



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학박사 학위논문

**Biomarker Development of Intraductal
Papillary Mucinous Neoplasm and
Breast Cancer using Quantitative
Proteomics and Bioinformatics**

정량 단백질체학 및 생물정보학을
이용한 췌관내 유두상 점액 종양 및
유방암의 바이오마커 개발

2021년 02월

서울대학교 대학원

의과학과 의과학전공

도 미 솔

A thesis of the Degree of Doctor of Philosophy

**정량 단백질체학 및 생물정보학을
이용한 췌관내 유두상 점액 종양 및
유방암의 바이오마커 개발**

**Biomarker Development of Intraductal
Papillary Mucinous Neoplasm and
Breast Cancer using Quantitative
Proteomics and Bioinformatics**

February 2021

Major in Biomedical Sciences

Department of Biomedical Sciences

Seoul National University

Graduate School

Misol Do

정량 단백질학 및 생물정보학을
이용한 체관내 유두상 점액 종양 및
유방암의 바이오마커 개발

지도교수 김 영 수

이 논문을 이학박사 학위논문으로 제출함

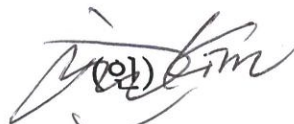
2020년 10월

서울대학교 대학원
의과학과 의과학전공
도 미 술

도미술의 이학박사 학위논문을 인준함

2020년 12월

위 원 장 김 종 서

(인) 

부위원장 김 영 수

(인) 

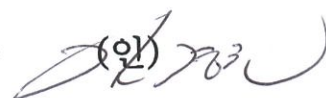
위 원 이 용 석

(인) 

위 원 유 한 석

(인) 

위 원 김 경 곤

(인) 

Biomarker Development of Intraductal Papillary Mucinous Neoplasm and Breast Cancer using Quantitative Proteomics and Bioinformatics

by

Misol Do

**A thesis submitted to the Department of Biomedical Sciences
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Biomedical Sciences at
Seoul National University Graduate School**

December 2020

Approved by Thesis Committee:

Professor Jong-Seo Kim Chairman

Professor Youngsoo Kim Vice chairman

Professor Yong-Seok Lee

Professor Han Suk Ryu

Professor Kyunggon Kim

ABSTRACT

Biomarker Development of Intraductal Papillary Mucinous Neoplasm and Breast Cancer using Quantitative Proteomics and Bioinformatics

Misol Do

Major in Biomedical Sciences

Department of Biomedical Sciences

Seoul National University

Graduate School

Introduction: Mass spectrometry (MS)-based proteomic approaches are being increasingly applied to identify markers that are related to specific diseases, based on their ability to screen thousands of proteins simultaneously to obtain hundreds of differentially expressed proteins (DEPs) in small amounts of samples. In general, pathologic specimens collected from clinical cohorts, such as body fluids and formalin-fixed paraffin-embedded (FFPE) tissues, are analyzed. For proteomic analysis, MS-based approach is a powerful tool in biomarker discovery and clinical diagnosis with its high throughput and high sensitivity. Also, a proteomics study will help understand biological mechanisms of diseases.

Methods: In chapter I and II, high-resolution mass spectrometry-based proteomics

was performed using pancreatic cyst fluid samples to discover marker candidates for predicting the degree of intraductal papillary mucinous neoplasm (IPMN) malignancy. In chapter II, samples were collected from the expanded cohort that included IPMNs and other PCLs (mucinous cystic neoplasm (MCN) and serous cystic neoplasm (SCN)) to better reflect actual clinical circumstances. In chapter III, a targeted proteomic technique, multiple reaction monitoring-mass spectrometry (MRM-MS), was applied to formalin-fixed paraffin-embedded (FFPE) tissues to establish a novel assay to determine human epidermal growth factor receptor 2 (HER2) status in breast cancer patients.

Results: In chapter I, a dataset of 2,992 proteins was constructed from pancreatic cyst fluid samples of IPMN patients. Eighteen biomarker candidates that were differentially expressed across histological grades of IPMN were discovered, and some of them were validated by western blot in an independent cohort, the results of which were consistent with our proteomic data. In chapter II, 5,834 proteins were identified using cyst fluid from patients with IPMN, MCN, and SCN. Among 364 proteins that differentially expressed between IPMN dysplasia, 19 final marker candidates consistently increased or decreased with greater IPMN malignancy. CD55 was validated in an independent cohort by ELISA, Western blot, and IHC, and the results were consistent with the MS data. In chapter III, we established an MRM-MS assay that improves on existing methods for differentiating HER2 status. The accuracy and precision of HER2 quantification were improved by simplifying the sample preparation through predicting the number of FFPE slides required to ensure an adequate amount of protein and using the expression levels of an epithelial cell-

specific protein as a normalization factor when measuring HER2 expression levels.

Conclusions: In chapter I and II, we have generated the largest proteomic dataset of pancreatic cyst fluid to date and discovered potential markers of IPMN dysplasia. The development of cyst fluid markers can facilitate an accurate assessment of the degree of IPMN malignancy and effectively guide surgical decision-making. Ultimately, if the developed marker is implemented in clinical practice, the accurate assessment of IPMN dysplasia will prevent unnecessary surgical resection for low-risk IPMN patients. In chapter III, our proposed protocol, which discriminates between equivocal HER2 subgroups, can potentially decrease the time and costs required for the diagnosis of breast cancer patients by reducing the number of cases that require ancillary fluorescence in situ hybridization (FISH) tests. In addition, the simplified assay procedure can reduce the barriers to entry for the clinical application of the MRM-MS assay. Our MRM-MS assay yields more accurate HER2 expression levels relative to immunohistochemistry and should help to guide clinicians toward the proper treatment for breast cancer patients, based on their HER2 expression.

Keywords: Proteomics; Mass spectrometry; Multiple Reaction Monitoring; Pancreatic cyst fluid; Intraductal papillary mucinous neoplasm (IPMN); Biomarker; Human epidermal growth factor receptor 2 (HER2); Formalin-fixed paraffin-embedded (FFPE)

Student number: 2014-25061

* This work is published in Clinical Proteomics Journal. Quantitative proteomic analysis of pancreatic cyst fluid proteins associated with malignancy in intraductal papillary mucinous neoplasms (M. Do, D. Han, Joseph I. Wang, H. Kim, W. Kwon, Y. Han, J.-Y. Jang and Y. Kim). Published 18 Apr 2018/ Clinical Proteomics 10.1186/s12014-018-9193-1.

* This work is published in Cancers Journal. Marker Identification of the Grade of Dysplasia of Intraductal Papillary Mucinous Neoplasm in Pancreatic Cyst Fluid by Quantitative Proteomic Profiling (M. Do, H. Kim, D. Shin, J. Park, H. Kim, Y. Han, J.-Y. Jang and Y. Kim). Published 23 Aug 2020/ Cancers 10.3390/cancers12092383.

* This work is published in Clinical Chemistry Journal. Clinical Application of Multiple Reaction Monitoring-Mass Spectrometry to Human Epidermal Growth Factor Receptor 2 Measurements as a Potential Diagnostic Tool for Breast Cancer Therapy (M. Do, H. Kim, I. Yeo, J. Lee, I.A. Park, H.S. Ryu and Y. Kim). Published 12 Sep 2020/ Clinical Chemistry 10.1093/clinchem/hvaa178.

CONTENTS

Abstract	i
Contents	v
List of Tables	vii
List of Figures	x
List of Abbreviations	xv
General Introduction	1
Chapter I	5
Quantitative Proteomic Analysis of Pancreatic Cyst Fluid Proteins Associated with Malignancy in Intraductal Papillary Mucinous Neoplasms	
 Introduction	6
 Materials and Methods	10
 Results	19
 Discussion	42
Chapter II	48

**Marker Identification of the Grade of Dysplasia of Intraductal Papillary
Mucinous Neoplasm in Pancreatic Cyst Fluid by Quantitative Proteomic
Profiling**

Introduction	49
Materials and Methods	53
Results	64
Discussion	99

Chapter III107

**Clinical Application of Multiple Reaction Monitoring-Mass
Spectrometry to Human Epidermal Growth Factor Receptor 2
Measurements as a Potential Diagnostic Tool for Breast Cancer Therapy**

Introduction	108
Materials and Methods	111
Results	124
Discussion	148

General Conclusion153

References156

Abstract in Korean176

LIST OF TABLES

Chapter I

Table 1. Demographic and clinical characteristics of the study population20

Table 2. Detailed statistical analysis of 18 final marker candidates38

Chapter II

Table 1. Demographic and clinical characteristics of the study population	54
Table 2. Detailed statistical analysis of 19 final marker candidates	91
Table 3. Results of upstream regulator analysis in IPA	92

Chapter III

Table 1. Demographic and clinical characteristics of the study population in the MRM-MS analysis (N = 210).....	112
Table 2. The final protein list and peptide sequences analyzed in the MRM-MS assay	119
Table 3. The discriminatory power of the normalized HER2 expression levels for 120 equivocal HER2 cases	140
Table 4. Cross-validation results of the single- and multi-marker analyses, as measured by MRM-MS assay, between the training and test sets.....	145

LIST OF FIGURES

Chapter I

Figure 1. Experimental workflow	21
Figure 2. The number of identified and quantified proteins in individual samples and the peptide library	23
Figure 3. Dynamic range of quantified proteins.....	25
Figure 4. Comparative analysis with other proteomic databases	27
Figure 5. Coefficient of variation (CV%) values of technical replicates in individual samples and scatterplots of Pearson correlation coefficients.....	29
Figure 6. Volcano plots and hierarchical heat map clusters of significant proteins by student's t-test.....	32
Figure 7. Gene ontology analysis	33
Figure 8. Ingenuity Pathway Analysis	34
Figure 9. 18 biomarker candidates that had expression patterns that were consistent with the degree of IPMN malignancy.....	39
Figure 10. Validation of HOOK1, PTPN6, and SERPINA5 as potential biomarker targets by western blot.....	41

Chapter II

Figure 1. Experimental workflow	67
Figure 2. Comparison of protein identification and quantification in total dataset and in sample groups	68
Figure 3. Dynamic range of quantified proteins.....	70
Figure 4. Coefficient of variation (CV%) values and Pearson correlation coefficients between technical triplicates in each individual sample.....	72
Figure 5. Comparative analysis with various proteome databases and other proteomic studies.....	74
Figure 6. Comparison of the expression patterns of the final marker candidates between our previous and present studies	76
Figure 7. Flowchart of the discovery of potential markers of IPMN dysplasia	78
Figure 8. Volcano plots of differentially expressed proteins in three comparison groups	79
Figure 9. Results of Gene Ontology and KEGG pathway analyses.....	82
Figure 10. Ingenuity Pathway Analysis.....	83
Figure 11. Overview of protein expression of the 19 final marker candidates of IPMN dysplasia.....	88

Figure 12. 19 potential markers that were differentially expressed in accordance with the degree of IPMN malignancy.....	89
Figure 13. The dynamic range of protein fold-changes in comparisons 1 and 3 and the results of upstream regulator analysis in Ingenuity Pathway Analysis (IPA)	90
Figure 14. Validation of CD55 as a potential biomarker target by ELISA.....	96
Figure 15. Validation of CD55 as a potential marker by Western blot	97
Figure 16. Immunohistochemical staining of CD55 and MPO	98

Chapter III

Figure 1. Determination of optimal incubation time for tryptic digestion	115
Figure 2. Schematic of overall procedure, from sample preparation to MRM-MS assay, for determining HER2 status.....	116
Figure 3. Calibration curve, constructed from plotting the extracted protein amount against the total cell count in the tumor area	127
Figure 4. Comparison of six histopathological characteristics between two sample groups.....	128
Figure 5. Evaluation of calibration curves.....	132
Figure 6. Measurement of purity of unpurified synthetic peptide.....	133
Figure 7. Evaluation of stability and reproducibility of HER2 and JAM1 surrogate peptides.....	137
Figure 8. Scatterplots of Spearman rank correlation coefficients between six HER2 peptides	141
Figure 9. Area under the receiver operating curve (AUROC) with respect to normalized quantified data on HER2 peptide in HER2 2+/FISH-negative versus HER2 2+/FISH-positive.....	143
Figure 10. Scatter dot plots and an area under the receiver operating curve (AUROC) of the light-to-heavy peptide peak area ratios for a HER2 surrogate peptide normalized by those for a JAM1 surrogate	

peptide plotted against IHC/FISH score	147
Figure 11. Conventional diagnostic strategy using immunohistochemistry (IHC) and the proposed strategy using MRM-MS assay for HER2-targeted therapy	152

LIST OF ABBREVIATIONS

IPMN: intraductal papillary mucinous neoplasm

LGD: low-grade dysplasia

HGD: high-grade dysplasia

MCN: mucinous cystic neoplasm

SCN: serous cystic neoplasm

PCL: pancreatic cystic lesion

MRI: magnetic resonance imaging

CT: computed tomography

EUS: endoscopic ultrasound

CEA: carcinoembryonic antigen

CA19-9: carbohydrate antigen 19-9

BCA: bicinchoninic acid

SDS: sodium dodecyl sulfate

DTT: dithiothreitol

FASP: filter-aided sample preparation

DEP: differentially expressed protein

GO: gene ontology

IPA: Ingenuity Pathway Analysis

TMHMM: transmembrane helices

HER2: Human epidermal growth factor receptor 2

FISH: fluorescence in situ hybridization

MRM-MS: multiple reaction monitoring-mass spectrometry

FFPE: formalin-fixed paraffin-embedded

JAM1: junctional adhesion molecule A

IHC: Immunohistochemistry

SIS: stable isotope-labeled internal standard

PAR: peak area ratio

LC: liquid chromatography

AUROC: area under the receiver operating curve

CI: confidence interval

GENERAL INTRODUCTION

Biomarkers are constantly modified or present at abnormal concentrations in specific diseases or other health conditions. Thus, disease biomarkers are important for determining the stage of disease or predicting disease progression, or might expect the effect of a particular treatment on clinical outcomes. Among various types of biomarkers, protein biomarkers are regarded as the most ubiquitously affected in disease, response, and recovery. There are two approaches to biomarker studies: (1) discovering novel biomarkers for the diagnosis of diseases that lack a definitive diagnostic method, (2) developing a novel technique for detecting existing biomarkers, such as human epidermal growth factor receptor 2 (HER2), more accurately and economically.

