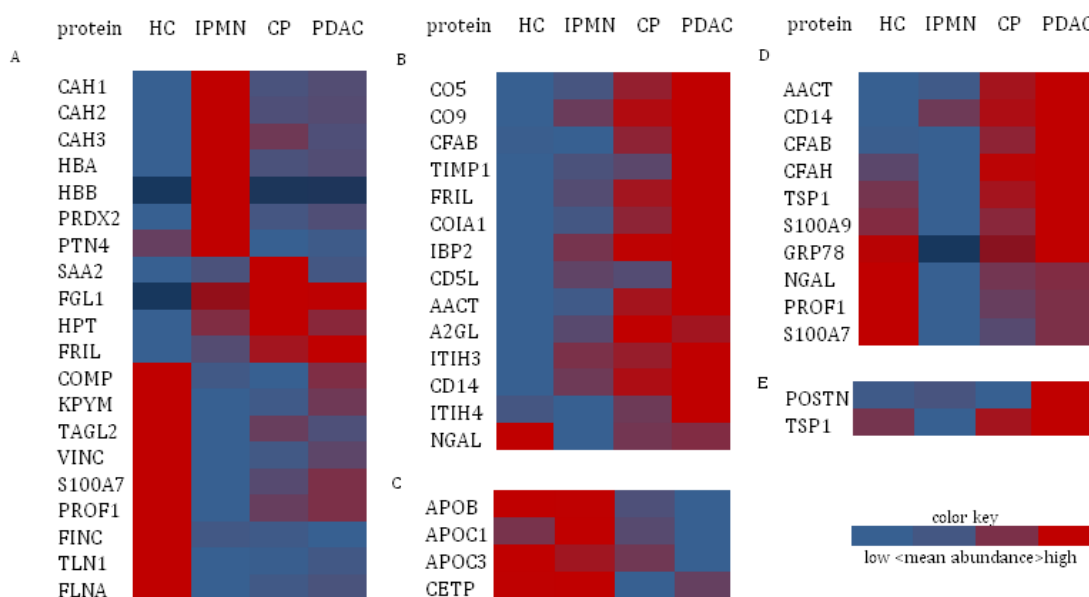


Fig. 4. Clusteri de proteine plasmatice specifice pacienților cu IPMN sunt reprezentate ca heatmaps: A) Proteine cu abundență crescută în IPMN față de HC. B) Proteine cu abundență scăzută în IPMN în comparație cu PDAC. C) Proteine cu abundență crescută în IPMN și HC în comparație cu PDAC D) Proteine cu abundență scăzută în IPMN în comparație cu CP și PDAC. E) Proteine raportate ca potențiali indicatori ai malignizării IPMN.

4.4 Concluzii

Acest studiu este primul care evaluează profilul proteic plasmatic al pacienților cu IPMN în comparație cu pacienți sănătoși, dar și cu pacienți diagnosticați cu pancreatită cronică și adenocarcinom ductal de pancreas. Profilurile proteice plasmatice ale pacienților cu IPMN sunt distincte de cele ale HC, CP și PDAC indicând alterări specifice bolii detectabile la nivel plasmatic. Proteinele identificate în acest studiu vor fi testate în cohorte mai mari de pacienți și evaluate din punct de vedere al utilității lor pentru implementarea în practica clinică ca potențiali biomarkeri de diagnostic. Nu au fost până în prezent disponibile date referitoare la evoluția IPMN spre PDAC pentru setul de pacienți testați. Seturile de proteine identificate merită validate în studii viitoare pentru identificarea de potențiali biomarkeri cu rol de diagnostic timpuriu al PDAC.

Concluzii generale

PDAC este o boală malignă extrem de agresivă, fiind a patra cauză de deces la nivel global. Chiar dacă au fost întreprinse numeroase studii în vederea identificării unor metode de diagnostic timpuriu al PDAC, acesta nu beneficiază încă de astfel de metode în practica clinică curentă. Acestea au condus la îndreptarea atenției către unul dintre precursorii PDAC, și anume IPMN. Astfel, obiectivul general al acestei lucrări a fost investigarea proteomului ca potențială sursă de biomarkeri de diferențiere în bolile pancreasului.

În primele două studii a fost investigat impactul a patru sisteme de colectare a sângelui asupra profilului proteic plasmatic utilizând o tehnică DIA nano LC UDMS^E. Au fost identificate aspectele de luat în considerare atunci când se realizează design-ul experimental în studiile de proteomică. Astfel, tipul de anticoagulat utilizat la colectarea sângelui: (i) nu are un impact major asupra concentrației și eficienței depleției proteinelor din sânge cu abundență ridicată investigate: albumina serică, serotransferrina, IgG, haptoglobina, IgA, α -1-antitrypsina, (ii) eficiența depleției acestora utilizând coloana MARS6 este similară, (iii) nu a influențat numărul total de proteine cuantificate, (iv) influențează variabilitatea la nivel proteic. Studiul evidențiază faptul că probele de plasmă colectate pe heparină și cele de ser au cea mai mică variabilitate la nivel proteic. Independent de modul de colectare a probelor au fost identificate (i) un set de 187 de proteine cu o stabilitate înaltă dar și (ii) proteine unice fiecărui tip de probă. Așadar, modul de colectare trebuie considerat ca fiind un factor pre analitic extrem de important în studii de profiling în vederea identificării de biomarkeri.

Al treilea studiu prezintă o metodă simplă și rapidă de cuantificare absolută a proteinelor. Într-o singură analiză, fără depleția prealabilă a proteinelor abundente, prin această metodă se poate cuantifica un număr mare de proteine, pe un domeniu dinamic de concentrații de cinci ordine de mărime și cu stabilitate mare la nivel proteic. Astfel, această metodă demonstrează un mare potențial de a fi aplicată în studiile biomarkerilor.

Studiul patru prezintă caracterizarea proteomului plasmatic al pacienților diagnosticați cu IPMN, ca precursor al PDAC. Bazat pe rezultatele anterioare, biofluidul utilizat în acest studiu a fost plasma recoltată pe heparină. Astfel, proteomul plasmatic al IPMN a fost caracterizat prin profiling proteic global și au fost evidențiate seturi distinctive de proteine pentru donatori sănătoși, pacienți diagnosticați cu pancreatită cronică și PDAC. Setul de proteine cu rol de diagnostic al IPMN în comparație cu donatori sănătoși a cuprins următoarele proteine: anhidrazele carbonice și subunități ale hemoglobinei. Componente ale complementului și proteina ferritin light chain fac parte din setul de proteine cu potențial de diferențiere al pacienților diagnosticați cu boli maligne și benigne. Proteina TSP1 a fost identificată cu potențial rol de indicator al progresului IPMN la PDAC. Aceste proteine vor fi verificate și validate în studii viitoare, iar valoarea lor în stratificarea riscului pacienților cu IPMN de a dezvolta PDAC va fi testată.

Studiile realizate aduc o contribuție importantă atât pentru îmbunătățirea metodologiei analitice de evaluare a proteomului în general dar și pentru investigarea potențialilor biomarkeri cu utilitate clinică în bolile pancreasului.