Among several protein assays for discovering disease biomarkers, mass spectrometry (MS)-based proteomic approaches has become the preferred method for biomarker studies of various human samples, based on their ability to screen thousands of proteins simultaneously in small amounts of samples. In addition, MS-based proteomics is increasingly regarded as an effective tool for complementing the limitations of conventional techniques, such as immunoassay, for detecting protein biomarkers due to its high analytical sensitivity, reproducibility, accuracy, and precision.

Various clinical specimens, such as blood, cyst fluid, frozen tissue, and FFPE tissue, are used for biomarker studies. Blood has been the most extensively studied body fluid for biomarker studies because proteins secreted by tumor cells are

transported to the circulation by drainage through lymphatic or capillary systems. However, proteins related to tumor exist in blood at very low levels, about 1 – 10 pg/ml or less. Therefore, it is difficult to discover biomarkers using blood and to identify the association between biomarkers and the specific disease. On the other hand, tumor fluid, such as cyst fluid, is regarded as a rich source of proteins related to the tumor because it contains secreted proteins from the surrounding tumor cells and neighboring stroma. The concentration of disease biomarkers in the local tumor microenvironment is estimated to be 1000 – 1500 times higher than in blood. In addition, FFPE tissues have been instrumental in MS-based proteomic studies due to their abundance and accessibility; vast archives of pathologically characterized clinical samples exist in abundance as FFPE tissue can be stored over extended periods without requiring expensive equipment.

In chapter I, pancreatic cyst fluid proteins were investigated by using high-resolution mass spectrometry to discover potential biomarkers of the degree of intraductal papillary mucinous neoplasm (IPMN) malignancy. In addition, using several bioinformatics tools, we detailed IPMNs at the molecular level. The incidence of patients with pancreatic cystic lesions, particularly IPMN, is increasing. However, the lack of a definitive diagnostic method has led to low-risk IPMN patients undergoing unnecessary surgeries. Thus, the development of cyst fluid markers can facilitate an accurate assessment of the degree of IPMN dysplasia and effectively guide surgical decision-making. Ultimately, if the developed marker is implemented in clinical practice, the accurate assessment of IPMN dysplasia will prevent unnecessary surgical resection for low-risk IPMN patients.

In chapter II, pancreatic cyst fluid from intraductal papillary mucinous neoplasm (IPMN) and other pancreatic cystic lesions (mucinous cystic neoplasm (MCN), and serous cystic neoplasm (SCN)) were investigated to discover more clinically relevant biomarker candidates of IPMN malignancy. To increase depth of proteome coverage, pooled cyst fluid (comprising equal amounts of individual samples), secreted proteins from PANC1, Mia Paca-2, BxPC3, and pooled cell lysates from the 3 cell lines were compiled to generate a peptide library. Consequently, we discovered potential markers of the histological grades of IPMN in a larger pool of proteins. We aimed to generate a logically sound process for discovering potential markers of IPMN dysplasia and discover reliable marker candidates in accordance with histological grades of IPMN. In addition, using several bioinformatics tools, such as gene ontology (GO) analysis and ingenuity pathway analysis (IPA), we intended to detail IPMN malignancy at the molecular level. The final potential markers, which reflect actual clinical circumstances, can help classify various PCLs and avoid unnecessary pancreatic resection for low-risk IPMN patients.

In chapter III, we established a novel MRM-MS assay that improves on existing methods for differentiating HER2 status by using FFPE tissue specimens. The accurate detection of HER2 is crucial for providing the appropriate measures for breast cancer patients. However, the current techniques used to detect HER2 status, immunohistochemistry and fluorescence in situ hybridization (FISH) have limitations. Our proposed protocol can improve the accuracy and precision of HER2 quantification by simplifying the sample preparation and applying a novel

normalization factor for better measuring HER2 expression levels. This MRM-MS assay yields more accurate HER2 expression levels relative to immunohistochemistry and should help to guide clinicians toward the proper treatment for breast cancer patients, based on their HER2 expression.

CHAPTER I

Quantitative Proteomic Analysis of Pancreatic Cyst Fluid Proteins Associated with Malignancy in Intraductal Papillary Mucinous Neoplasms

INTRODUCTION

Intraductal papillary mucinous neoplasms (IPMNs) are precancerous lesions that grow in the pancreatic ducts and are characterized by papillary growth of the ductal epithelium. The production of thick mucinous fluid, another hallmark of IPMNs, causes cystic dilation and can progress into pancreatic ductal adenocarcinoma (1). Depending on the malignancy, IPMN is classified as low-grade dysplasia (LGD), intermediate-grade dysplasia (IGD), high-grade dysplasia (HGD), and invasive IPMN. According to the official guidelines for managing pancreatic IPMN, only patients with HGD or invasive IPMN require surgery, because they are at higher risk of their disease developing into cancer (2). Milder forms of IPMN can be managed with active surveillance and do not warrant surgical intervention. However, current methods for assessing the histological grades of IPMNs are unreliable, and as a result, patients with milder IPMN are often subjected to unnecessary operations (3).

In clinical practice, MRI and CT scans, cytological examination of cyst fluid, measurement of tumor markers such as carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA 19-9), and analysis of GTPase Kras (KRAS) and guanine nucleotide-binding protein alpha subunit (GNAS) mutations are used to categorize patients with pancreatic cysts (3-6). Features of pancreatic images in MRI or CT scans are generally used to assess the potential malignancy of cysts but have low diagnostic accuracy—up to 40% of neoplastic cysts are misdiagnosed as pseudocysts, and the overall accuracy ranges from 20% to 80% (7). Cytological

examination of pancreatic cyst fluid is an alternative approach, but it has difficulties in identifying the existence of malignancy when sufficient sample volumes are unavailable (8). Differentiating mucinous cysts from other cystic lesions by measuring carcinoembryonic antigen levels in cyst fluid has relatively low accuracy (79% sensitivity, 73% specificity) (9). Similarly, as shown by Frossard et al., CA 19-9, a pancreatic cancer marker, also performs poorly in distinguishing mucinous cysts and other lesions, with 15% sensitivity and 81% specificity (10). Analyzing GNAS mutations are only applicable for samples that are acquired during the early stages of IPMN (11). The general consensus is that existing methods for diagnosing IPMN histological grades are imprecise and unreliable, even when used in tandem (12).

Because pancreatic cyst fluid contains secreted proteins from tumor cells at higher proportions, several groups, such as Poersch et al., have concluded that it is a better experimental model of IPMN histological grades than serum and plasma (13). Consequently, pancreatic cyst fluid has been widely favored in recent research on IPMN, because it is obtainable by endoscopic ultrasound-guided fine needle aspiration biopsy, which is minimally invasive (10). Many studies have focused on discovering protein markers that differentiate mucinous from nonmucinous cyst fluid and cyst fluid that is related to IPMN dysplasia, based on DNA methylation and telomerase activity, as demonstrated by Hata et al. (14). Diagnosing histological grades of IPMN using pancreatic cyst fluid by proteomic analysis is a relatively unexplored area (15). Thus, the IPMN dysplasia proteome has not been characterized extensively.

Cuoghi et al. performed a cursory profiling study of the proteomic patterns

of pancreatic cyst fluids from various cystic lesions, including IPMN, MCN, serous cystadenomas, pancreatic neuroendocrine tumors, and pseudocysts, identifying 220 to 727 proteins in these fluids. Specifically, 243 proteins were identified in the IPMN groups (16). Gbormittah et al. characterized glycoproteins and nonglycoproteins in mucinous and nonmucinous pancreatic cyst fluid to identify DEPs as potential biomarker targets. They found 230 proteins in mucinous subtypes and 290 proteins in nonmucinous subtypes; the DEPs between mucinous and nonmucinous cyst fluid were associated with lipid metabolism, energy metabolism, and stress responses (17). These studies were unable to determine the IPMN histological grades, merely differentiating between mucinous and nonmucinous cyst fluid. These recent studies demonstrate that the current cyst fluid proteome lacks the coverage to extrapolate meaningful conclusions on the molecular and biological activities of the identified proteins, which ultimately impedes our understanding of IPMN histology in terms of proteomic differences and biological functions.

In this report, we aimed to comprehensively identify pancreatic cyst fluid proteins and discover differentially expressed proteins in accordance with histological grades of IPMN. Recently, we reported a platform for in-depth profiling of pancreatic cyst fluid (18). Using this platform, the protein expression patterns of pancreatic cyst fluid were analyzed on a high-resolution mass spectrometer to discover potential biomarkers of IPMN histological grades. Subsequently, we validated some of the 18 candidate markers by western blot. We report here that pancreatic cyst fluid is a valuable source for biomarker studies as it contains putative markers related to IPMNs and that bioinformatics analyses using identified proteins

of cyst fluid enhance our understanding of IPMNs at the molecular level. We ultimately intend to discover marker candidates that can help patients avoid unnecessary operations.

MATERIALS AND METHODS

1. Clinical samples

Cyst fluid samples were collected from 9 IPMN patients during their pancreatectomies at Seoul National University Hospital (Seoul, South Korea) from April 2013 to December 2015. At least 200 μ L of cyst fluid was aspirated from each patient. The samples were then snap-frozen in liquid nitrogen and stored at -80°C . All patients consented to participation in the study in accordance with Institutional Review Board guidelines (IRB No. 1301-095-458). IPMN samples were divided into low-grade dysplasia (LGD, $n=3$), high-grade dysplasia (HGD, $n=3$), and invasive IPMN ($n=3$).

2. Pancreatic cyst fluid protein sample preparation

Each pancreatic cyst fluid sample was transferred to an Eppendorf tube. Viscous samples that could not be pipetted were sonicated briefly (Sonics & Materials Inc., USA) to remove the mucus. All samples were centrifuged at 15,000 rpm for 20 min at 4°C , and the supernatant was placed into a new tube. The protein concentration was estimated using a BCA reducing agent compatibility assay kit (Thermo Scientific, Rockford, IL, USA). Equal portions of each sample were pooled to create a peptide library from 600 μ g of proteins. One hundred micrograms of individual protein samples were used for label-free quantification. Cold acetone (Sigma-Aldrich, USA) was added to the supernatant to the ratio of 5:1 (v/v) to precipitate

the proteins. The mixture was vortexed gently and incubated overnight at -20 °C. The precipitate was centrifuged for 10 minutes (15,000 rpm at 4 °C), and the supernatant was carefully decanted, after which 500 µL cold acetone was added to the pellet. After this wash step, the pellet was centrifuged for 10 min (15,000 rpm at 4 °C). The remaining acetone was poured off, and the pellet was air-dried for 2 h.

3. Protein digestion and desalting

The pellet was dissolved in 30 µL of lysis buffer (4 % SDS, 0.1 M DTT, 0.1 M Tris-Cl, pH 7.4). The mixture was gently vortexed and boiled for 30 min at 95 °C. The boiled mixture was then transferred through a 30-kDa cutoff filter (Amicon® Ultra, Millipore, USA) with 300 µL 8 M urea (8 M Urea, 0.1 M Tris-Cl, pH 8.5) and centrifuged (14,000 g, 15 min, 20 °C). This filtration step was repeated twice to dilute and lower the SDS concentration. Next, 200 µL 50 mM IAA (50mM IAA, 8 M urea, 0.1 M Tris-Cl, pH 8.5) was added to each sample and incubated for 1 h at 25 °C. Each sample was then centrifuged and washed twice with 300 µL 8 M urea and then 3 times with 300 µL 40 mM ammonium bicarbonate (ABC).

After the samples were centrifuged, 100 µL 40 mM ABC and 0.1 µg/µL trypsin (at a trypsin:sample ratio of 1:80, wt/wt) were added to each sample and incubated for 18 h at 37 °C. Next, the filters (9 individual samples, 1 pooled sample) were transferred to new collection tubes, which were centrifuged after 100 µL 40 mM ABC was added. Fifty microliters NaCl was added to each individual sample, and 50 µL water was added to the pooled sample. The pooled sample underwent an

additional digestion step (19, 20). Again, the filter unit was transferred to a new tube and centrifuged after 200 μ L 8 M urea was added. Then, the unit was centrifuged twice with 300 μ L 40 mM ABC. One-tenth of the concentration of trypsin that was used in the first digestion step was added with 100 μ L 40 mM ABC, and the unit was incubated for 18 h at 37°C. Next, the filter was transferred to another tube, and the peptides were collected by sequential centrifugation with 100 μ L 40 mM ABC and 50 μ L 0.5 M NaCl.

Prior to acidification and desalting, all tryptic peptides were measured by tryptophan fluorescence assay to determine the volume that was required to extract the same amount of peptides from each sample (21). The equalized amounts of peptides were then set aside for label-free quantification. The measured peptides were acidified with 10 μ L 10% TFA and desalted with homemade C18-StageTip columns as described (22). The desalted peptides were then lyophilized on a speed-vacuum centrifuge and stored at -80°C.

4. Peptide fractionation by high-pH reverse phase fractionation

To increase the number of identified proteins, the pooled cyst fluid sample was fractionated using 2 methods: modified Stage-tip-based high-pH peptide fractionation (23) and offline HPLC high-pH fractionation on an Agilent 1260 Bio-inert. For Stage-tip fractionation, half of the lyophilized peptides were dissolved in 200 μ L of loading buffer (15 mM ammonium hydroxide solution, pH 10, and 2% acetonitrile) and separated on a pipette-based C18 RP microcolumn. The column was constructed by plugging the bottom of a 200 μ L transparent pipette tip with C18

Empore disk membrane (3M, Bracknell, UK) and packing the tip with POROS 20 R2 resin. The plugged tip was rinsed 3 times with 100 μ L 100% methanol and then 3 times with 100 μ L 100% acetonitrile (ACN). The column was then conditioned with 100 μ L of loading buffer using a syringe. The peptides were loaded onto the column at pH 10. An ACN gradient of 2, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5, 30, 32.5, 35, 40, 50, 60, 70, 80, and 100% was used to elute 20 fractions, which were collected into 6 tubes discontinuously to distribute eluents of varying hydrophobicity. These 6 fractions were lyophilized in a speed-vacuum centrifuge and stored at -80°C. The remaining half of the lyophilized peptides was dissolved in 80 μ L of loading buffer (15 mM ammonium hydroxide in water, pH 10). The peptides were loaded onto the column, and 96 (2mL Square Collection Plate, Waters, UK) fractions were eluted by applying an ACN gradient (pH 10, 5%~35%) for 40 minutes at a flow rate of 0.2 mL/min and washing the column with 90% ACN for 10 minutes at 0.2 mL/min. The ACN gradient was established by mixing varying proportions of solution A (0.1% formic acid in HPLC-grade distilled water) and solution B (0.1% formic acid in ACN). The 96 fractions were concatenated according to the column number of the plate to produce 12 pooled fractions. The resulting 12 tubes were lyophilized in a speed-vacuum centrifuge and stored at -80°C.

5. LC-MS/MS analysis

The peptide samples were analyzed using an LC-MS/MS configuration, comprising an Easy-nLC 1000 (Thermo Fisher Scientific, Waltham, MA) that was coupled to a Q Exactive mass spectrometer with a nanoelectrospray ion source (Thermo Fisher

Scientific, Waltham, MA), per our established protocol (18). Peptides were separated on a 2-column system that was composed of a trap column (75 μm I.D. x 2 cm, C18 3.0 μm , 100 Å) and an analytical column (50 μm I.D. x 15 cm, C18 3.0 μm , 100 Å). Fractionated peptides were subjected to an ACN gradient (6% to 60%) for 235 minutes. The gradient was created by mixing solvent A (2% ACN and 0.1% v/v formic acid) and solvent B (100% acetonitrile and 0.1% v/v formic acid) at various proportions. The spray voltage was set to 2.0 kV in positive ion mode, and the temperature of the heated capillary was set to 320 °C. Mass spectra were acquired in data-dependent mode by top 20 method on an Orbitrap analyzer with a mass range of 350-1700 m/z and a resolution of 70,000 at m/z 200. HCD scans were acquired at a resolution of 17,500. HCD peptide fragments were acquired at a normalized collision energy (NCE) of 27. The maximum ion injection time for the survey scan and MS/MS scan was 20 ms and 80 ms, respectively. All samples were analyzed in 3 technical replicates.

6. Raw data search

The MS data from the Q Exactive were processed in MaxQuant (version 1.5.5.1 with built-in Andromeda search engine) (24). Precursor MS signal intensities were determined, and HCD MS/MS spectra were de-isotoped and filtered, such that only the 20 most abundant fragments per 100 m/z range were retained. Protein groups were identified by searching the MS and MS/MS data of the peptides against the Uniprot human database (2014 December, 88,717 entries). Both the forward and reverse amino acid sequences were taken into account when calculating the false

discovery rate (FDR). Following established target-decoy search procedures (25), search results were filtered at FDR <1% for identifying peptides, modification sites, and proteins. The search was conducted in digestion mode trypsin/P, which assumes cleavage at carboxyl sides of lysine and arginine, including cases where the subsequent residue is a proline.

The following parameters were used in the database search: precursor and HCD fragment mass tolerances of 6 ppm and 20 ppm, respectively; tolerance of up to 2 missed cleavages; carbamidomethylation of cysteine as a fixed modification; and oxidation of Met and acetylation of protein N-term as variable modifications. The minimum peptide length was set to 6 residues. Peptides were assigned to protein groups by the principle of parsimony (26). The principle is applied to derive the smallest list of probable protein groups that adequately represent the identified peptides, which reduces sequence redundancy issues. All proteomics data in this report have been deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org/>) through the PRIDE partner repository: dataset identifier PXD008302 (27).

7. Label-free quantification and statistical analysis

Label-free quantification (LFQ) and statistical analysis were performed in MaxQuant (version 1.5.5.1) and Perseus (version 1.5.8.5), respectively, according to our previous studies (18). Protein abundance was obtained from LFQ intensity values. LFQ intensity was calculated as described by the equation by Cox et al (28). Each of the three histological groups in this study had three biological replicates,

which in turn had three technical replicates each. Thus, a total of 9 LFQ intensity values exist per histological group (3 biological replicates \times 3 technical replicates). LFQ intensity values greater than zero were deemed valid. Proteins with at least 6 valid values within a histological group were used in statistical analysis for label-free quantification. This criterion was used to reduce the possibility of analyzing proteins that are nonspecific to histological grades. After log₂-transformation of protein intensities, the missing values were replaced with expected intensities based on the normal distribution (imputation width = 0.3, shift = 1.8) of log₂-transformed LFQ intensities (23). Student's t-test was applied to the preprocessed dataset of matched proteins to detect DEPs across grades of IPMN dysplasia. The comparative pairs for the statistical analysis were LGD versus HGD (comparison 1), HGD versus invasive IPMN (comparison 2), and LGD versus invasive IPMN (comparison 3). A Benjamini-Hochberg FDR threshold of 0.05 was applied to each pair to find significantly changed proteins. Subsequently, the expression patterns of overlapping DEPs across 2 or more pairs were analyzed to screen for biomarker candidates. DEPs that had expression patterns that varied based on the malignancy of IPMN were selected as final biomarker candidates. The resulting DEPs were subjected to hierarchical clustering in Perseus (version 1.5.8.5) with the following parameters: Euclidean distance, average linkage, the number of clusters of 100, maximal number of iterations of 10, the number of restarts of 1, and k-means preprocessing prior to clustering.

8. Bioinformatics analysis

The gene ontologies (GOs) of all DEPs were annotated using the DAVID bioinformatics resource tool (<https://david.ncifcrf.gov/>) and the UniprotKB database (<http://www.uniprot.org/>). The GO analysis included information on biological process (BP), cellular component (CC), and molecular function (MF). Pathway analysis was performed using the KEGG database (<http://www.genome.jp/kegg/>). Secretory protein prediction and functional annotation were performed using SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>), SecretomeP 2.0 (<http://www.cbs.dtu.dk/services/SecretomeP/>), and TMHMM, server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). Ingenuity Pathway Analysis (IPA) was used to conduct functional analysis (Ingenuity Systems, <http://www.ingenuity.com/>). The Plasma Proteome Database (PPD) was used to confirm the association between the proteins that were identified in human plasma and the proteins that were identified in this study (29). The proteins that were identified in our dataset were crossreferenced with mRNA and protein expression in pancreatic sections in the Human Protein Atlas (<http://www.proteinatlas.org/>).

9. Western blot analysis

A total of 19 pancreatic cyst fluid samples—10 LGD, 4 HGD, and 5 invasive IPMN—were used to validate the candidate markers. Equal volumes of a pooled cyst fluid sample were loaded onto each gel to correct for the intensity of the blots. Pancreatic cyst fluid samples were mixed with 5X SDS loading dye (250 mM Tris-Cl, pH 6.8, 10% SDS, 50% glycerol, 0.5 M DTT, 0.1% bromophenol blue). Proteins (40 µg, as measured by BCA assay) were separated on 10% SDS-PAGE gels and

transferred to polyvinylidene fluoride (PVDF) membranes (Hybond-P, GE Healthcare, Pittsburgh, PA). The membranes were stained with Ponceau S dye (P7170, Sigma-Aldrich, USA), blocked with 5% BSA for 2 h at RT, and incubated overnight at 4°C with the following primary antibodies: rabbit monoclonal anti-HOOK1 (ab150397, Abcam, Cambridge, U.K.) at 1:250, mouse monoclonal anti-PTPN6 (sc-7289, Santa Cruz Biotechnology, USA) at 1:1000, and mouse polyclonal anti-SERPINA5 (ab67368, Abcam, Cambridge, U.K.) at 1:100. The membranes were then washed 5 times with Tris-buffered saline and Tween-20 (TBS-T) before being incubated with the following HRP-conjugated secondary antibodies: anti-rabbit (ab6721, Abcam, Cambridge, U.K.) at 1:1000 and anti-mouse (ab6789, Abcam, Cambridge, U.K.) at 1:2500 for 2 h at RT. The membranes were developed with ECL solution (West-Q chemiluminescent substrate Kit-plus, GenDEPOT, Barker, TX), and the signals were visualized on an LAS-4000 (Fujifilm, Japan).

RESULTS

1. Clinical sample characterization

Pancreatic cyst fluid samples from 9 patients were classified into 3 groups: LGD (n=3), HGD (n=3), and invasive IPMN (n=3). The samples did not differ significantly in composition, with the exception of serum CEA level and CA 19-9 concentration measured by chemiluminescent microparticle immunoassay and cyst size (Table 1). The invasive IPMN patient group had the highest average CEA and CA19-9 concentrations at 7.67 ± 7.06 mg/L and 117.17 ± 142.78 mg/L, respectively. CEA and CA19-9 levels were generally higher in the more severe forms of IPMN. The average CEA concentration was approximately 3-fold higher for HGD than LGD subjects and 7-fold higher in invasive IPMN versus LGD. In addition, the average CA19-9 level was approximately 2-fold and 30-fold greater for these comparisons. Our samples were consistent with several publications that have reported that malignant cysts tend to be larger, as evidenced by our invasive IPMN samples (6.63 ± 3.74 cm) being twice as large as LGD (2.93 ± 0.54 cm) and HGD (2.50 ± 0.41 cm) samples on average (30).

Table 1. Demographic and clinical characteristics of the study population.

Group	Pancreatic cyst fluids		
	LGD (N = 3)	HGD (N = 3)	Invasive IPMN (N = 3)
Age (years)			
mean \pm SD	69.00 \pm 1.41	66.33 \pm 8.58	58.33 \pm 11.09
Gender			
Male/Female	1/2	2/1	1/2
Gland Type			
Gastric	2	2	1
Intestinal	0	1	1
Oncocytic	0	0	1
Unknown	1	0	0
Duct Type			
Main	0	0	1
Branch	2	1	0
Mixed	0	2	2
Unknown	1	0	0
Cyst Focality			
Single	2	3	3
Multiple	1	0	0
Mural Nodule			
Y	0	3	3
N	3	0	0
Cyst Location			
Head	1	0	2
Body/Tail	1	3	1
Mixed	1	0	0
CEA Concentration (mg/L)	1.13 \pm 0.53	3.07 \pm 1.27	7.67 \pm 7.06
CA 19-9 Concentration (mg/L)	4.00 \pm 1.48	6.87 \pm 7.17	117.17 \pm 142.78
Cyst Size			
Cyst Size (cm)	2.93 \pm 0.54	2.50 \pm 0.41	6.63 \pm 3.74
< 3.0 cm	1	2	1
\geq 3.0 cm	2	1	2

LGD, low-grade dysplasia; HGD, high-grade dysplasia.

2. In-depth analysis of pancreatic cyst fluid

The overall scheme of the study was based on a proteomic platform of cyst fluids that we established earlier (18). In this study, 9 individual pancreatic cyst fluid samples of various types [LGD (n=3), HGD (n=3), and invasive IPMN (n=3)] were used for label-free quantification. All samples were centrifuged, and only the supernatant was used. The same portions of individual samples were pooled and fractionated to generate a peptide library, which increased the depth of the identified proteins. In contrast, the 9 individual samples were not fractionated. After a series of sample preparation steps, LC-MS/MS analysis was performed on a Q Exactive mass spectrometer. The MS data were processed in MaxQuant (version 1.5.5.1), and the statistical analysis was performed in Perseus (version 1.5.8.5) (Figure 1).

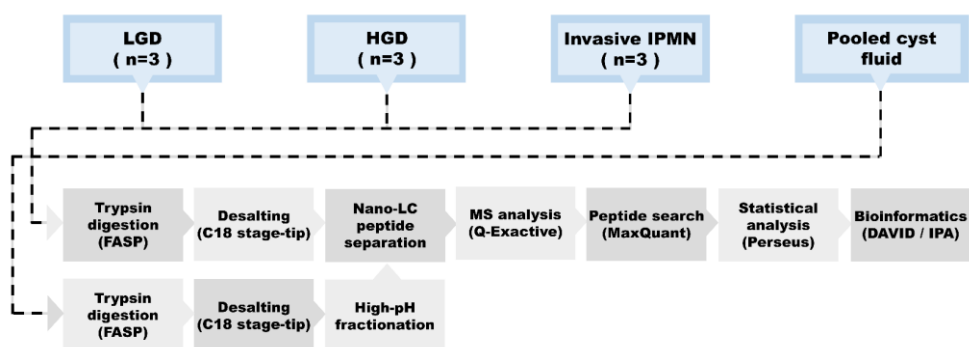


Figure 1. Experimental workflow.

Cyst fluids of three types of dysplasia of IPMN and pooled cyst fluid were used for the analysis. The sample preparation and data analysis are shown in the flowchart.

In total, 2992 proteins were identified and 2938 proteins were quantified. A total of 28,728 peptides were identified, and 553,199 peptide spectra matches were found. In the peptide library and the 9 individual samples, 2778 and 2963 proteins were identified, respectively. Comparing the peptide library with the individual samples, 2749 proteins (91.9% of all identified proteins) were shared (Figure 2A). In the 9 individual samples, most of the identified proteins (95.7%) were usable for quantitative analysis, as evidenced by the 2963 and 2837 proteins that were identified and quantified (Figure 2B). Approximately 2200 to 2500 proteins were quantified in each sample group. The 3 IPMN groups were similar with regard to the number of quantified proteins (Figure 2C). In contrast, there was substantial individual variation in the number of identified and quantified proteins within the same histological subgroups. This pattern was observed across all 9 samples (Figure 2D).

On average, the number of identified and quantified proteins increased by 129 and 83, respectively, in individual samples when matched with the peptide library. In addition, the number of identified peptides rose by 752 on average in individual samples with HGD 1 displaying the greatest improvement of 2109. As shown by the Venn diagram, approximately 77% of identified and 63% of quantified proteins overlapped in all histological groups and 337 additional proteins were identified exclusively when the search was performed with the generated peptide library. Whereas the number of proteins that overlapped in the 3 histological groups decreased by approximately 6% when searched without the peptide library (Figure 2E, F). This result implies that the number of proteins that were common between individual samples rose due to the contribution of the peptide library. The dynamic

range of the proteome spanned over 7 orders of magnitude overall, but most proteins (95%) were expressed within 4 orders (Figure 3). Overall, the proteins with lower orders of magnitude were analyzed, and tumor marker proteins, such as MUC5AC, MUC1, and CEA, were quantified with high intensity in the dataset.

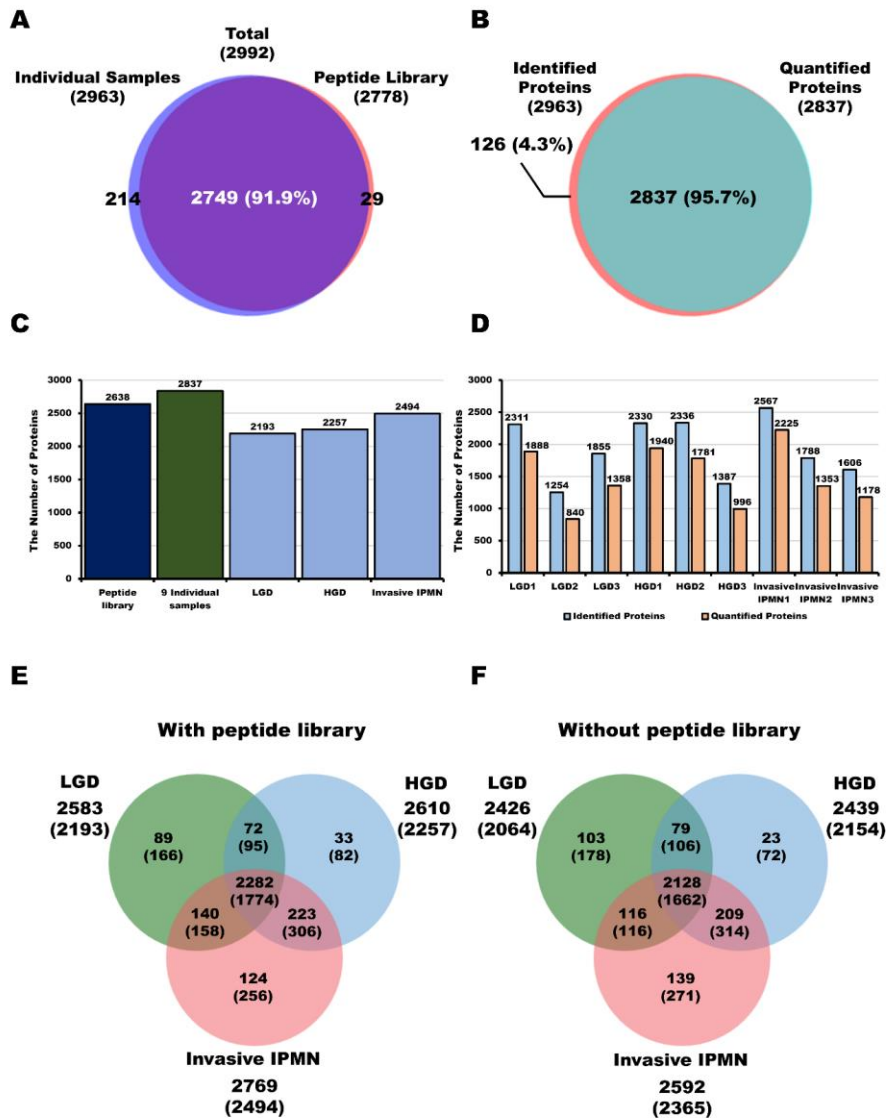


Figure 2. The number of identified and quantified proteins in individual samples and the peptide library.

(A) All identified proteins in the 9 individual samples and peptide library. **(B)** All identified proteins and quantified proteins in the 9 individual samples. **(C)** Quantified proteins in the peptide library, 9 individual samples, and each histological group of IPMN. **(D)** Total identified and quantified proteins in each of the 9 individual samples. Identified and quantified proteins in LGD, HGD, and invasive IPMN in the search results, including the peptide library **(E)** and excluding the peptide library **(F)**. The number of quantified proteins is noted in parentheses.

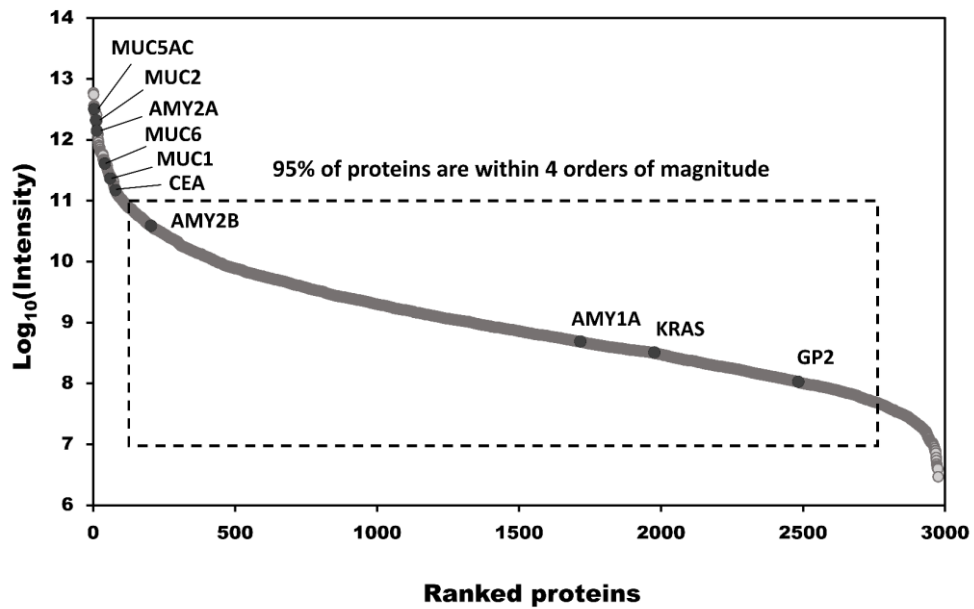


Figure 3. Dynamic range of quantified proteins.

Distribution of expression intensities of quantified proteins show a large dynamic range of abundance, but 95% of the proteins were expressed within 4 orders of magnitude. Several tumor marker proteins, such as MUC2, CEA, and KRAS, were quantified.

3. Comparative analysis using other proteome databases

Our bioinformatics analysis showed that secreted proteins accounted for 60.5% (1810 proteins) of the 2992 identified proteins (Figure 4A). Across SecretomeP, SignalP, and TMHMM, 1527, 682, and 381 proteins were identified, respectively (Figure 4B). Protein accession numbers were converted into gene names, and the resulting redundancy was discarded prior to comparative analysis. We compared our dataset with the Human Plasma Proteome Database to assess the likelihood that the discovered proteins are potential blood markers (29). As a result, 2299 (79.7%) of the identified proteins were plasma proteins (Figure 4C). To determine whether the discovered proteins are expressed in the pancreas, the dataset was crossreferenced with The Human Protein Atlas (<http://www.proteinatlas.org>, May 31, 2017)—2613 genes had corresponding mRNA entries and 2021 genes had corresponding protein entries in the pancreas (Figure 4D).

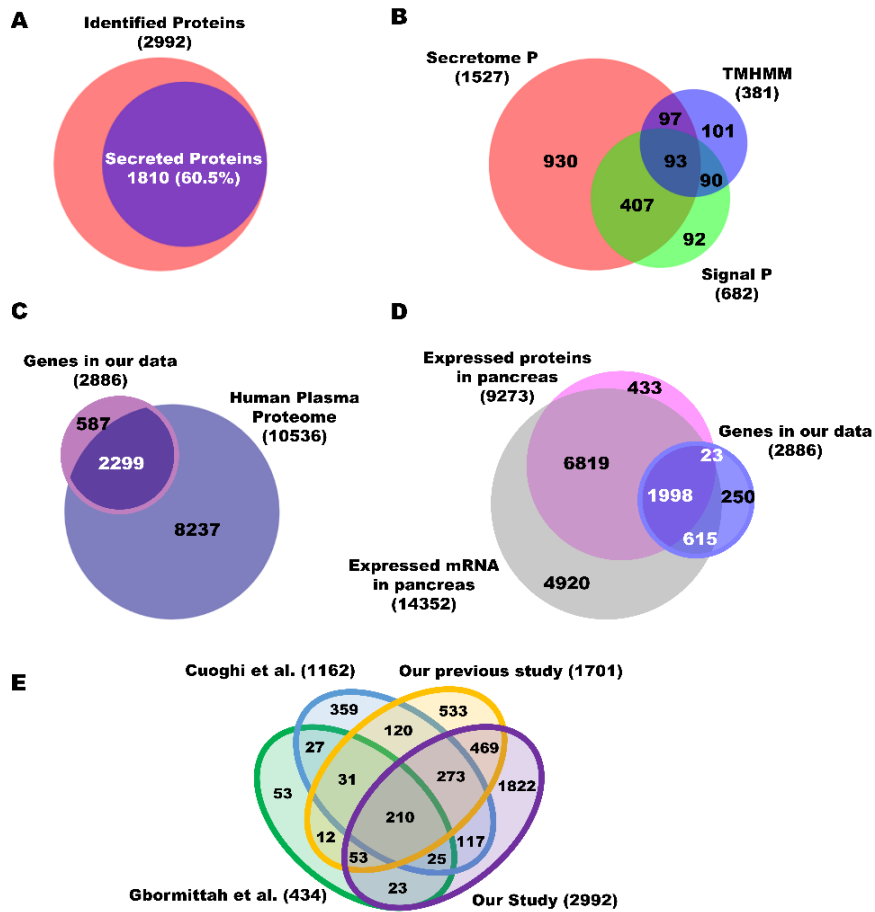


Figure 4. Comparative analysis with other proteomic databases.

(A) A total of 1810 of the identified proteins were secreted proteins; (B) 1810 secreted proteins were annotated in Secretome P, Signal P, and TMHMM. (C) Comparing the dataset with the Human Plasma Proteome Database, 2299 of the identified proteins were plasma proteins. (D) Our data had greater depth than those of other proteomics studies on pancreatic cyst fluid (Cuoghi et al., Gbormittah et al.) and our previous report. (E) Comparing the dataset of 2886 genes to the Human Protein Atlas, 2613 genes had corresponding mRNA entries and 2021 genes had corresponding protein entries in the pancreas.

4. Variation in individual cyst samples

Coefficient of variation (CV%) values were calculated to evaluate the reproducibility of the technical and biological replicates. The CV% values of log₂-transformed LFQ intensity sums of technical triplicates of individual samples ranged from 0.32% to 6.45%. All CV% values of log₂-transformed LFQ intensities of each quantified protein in technical triplicates of individual samples were less than 20%. Moreover, the average CV% value of individual samples ranged from 1.085% to 1.524% (Figure 5A). Pearson correlation coefficients of the LFQ intensities of technical triplicates and their averages were greater than 0.9 (Figure 5B–D). These data suggest that the variation between technical replicates was low. In contrast, the variation between biological triplicates was generally high, based on the Pearson correlation coefficients, which ranged from 0.370 (between LGD1 and LGD2) to 0.789 (between HGD1 and HGD2) (Figure 5B–D).

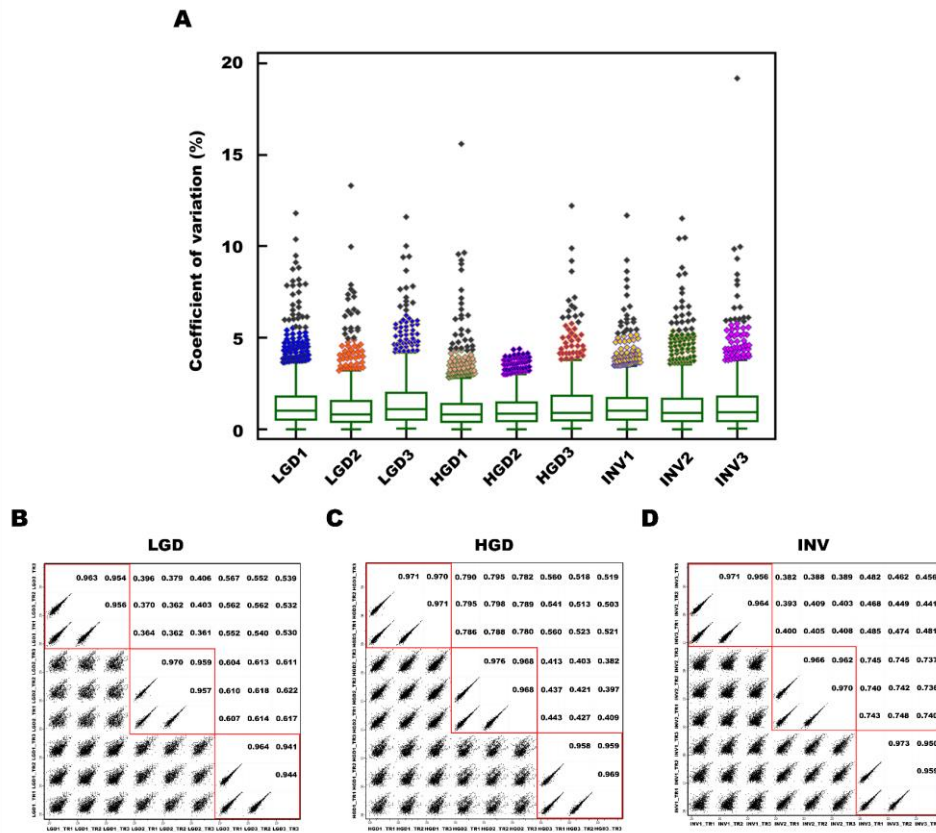


Figure 5. Coefficient of variation (CV%) values of technical replicates in individual samples and scatterplots of Pearson correlation coefficients.

(A) Intensity values of label-free quantification by MaxQuant were transformed to base-2 logarithms, and the coefficient of variation (CV%) values in the technical replicates of the 9 individual cyst fluids were calculated and represented as box plots. Scatterplots of Pearson correlation coefficients of each technical replicate in 3 biological replicates of LGD (B), HGD (C), and invasive IPMN (D). TR, technical replicate.

5. Differentially expressed proteins in IPMN dysplasia

The 1751 proteins that had at least 6 valid values within a histological group were used for the statistical analysis. By student's t-test (Benjamini-Hochberg FDR = 0.05), 149, 48, and 98 proteins were differentially expressed between comparisons 1 (LGD versus HGD), 2 (HGD versus invasive IPMN), and 3 (LGD versus invasive IPMN), respectively, 75, 32, and 64 of which were upregulated (Figure 6A–C). By unsupervised hierarchical clustering, the DEPs clustered by IPMN histology (Figure 6D–F).

There were 243 DEPs across comparisons 1, 2, and 3. Among the 243 DEPs, 142 were upregulated and 91 were downregulated in at least 1 comparison group. Enriched DEPs were used to conduct GO and KEGG pathway analyses to identify their overrepresented terms in biological process, molecular function, and cellular component. The DEPs from comparisons 1 and 3 were analyzed by Ingenuity Pathway Analysis (IPA) bioinformatics tool to track biological processes that became activated or more pronounced in aggressive malignancy.

By GO enrichment analysis, 243 DEPs were involved primarily in vesicle-mediated transport and cell surface receptor signaling with regard to biological process. The analysis also found that 76.6% of DEPs were extracellular region proteins. The molecular functions of the DEPs were primarily associated with peptidase activity and regulation (Figure 7A–C). By KEGG pathway analysis, the 142 upregulated proteins were associated with the ribosome, oxidative phosphorylation, and endocytosis, whereas the 91 downregulated proteins were

linked to leukocyte transendothelial migration, focal adhesion, and ECM-receptor interaction (Figure 7D).

The significantly changed proteins from comparison 1 and 3 were examined by IPA with regard to biological processes that are related to pancreatic cysts and aggressive malignancy. Core analysis in IPA was used to evaluate the biological functions that were most likely to be affected by changes in expression of proteins in our dataset. As a result, 149 DEPs in comparison 1 and 98 DEPs in comparison 3 were associated with such terms as cellular movement and angiogenesis in Diseases and Functions, which are indicative of malignancy; the biological terms that correlated with aggressive malignancy are highlighted in yellow (Figure 8A, B). Cell growth and vasculogenesis were upregulated among the DEPs in comparison 1. A total of 98 DEPs in comparison 3 were upregulated in most Diseases and Functions terms, except for apoptosis of tumor cell lines—particularly metastasis-related terms, such as cell spreading and angiogenesis.

Comparison analysis is usually performed to visualize trends in protein expression across several analyses. Consistent with our expectations, the Diseases and Bio functions heat map of the comparison analysis demonstrated that the DEPs that were associated with cell movement of endothelial cells and angiogenesis were more highly expressed in comparison 3 versus 1. The term “apoptosis of tumor cell lines” was downregulated in comparison 3 but unchanged in comparison 1 (Figure 8C). A higher percentage of DEPs in comparison 3 was associated with pancreas-specific diseases, such as chronic pancreatitis and associated diseases than DEPs in comparison 1 (Figure 8D).

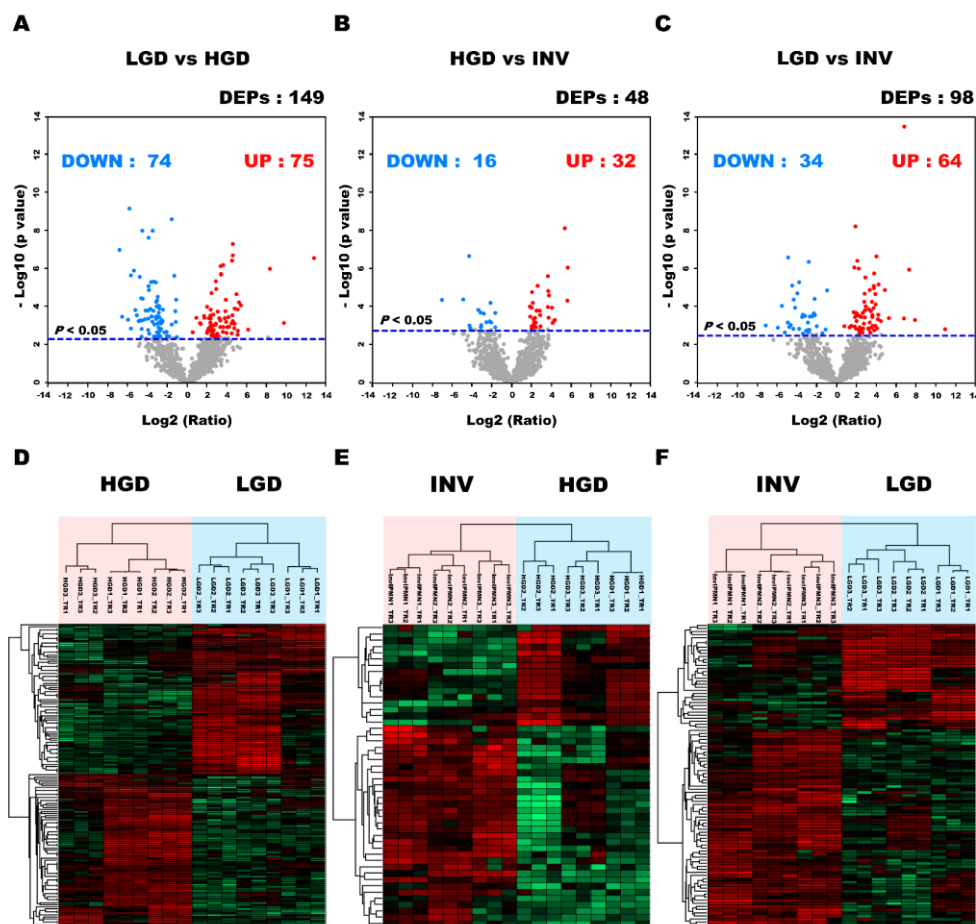


Figure 6. Volcano plots and hierarchical heat map clusters of significant proteins by student's t-test.

The colored dots indicate the proteins that passed the t-test for significance between LGD versus HGD (A), HGD versus invasive IPMN (B), and LGD versus invasive IPMN (C). The blue dots represent downregulated proteins, and the red dots denote upregulated proteins. Hierarchical clustering of differentially expressed proteins after student's t-test of LGD versus HGD (D), HGD versus invasive IPMN (E), and LGD versus invasive IPMN (F). By student's t-test, the DEPs clustered in accordance with the histological groups of IPMN.

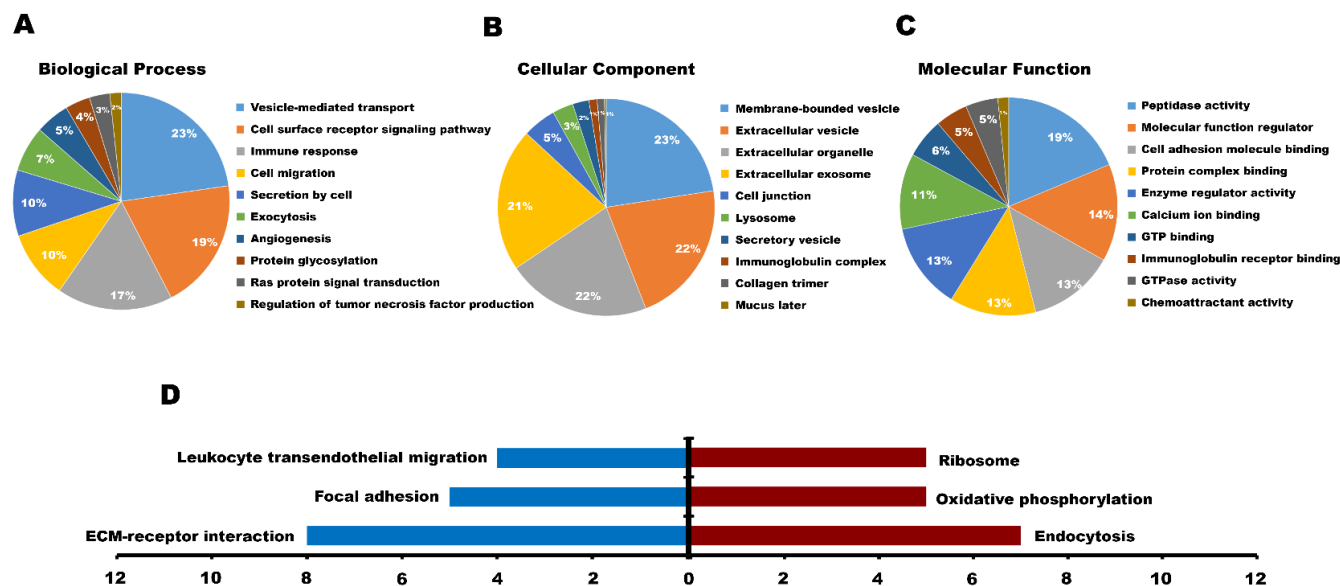


Figure 7. Gene ontology analysis.

The percentage of differentially expressed proteins belonging to (A) Biological process, (B) Cellular component, and (C) Molecular function are shown in pie charts. (D) A total of 142 upregulated proteins in all comparison groups were associated with the ribosome, oxidative phosphorylation, and endocytosis; the 91 downregulated proteins in all comparison groups were associated with leukocyte transendothelial migration, focal adhesion, and ECM-receptor interaction.

Figure 8. Ingenuity Pathway Analysis.

(A-B) The heat maps represent upregulation and downregulation of biological function based on z-score as squares of various sizes and colors. Larger squares reflect greater z-scores, with orange signifying positive values and blue signifying negative values. (A) The biological terms associated with 149 DEPs in LGD and HGD. (B) The biological terms associated with 98 DEPs in LGD and invasive IPMN. The terms related to malignancy are highlighted in yellow. (C) Various terms in Diseases and Functions were represented in accordance with the 2 comparison groups. Orange represents a positive z-score, and blue represents a negative z-score. (D) The p-values, which represent the correlation between DEPs and pancreatic diseases, are shown in accordance with the 2 comparison groups. Darker colors reflect greater association with the disease.

6. Biomarker candidates for IPMN malignancy

Proteins that were shared by at least 2 comparison groups were chosen as the initial marker candidates. Our rationale was that significantly changed proteins that are common to several comparison groups are more likely to be associated with the malignancy of IPMNs (31). A total of 49 candidates expressed in at least 2 comparison groups were selected from 243 DEPs. Then, the DEPs that had expression patterns that were consistent with the degree of IPMN malignancy were selected as the final candidates. Table 2 details the results of the statistical analysis of the 49 DEPs, including the p-value, fold-change, and t-test significance for each comparison group. Of the 49 DEPs, 38 were secreted proteins and 33 were confirmed to be expressed in the pancreas as mRNA or proteins in The Human Protein Atlas. In addition, 35 proteins were confirmed to be expressed in plasma, according to the Human Plasma Proteome Database (Table 2).

Of the 49 shared DEPs between groups, 18 had expression patterns that were consistent with the degree of malignancy. PTPN6, MUC2, TLN1, and YBX1 were expressed in lower amounts in LGD but gradually elevated in HGD and invasive IPMN. Conversely, SERPINA5, AKR1B10, and TFF1 expression decreased as IPMN histological grade progressed. Other proteins, such as HOOK1, TYMP, TEX12, FBN1, CLDN18, THY1, MUC5AC, CST6, WFDC2, PIK3IP1, and SERPINA4, were predominantly expressed in LGD or invasive IPMN but not in other groups (Figure 9). Based on these results, these 18 proteins were selected as potential biomarkers of IPMN dysplasia.

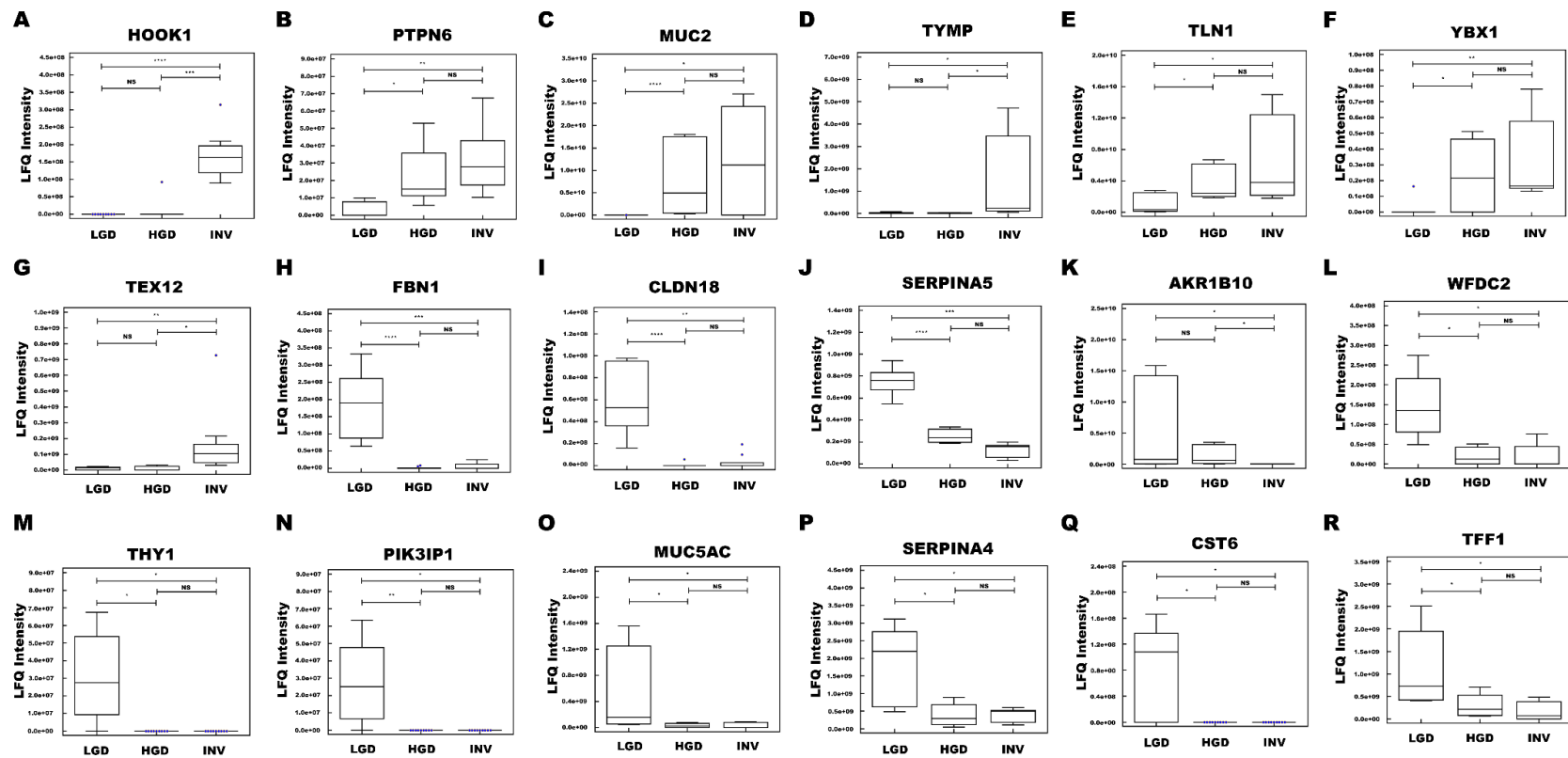
7. Validation by western blot

Three DEPs (HOOK1, PTPN6, and SERPINA5) were validated by western blot using 19 pancreatic cyst fluid samples (10 LGD, 4 HGD, and 5 invasive IPMN) (Figure 10). The results were then compared with the MS analysis findings. Although not every western blot sample followed the trend in the MS analysis, the expression patterns of HOOK1 and PTPN6 generally correlated with the LFQ intensity values. HOOK1 was significantly upregulated in high-risk IPMN ($P < 0.01$), and PTPN6 was detected at higher levels in high-risk IPMN ($P < 0.05$). The expression pattern of SERPINA5 was not consistent with the MS analysis and was higher in high-risk IPMN.

Table 2. Detailed statistical analysis of 18 final marker candidates.

LGD vs HGD				HGD vs INV				LGD vs INV				SignalP	SecretomeP	TMHMM	The Human Protein Atlas		Human plasma proteome
Gene name	<i>T</i> test Significance	adjusted <i>P</i> value	Log ₂ (Fold Change)	<i>T</i> test Significance	adjusted <i>P</i> value	Log ₂ (Fold Change)	<i>T</i> test Significance	adjusted <i>P</i> value	Log ₂ (Fold Change)	Passed	Passed	Passed	Passed	Passed	RNA expression	Protein expression	Included
Up	TYMP	0.890	-0.327	+	0.008	5.612	+	0.013	5.285	-	Y	-	Y	-	Y	-	Y
	TEX12	0.756	-0.484	+	0.012	4.061	+	0.004	3.577	-	Y	-	-	-	-	-	Y
	HOOK1	0.336	1.194	+	0.001	5.655	+	0.000	6.839	-	-	-	Y	-	Y	-	Y
	PTPN6	+	0.024	2.133	0.501	0.664	+	0.004	2.798	-	-	-	Y	-	Y	-	Y
	YBX1	+	0.042	2.626	0.621	0.738	+	0.001	3.365	-	Y	-	Y	Y	Y	Y	Y
	MUC2	+	0.000	12.787	0.768	-1.824	+	0.033	10.964	Y	-	-	-	-	-	-	Y
	TLN1	+	0.037	3.014	0.601	0.545	+	0.030	3.559	-	-	-	Y	Y	Y	Y	Y
Down	AKR1B10	0.979	-0.105	+	0.008	-7.034	+	0.027	-7.139	-	-	-	-	-	-	-	-
	PIK3IP1	+	0.007	-2.889	0.000	0.000	+	0.020	-2.877	Y	-	-	Y	-	Y	-	Y
	THY1	+	0.010	-3.068	0.000	0.000	+	0.012	-3.390	Y	-	-	Y	-	Y	-	Y
	TFF1	+	0.031	-2.095	0.324	-2.247	+	0.032	-4.342	Y	Y	-	Y	-	Y	-	Y
	SERPINA5	+	0.000	-1.585	0.070	-1.222	+	0.000	-2.807	Y	Y	-	Y	-	Y	-	Y
	SERPINA4	+	0.015	-2.655	0.687	0.502	+	0.012	-2.154	Y	Y	-	Y	Y	Y	Y	Y
	FBN1	+	0.000	-5.833	0.437	0.951	+	0.000	-4.882	Y	-	-	Y	-	Y	-	Y
	CLDN18	+	0.000	-3.901	0.675	-0.571	+	0.001	-4.472	-	-	Y	Y	-	Y	-	-
	MUC5AC	+	0.027	-1.608	0.697	-0.012	+	0.020	-1.620	Y	-	-	-	-	-	-	-
	WFDC2	+	0.006	-3.726	0.830	-0.462	+	0.007	-4.188	Y	Y	Y	Y	Y	Y	-	Y
	CST6	+	0.012	-4.284	0.000	0.000	+	0.032	-3.778	Y	Y	-	-	-	-	-	Y

LGD, low-grade dysplasia; HGD, high-grade dysplasia; TMHMM, Transmembrane Helices.



* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, NS not significant

Figure 9. 18 biomarker candidates that had expression patterns that were consistent with the degree of IPMN malignancy.

HOOK1 (A), PTPN6 (B), MUC2 (C), TYMP (D), TLN1 (E), YBX1 (F), and TEX12 (G) were predominantly expressed in invasive IPMN. FBN1 (H), CLDN18 (I), SERPINA5 (J), AKR1B10 (K), WFDC2 (L), THY1 (M), PIK3IP1 (N), MUC5AC (O), SERPINA4 (P), CST (Q), and TFF1 (R) were primarily expressed in LGD. *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$, ****, $P < 0.0001$, NS: not significant.

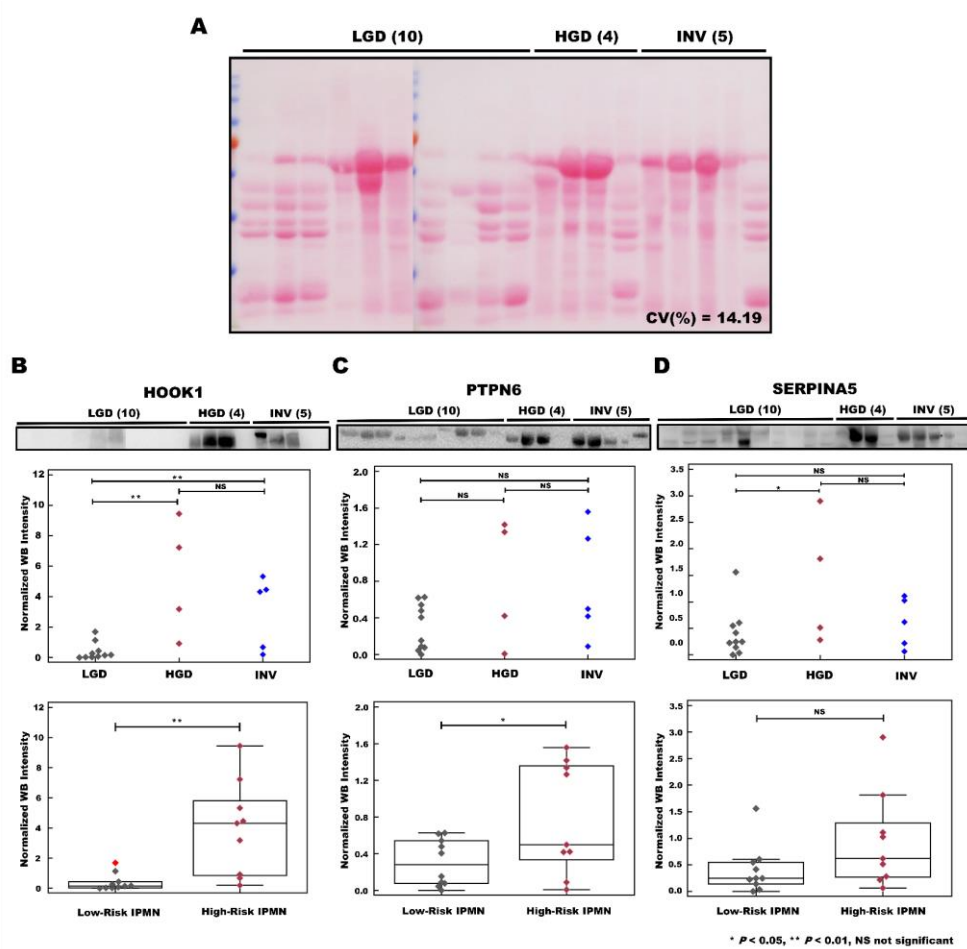


Figure 10. Validation of HOOK1, PTPN6, and SERPINA5 as potential biomarker targets by western blot.

(A) Ponceau S staining as an alternative loading control. Western blot band and scatter dot plots of HOOK1 (B), PTPN6 (C), and SERPINA5 (D).

DISCUSSION

Most pancreatic neoplasms, which are predominantly IPMN, are discovered incidentally during routine check-ups (3). Nevertheless, the lack of a standardized guideline adds subjectivity and undesired variability in diagnosing the malignancy of IPMN lesions. Because the concentrations of tumor biomarkers are higher in cyst fluid than in blood, pancreatic cyst fluid of IPMN patients was analyzed to discover biomarker candidates that could address these inconsistencies in diagnosing IPMN malignancy (12, 13). Thus, analyzing proteins that vary significantly, depending on the malignancy of IPMN, can identify biomarkers that improve the diagnostic performance of current methods and decrease the number of patients who undergo unnecessary operations (14).

As shown in our results, we generated a pancreatic cyst fluid proteome that comprised 2992 proteins (Figure 2A). Our proteome had 3 and 7 times the number of proteins versus studies by Cuoghi (16) and Gbormittah (17), respectively. Further, 1291 additional proteins were identified over our previous study (18) by optimizing the standard proteomic profiling platform by constructing a peptide library of a pooled sample, methodically preparing samples, and reproducibly performing label-free quantitative analysis in triplicates (Figure 4E). Normally, DDA acquisition cannot detect low-abundance proteins in individual samples, because high-abundance proteins saturate the signal. By pooling and fractionating individual samples, these low-abundance proteins became distinct and detectable, as evidenced

by a dynamic range that spanned 7 orders of magnitude (Figure 3). Consequently, the number of identified and quantified proteins that were common to all individual samples rose substantially when the mass spectra of individual samples were matched to those of the peptide library (Figure 2E, F) (32, 33). This increase enabled us to select biomarker candidates from a larger pool of DEPs.

Most identified proteins (79.7%) that had entries in the Plasma Proteome Database and all marker candidates in our study, except AKR1B10, CLDN18, and MUC5AC, were confirmed to be expressed in plasma (Figure 4C). This result suggests that the discovered candidates are potential blood marker candidates. Taking into account that 70.0% of proteins were expressed in the pancreas, according to The Human Protein Atlas, it is probable that the biomarker candidates are specific to the pancreas (Figure 4D). Considering the bioinformatics analysis results of secreted proteins, we conclude that secreted proteins that originate from pancreatic epithelial cells constitute a significant portion of cyst fluid (Figure 4A). The high percentage of matches in these comparative analyses confirms that virtually all of the debris was discarded and that only cyst fluid was collected during sample preparation.

The high Pearson correlation coefficients (>0.9) that were obtained from the pairwise correlation analysis of LFQ intensity values indicated a strong correlation between technical triplicates and that the MS data were acquired without bias (Figure 5B–D). In contrast, the Pearson correlation coefficient of the biological replicates of the histology groups was low, as shown in Figure 5B–D, primarily due to the wide variety of cyst types, the variations in cyst size, and the presence of blood

contaminants (34). One possible source of variation is the contamination of cyst fluid by blood. Fortunately, the samples in this experiment were relatively clean, as evidenced by the inability to detect albumin and low (intensity rank 1475) IgG levels (Figure 3). Despite using relatively clean cyst fluid, the variation between individual cyst fluid samples remained large (Figure 2D). Based on this result, we infer that using contaminated samples will result in even greater individual variation.

Selecting proteins that had at least 6 valid values within a histological group mitigated the likelihood of analyzing proteins that are not representative of their histology group, as evidenced from the low p-value of the t-test, the high fold-change value, and the clear division between clusters shown in the heat map (Figure 6D–F). After eliminating DEPs that were unique to single comparison group, 18 proteins that changed expression levels in accordance with the degree of IPMN malignancy were selected as biomarker candidates (Table 2, Figure 9). Overall, our stringent criteria—requiring at least 6 valid values in a histological group, rigorous statistical analysis parameters, and a consistent expression pattern across histology groups—significantly increased the probability of finding more credible biomarker candidates.

All 18 biomarker candidates were associated with pancreatic disease and malignancy. With the exception of HOOK1, TEX12, TLN1, and PIK3IP1, all candidates are expressed in pancreatic tissue. Twelve candidates were associated with pancreatic diseases, such as IPMN, pancreatic ductal adenocarcinoma (PDAC), and pancreatitis (35–46). According to Tanaka, CLDN18 is an early-stage marker of PDAC and is elevated in high-grade versus low-grade lesions, consistent with our data (2). Two types of mucin proteins were selected as biomarker candidates and

have been examined in studies on IPMN and pancreatic cancer. Our protein expression patterns were consistent with those of prior studies. One of the 2 mucin biomarkers, MUC5AC, is expressed at high levels during the early stages of pancreatic ductal dysplasia but is low in high-grade tumors (38). MUC2 is expressed in IPMNs but not normal pancreatic tissue or PDAC (47).

PTPN6, YBX1, TYMP, CLDN18, WFDC2, SERPINA4, TFF1, MUC2, MUC5AC, CST6, THY1, and AKR1B10 overexpressed in PDAC and pancreatitis. PTPN6 has not been reported in human pancreatic samples but has been observed in PANC-1 cell lines and a rat model of pancreatitis (48). The upregulated proteins, YBX1 and TYMP, are expressed at higher levels in PDAC versus normal tissue, a pattern that is consistent with our proteomic data (43, 44). In addition, these candidates are overexpressed in other types of cancer, such as breast and bladder cancer (49, 50). The remaining 6 candidates, except SERPINA4 and MUC2, are overexpressed in PDAC (35-37, 41, 45, 51, 52). These proteins are involved in tumor progression and differentiation. Accordingly, they are regarded as marker candidates of various cancer types. WFDC2 is a potential early diagnostic marker of gynecological cancers, such as ovarian and endometrial cancer (53). Moreover, serum levels of WFDC2 are indicative of ovarian cancer (54). TFF1, THY1, and AKR1B10 are associated with various cancers and have been implicated as biomarker candidates (55-57). Although it is unknown whether SERPINA4 mediates the progression of pancreatic cancer, it is an early marker of severity in acute pancreatitis (58). These studies have demonstrated that our final list comprises bona fide candidate markers for IPMN. Our report has significance as the first study to

discover the following marker candidates of IPMN: HOOK1, TEX12, TLN1, SERPINA5, FBN1, and PIK3IP1. With the exception of TEX12, these proteins are associated with other cancers, such as hepatocellular carcinoma, breast cancer, and prostate cancer (59-64). Considering the literature regarding the 18 candidates, it is likely that they are related to IPMN malignancy, except for TEX12.

In order to confirm the validity of the aforementioned marker candidates, we compared our MS analysis results with western blot results. Western blot with cyst fluids is difficult due to the lack of housekeeping proteins, such as alpha-tubulin and beta-actin. To address this issue, we used 0.1% Ponceau S solution as a loading control (Figure 10A) (65, 66). The CV% of the intensities of individual samples was 14.19%, indicating that approximately equal amounts had been loaded onto the SDS-PAGE gels. Three DEPs were selected for further validation: 2 upregulated (HOOK1 and PTPN6) and 1 downregulated protein (SERPINA5). The selection criteria for validation were a low p-value, high LFQ intensities, and a lack of an association with IPMN in the literature (which suggests novelty).

HOOK1 was highly expressed in HGD and invasive IPMN compared with LGD ($P < 0.01$). Although the difference in PTPN6 was not statistically significant between the 3 comparison groups, its overall expression pattern underwent similar changes as in the MS results (Figure 10B, C). The expression pattern of SERPINA5 was not consistent with the MS analysis and was higher in high-risk IPMN (Figure 10D). This inconsistency might have resulted from the inherent property of western blots, which depends on the affinity between an antibody and a single antigenic epitope (67, 68). Thus, if the antibody has weak affinity for the epitope, the western

blot results would not be an accurate measure of protein abundance. In this regard, although western blot has been the standard assay in proteomics, targeted proteomic analysis might be a better alternative for verifying our quantitative MS data.

In summary, we have identified 2992 proteins in IPMN cyst fluid samples using mass spectrometry techniques. Our investigation demonstrates that the use of a peptide library is beneficial, because the increased number of identified proteins provides a wider selection to choose from as biomarkers. This is evident from our dataset, which contains the largest number of proteins for pancreatic cyst fluid. Our in-depth data on the pancreatic cyst fluid proteome will be a valuable resource for pancreatic disease research.

Our bioinformatics analysis concluded that upregulated DEPs were associated with pancreatic cell proliferation and aggressive malignancy. Through statistical analysis, we designated 18 biomarker candidates that changed expression levels, depending on the histological grade of IPMN. Among them, 2 upregulated DEPs were consistent with our western blot analysis. The literature has also concluded that these proteins are involved primarily in pancreatic diseases and malignancy, rendering them promising biomarker candidates of IPMN malignancy. In future studies, we plan to collect a sufficient amount of cyst fluid samples from more patients to test the performance of these biomarkers by immunoassay and multiple reaction monitoring (MRM).

CHAPTER II

Marker Identification of the Grade of Dysplasia of Intraductal Papillary Mucinous Neoplasm in Pancreatic Cyst Fluid by Quantitative Proteomic Profiling

INTRODUCTION

The incidental detection of pancreatic cystic lesions (PCLs) has increased in recent years due to the implementation of various screening methods and the advancement of medical imaging technologies, such as magnetic resonance imaging (MRI), computed tomography (CT), and endoscopic ultrasound (EUS) (69-71). In response, many studies have attempted to develop screening methods that aid in the therapeutic decision-making with regard to PCLs, including intraductal papillary mucinous neoplasm (IPMN), which has been detected most frequently as a precursor lesion of pancreatic cancer (71).

IPMN stages vary significantly as the malignancy progresses from benign to malignant—low-grade dysplasia (LGD), high-grade dysplasia (HGD), and invasive IPMN (72). LGD is considered primarily to be amenable to active surveillance, whereas the lesions in HGD and invasive IPMN require surgical intervention (2), necessitating the accurate classification of cystic lesions for appropriate patient management. Currently, 3 guidelines are used widely for establishing the treatment strategy for IPMN patients (73-75). However, the standard for determining whether to conduct active surveillance or surgical intervention and the diagnostic accuracy in determining IPMN grade differ between guidelines. Thus, the same patient can be treated differently, depending on which guideline is followed by the clinician. Consequently, establishing a treatment strategy that is based solely on these guidelines is problematic in actual clinical practice (76).

To examine the discrepancy between these guidelines, nomograms have been developed to predict low-risk and high-risk IPMNs (77, 78). However, the nomogram-derived objective risk score has limited diagnostic accuracy. The following factors are used to calculate the risk score: (1) Abdominal imaging, such as MRI, CT, and EUS; (2) carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) levels; (3) analysis of KRAS and GNAS mutations; and (4) cyst fluid cytology. MRI and CT scans are inconsistent in differentiating between cyst types, as evidenced by their wide range in diagnostic accuracy (20% to 80%) (79). EUS also suffers from poor accuracy (51% to 59%) and cannot distinguish benign cystic lesions from malignant cysts (80). The most extensively studied biomarker, CEA, has low accuracy (60% to 80%) when used to discriminate between mucinous and nonmucinous cystic lesions (6). Another tumor marker, CA19-9, has a specificity of 81%, which is offset by its low sensitivity (15%) in differentiating mucinous and nonmucinous cystic lesions (10). Analysis of KRAS mutations has 100% specificity but is not sufficiently sensitive (50%) to determine IPMN dysplasia (81). Similarly, although analyzing GNAS mutations is adequate for distinguishing IPMNs from other types of pancreatic cysts, they cannot predict the grade of dysplasia, because they generally occur early in IPMN development (5). Cyst fluid cytology for mucinous cysts has low diagnostic accuracy (54% sensitivity and 93% specificity) (82). Further, cytological examination is only applicable to cases in which there is a sufficient concentration of observable cells in the cyst fluid (83). Thus, it is likely that patients have undergone unnecessary surgical interventions due to the absence of an accurate method for determining the malignancy of IPMN (3),

necessitating novel biomarkers that improve the accuracy of the diagnosis of IPMN.

Pancreatic cyst fluid has several advantages over serum and plasma with regard to the discovery of markers for IPMN (84, 85). Because cyst fluid is composed of secreted proteins from surrounding tumor cells, the concentration of potential biomarkers in cyst fluid is approximately 1000 times higher than in blood. In addition, these candidates closely represent changes in the dysplastic epithelium (86), and cyst fluid can be collected by endoscopic ultrasound-guided fine needle aspiration (EUS-FNA), a safe and minimally invasive method (87).

Mass spectrometry (MS)-based proteomic approaches are being increasingly applied to identify markers that are related to specific diseases, based on their ability to screen thousands of proteins simultaneously to obtain hundreds of differentially expressed proteins (DEPs) in small amounts of samples (88). For instance, a study by Jabbar et al. concluded that quantifying CEA by a conventional method requires 1000 times more cyst fluid than MS analysis (89). Thus, an MS-based approach is the most suitable platform for screening biomarkers in cyst fluid.

Existing diagnostic modalities (CT, MRI, EUS, CEA, and CA19-9 levels; KRAS and GNAS mutations; and cyst fluid cytology) are insufficient for accurately classifying IPMN patients due to their low diagnostic accuracy (5, 6, 10, 79-82). Consequently, many reports, including proteomic studies, have examined methods of discovering biomarkers to improve the diagnostic accuracy for PCLs. In our previous study, we identified potential biomarkers of the histological grades of IPMNs using cyst fluid that was obtained exclusively from IPMN patients by LC-MS/MS (90). In the current study, we aimed to discover marker candidates for IPMN

dysplasia from an expanded cohort that included IPMNs and other PCLs (mucinous cystic neoplasm (MCN) and serous cystic neoplasm (SCN)) by mass spectrometry to better reflect actual clinical circumstances to help classify various PCLs and avoid unnecessary pancreatic resection for low-risk IPMN patients.

MATERIALS AND METHODS

1. Patients and cyst fluid samples

Cystic fluid samples were collected from 30 patient specimens (20 IPMN, 5 MCN, and 5 SCN) immediately after pancreatectomy at Seoul National University Hospital between April 2013 and June 2017. IPMN samples were classified as low-grade dysplasia (LGD, n = 10), high-grade dysplasia (HGD, n = 5), and invasive IPMN (n = 5). The same samples were also categorized as low-risk IPMN (LGD) and high risk-IPMN (HGD and invasive IPMN). The patient data and characteristics of the cystic lesions are summarized in Table 1. At least 200 μ L of cyst fluid was aspirated from patients to acquire sufficient protein for analysis. The aspirated cyst fluid samples were stored at -80°C until sample preparation. All contents of this research were approved by the Institutional Review Board (IRB No. 1304-121-486), and all participants provided written informed consent.

Table 1. Demographic and clinical characteristics of the study population.

Group	Pancreatic cyst fluids				
	LGD	HGD	Invasive IPMN	MCN	SCN
	(N = 10)	(N = 5)	(N = 5)	(N = 5)	(N = 5)
Age (years)					
Mean \pm SD	65.80 \pm 5.55	67.80 \pm 9.88	50.80 \pm 14.45	49.00 \pm 11.60	51.60 \pm 17.08
Gender					
Male/Female	5/5	4/1	3/2	1/4	1/4
Gland Type					
Gastric	9	2 (1)	2 (1)		
Intestinal	0	1 (1)	1		
Oncocytic	0	0	1		
Pancreatobiliary	0	1	0		
Pancreatic	0	0	(1)		
Unknown	1	0	0		
Duct Type					
Main	0	0	1		
Branch	4	4	1		
Mixed	4	1	3		
Unknown	2	0	0		
Cyst Focality					
Single	8	5	4	2	0
Multiple	2	0	1	0	0
Unknown	0	0	0	3	5
Mural Nodule					
Y	2	3	5	0	0
N	8	2	0	2	0
Unknown	0	0	0	3	5
Cyst Location					
Head	4	3	3	0	1
Body/Tail	6	2	2	5	4
CEA Concentration (mg/L)	1.44 \pm 0.74	1.52 \pm 1.04	5.48 \pm 6.82	1.32 \pm 0.96	1.44 \pm 0.48
CA 19-9 Concentration (mg/L)	11.86 \pm 8.69	22.80 \pm 29.77	90.28 \pm 129.71	20.20 \pm 30.35	19.76 \pm 17.81
Cyst Size	3.36 \pm 1.33	3.56 \pm 1.66	5.74 \pm 3.69	7.50 \pm 2.18	3.98 \pm 1.93
<3.0 cm	4	3	2	0	1
\geq 3.0 cm	6	2	3	5	4

LGD, low-grade dysplasia; HGD, high-grade dysplasia; MCN, mucinous cystic neoplasm; SCN, serous cystic neoplasm; * The number of patients with two different gland types is shown in parentheses.

2. Protein isolation from cell pellets and secreted protein

PANC1, Mia Paca-2, and BxPC3 cells pellets were rinsed with cold PBS and homogenized by sonication for 30 s (Sonics & Materials Inc., Newtown, CT, USA) in lysis buffer (4% SDS, 1 mM TECP, 0.1 M Tris-Cl, pH 7.4). The samples were boiled in a water bath for 30 min at 100°C and then centrifuged (15,000 rpm, 20°C, 20 min) to remove cell debris. Next, 100 µg of proteins from the 3 cell types were pooled for further processing. The pooled cell lysate and 300 µg of each secreted protein were precipitated with cold acetone (Sigma-Aldrich, St. Louis, MO, USA) at a ratio of 1:5 (sample:acetone, v/v). The mixture was incubated overnight at -20°C after being vortexed thoroughly. The precipitate was centrifuged for 10 minutes at 15,000 rpm at 4°C, and the acetone was gently removed. After an additional rinse step with 500 µL cold acetone and centrifugation, the protein pellet was air-dried for 2 h and stored for digestion.

3. Pancreatic cyst fluid sample preparation

In cases in which the cyst fluid was too viscous to be extracted with a pipette, the cyst fluid was sonicated briefly prior to mucus removal in a 1.5-mL Eppendorf tube (90). The samples were centrifuged (15,000 rpm, 20 min, 4°C) to separate the supernatant from the cellular debris and other solid contents. Only the supernatant was used in this study. The protein concentration was measured by BCA assay, and 20 µg of proteins from each sample were pooled for further fractionation. One hundred fifty micrograms of proteins from each sample were precipitated with cold acetone in the same manner as the cell pellets.

4. Protein digestion and desalting

The acetone-precipitated samples (1 pooled cyst fluid, 30 individual cyst fluids, 1 pooled cell lysate, secreted proteins of 3 cell lines) were mixed with 30 μ L SDT lysis buffer (4 % SDS, 0.1 M DTT, 0.1 M Tris-Cl, pH 7.4). After being vortexed gently, the mixture was boiled for 30 min at 100°C to denature the proteins. The denatured samples were mixed with 300 μ L 0.22- μ m-pore filtered UA buffer (8 M urea, 0.1 M Tris-Cl, pH 8.5) and then transferred to a 30-kDa centrifugal filter (Millipore, Billerica, MA, USA). The sample was then centrifuged 3 times (14,000 g, 15 min, 20°C) to remove SDS. The washed samples were incubated in 200 μ L 50 mM iodoacetamide (IAA) in UA buffer at room temperature (RT) for 1 h to alkylate the reduced cysteine. After exchanging the UA buffer with 40 mM ammonium bicarbonate (ABC), the samples were digested with 0.1 μ g/ μ L trypsin at a ratio of 1:50 (enzyme:substrate, wt/wt) for 18 h at 37°C.

The pooled samples were subjected to a second digestion step (90). After being transferred to a new centrifuge tube, the filters were washed sequentially with 200 μ L UA buffer once and 300 μ L 40 mM ABC twice. Then, the proteins were digested with 0.1 μ g/ μ L trypsin (trypsin:sample ratio of 1:100, wt/wt). The 30 individual cyst fluid digests were measured by tryptophan fluorescence assay to estimate the amounts of peptides. Equal amounts of peptides were acidified and desalted with homemade StageTips as described (21, 22). In contrast to the individual samples, all pooled samples were desalted without conserving any spare volume. The desalted samples were then lyophilized to dryness in a speed-vacuum centrifuge and stored at -80°C until fractionation and analysis.

5. High-pH reverse-phase peptide fractionation

A total of 5 library samples (pooled cyst fluid; pooled cell lysate; and the secreted proteins of PANC-1, Mia PaCa-2, and BxPC3 cells) were further fractionated on a modified Stage-tip column in a high-pH environment to generate a peptide library (23). The desalted peptides from the first digest were reconstituted in 200 μ L loading solution (15 mM ammonium hydroxide, pH 10, 2% acetonitrile) and separated on a pipette-based RP microcolumn, prepared by plugging the bottom of a 200- μ L pipette tip with C18 Empore disk membrane (3M, St. Paul, MN, USA) and filling the tip with POROS 20 R2 resin. After 3 rinses each of 100% methanol, 100% acetonitrile (ACN), and loading buffer, the dissolved peptides were loaded onto the column and eluted into 20 fractions on a discontinuous ACN gradient (2, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5, 30, 32.5, 35, 40, 50, 60, 70, 80, and 100%). These 20 fractions were concatenated into 6 fractions with varying hydrophobicities to optimize coverage and liquid chromatography-tandem mass spectrometry (LC-MS/MS) run time. The 6 fractions were dried in a vacuum centrifuge and stored at -80°C until LC-MS/MS analysis.

6. LC-MS/MS analysis

The peptides were analyzed by a Q Exactive mass spectrometer that was equipped with an EASY-Spray ion source (Thermo Fisher Scientific, Waltham, MA, USA), coupled to an Easy-nano LC 1000 (Thermo Fisher Scientific, Waltham, MA, USA), following our established protocol (90). The peptide samples were separated on a 2-column setup that comprised a trap column (75 μ m I.D. x 2 cm, C18 3.0 μ m, 100 Å)

and an analytical column (Easy-Spray Pepmap RSLC, 75 μm I.D. x 50 cm, C18 2.0 μm , 100 Å).

An ACN gradient (6% to 40%) run of 180 minutes was achieved by mixing solvent A (2% ACN and 0.1% v/v formic acid) and solvent B (100% acetonitrile and 0.1% v/v formic acid) in varying proportions. Peptides that were eluted from the analytical column were ionized at a spray voltage of 2.0 kV in positive ion mode. MS1 spectra were collected in data-dependent acquisition (DDA) mode using a top 15 method with a resolution of 70,000 at m/z 200 with a mass range of 350-1700 m/z . The 15 most abundant ions were fragmented by higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 27 at a resolution of 17,500 at m/z 200. The maximum ion injection times for the survey and MS/MS scans were 20 ms and 80 ms, respectively. The dynamic exclusion was set to 30 s to prevent repeated sequencing.

7. Raw data search

All raw MS files (120 files) from the Q Exactive were processed in MaxQuant, version 1.6.0.16 (24) with the built-in Andromeda search engine (91) against the Uniprot human database (88,717 entries, version from December 2014), containing the forward and reverse amino acid sequences. In accordance with the established target-decoy search procedures (25), the search results were filtered at a false discovery rate (FDR) < 1% for identifying peptides, modification sites, and proteins.

The database search was performed with the following parameters: digestion mode trypsin/P; main search and first search tolerances of 6 ppm and 20

ppm, respectively; tolerance of up to 2 missed cleavages; carbamidomethylation of cysteine as a fixed modification; oxidation of methionine and protein N-terminal acetylation as variable modifications; and peptide length of at least 6 residues. Peptides were assigned to protein groups by the principle of parsimony (26). The principle was applied to reduce the number of ambiguous proteins, the identified peptides of which could belong to several proteins. The retention times of all raw files were aligned through the “match between runs” feature in MaxQuant, which allows the transfer of MS/MS spectra and sequence information within a retention time window of 0.7 min to other raw files that have insufficient MS/MS spectra to identify the sequences (92).

All generated proteomic data have been submitted to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org/>) via the PRIDE partner repository, with PXD016127 as the identifier (27).

8. Label-free quantification and statistical analysis

Label-free quantification (LFQ) and statistical analysis were performed in MaxQuant (version 1.6.0.16) and Perseus (version 1.6.1.1), respectively, in accordance with our previous studies (90). Normalized spectral protein intensity values (LFQ intensity values) were used to estimate the protein abundance (28). LGD had 30 LFQ intensity values (10 biological replicates \times 3 technical replicates), whereas the others had 15 LFQ intensities (5 biological replicates \times 3 technical replicates). LFQ intensity values that were greater than 0 were deemed to be valid.

From the list of identified peptides, proteins that had 70% or more valid

values in at least 1 histological group were used for the statistical analysis. This standard was used to exclude proteins that could not characterize at least 1 sample group. The missing values were estimated, based on a normal distribution (imputation width = 0.3, shift = 1.8) of log2-transformed LFQ intensities (90). Student's t-test ($P < 0.05$) was applied to identify significantly changed proteins. The seven comparative pairs that were used in the statistical analysis were LGD versus HGD (comparison 1), HGD versus invasive IPMN (comparison 2), and LGD versus invasive IPMN (comparison 3), SCN versus LGD (comparison 4), MCN versus LGD (comparison 5), SCN versus invasive IPMN (comparison 6), and MCN versus invasive IPMN (comparison 7). Proteins that were differentially expressed in at least 2 comparative pairs within comparisons 1 to 3 were considered initial biomarker candidates.

Subsequently, DEPs that increased or decreased consistently with greater malignancy of IPMN but not in the MCN and SCN groups were selected as final biomarker candidates of IPMN progression—i.e. the final candidates of Invasive IPMN must have been statistically significant in comparisons 6 and 7 and increased sequentially with greater IPMN malignancy. Similarly, the final candidates of LGD must have been statistically significant in comparisons 4 and 5 and decreased sequentially with greater IPMN malignancy.

9. Bioinformatics analysis

The gene ontologies (GOs) of the analyzed DEPs were explicated with the DAVID bioinformatics tool (<http://david.abcc.ncifcrif.gov/>) and UniprotKB database

(<http://www.uniprot.org/>). Pathway analysis was performed using the KEGG database (<http://www.genome.jp/kegg/>). Putative secretory proteins were confirmed using SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>), SecretomeP 2.0 (<http://www.cbs.dtu.dk/services/SecretomeP/>), and TMHMM, server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (93-95). The Plasma Proteome Database (PPD) was used to estimate the percentage of proteins that were identified simultaneously in human plasma and in this study (29). The proteins that were identified in this study were crossreferenced with mRNA and protein expression in the “pancreatic category” of the Human Protein Atlas (<http://www.proteinatlas.org/>). Ingenuity Pathway Analysis (IPA) was used for the functional analysis (Ingenuity Systems, <http://www.ingenuity.com/>). Fisher’s exact test ($P < 0.05$) was used in IPA to estimate the probability that a specific set of proteins was related to a pathway.

10. Enzyme-Linked Immunosorbent Assay (ELISA)

CD55 protein was measured using a commercial quantikine ELISA kit (CSB-E05121h, CUSABIO, China) per the manufacturer’s instructions. Seventy cyst fluid samples—22 LGD, 5 HGD, 14 invasive IPMN, 13 MCN, and 16 SCN—were centrifuged to isolate the supernatant for the ELISA. Equal amounts of proteins (298 µg, as measured by BCA assay) were loaded into each well of a 96-well plate. The protein concentration data were analyzed statistically by student’s t-test.

The intraplate repeatability of the CD55 ELISA was calculated by measuring 3 replicates of 10 positive control (cyst fluid from 2 HGD and 8 invasive IPMN) and 11 negative control samples (cyst fluid from 4 LGD and 7 SCN) on a

single plate. Three independent ELISA analyses were conducted at different times to analyze the interplate repeatability using a total of 21 samples (10 positive and 11 negative controls). The optical density (OD) was measured at a wavelength of 450 nm to calculate the standard deviation (SD) and coefficient of variation (CV) for each sample.

11. Western blot of CD55

Among the 70 cyst fluid samples that were used for ELISA, a portion of the samples (8 LGD, 4 HGD, 8 invasive IPMN, 5 MCN, and 5 SCN) were selected, based on their suitable protein concentration and remaining protein content, for further validation by western blot. Forty micrograms of cyst fluid samples, mixed with 5X SDS loading dye, were separated on 7% SDS-PAGE gels and transferred to PVDF membranes. The membranes were stained with Ponceau S (P7170, Sigma-Aldrich, MO, USA), blocked with 5% BSA for 2 h at RT, and incubated overnight at 4°C with CD55 rabbit monoclonal antibody (38730, Cell Signaling Tech., MA, USA) at 1:1000 (90). The membranes were washed 5 times with TBS-T for 10 minutes each and then incubated with goat anti-rabbit IgG (HRP) (ab6721, Abcam, Cambridge, UK) at 1:1000 for 2 h at RT. All signals were detected by LAS-4000 (Fujifilm, Tokyo, Japan) after incubation with ECL solution (West-Q Chemiluminescent Substrate Kit-plus, GenDEPOT, TX, USA).

12. Immunohistochemistry

The immunohistochemical analysis was performed on 4- μ m-thick unstained sections,

cut from formalin-fixed paraffin-embedded tissues, with CD55 rabbit monoclonal antibody (31759, Cell Signaling Tech., MA, USA) at 1:600 and myeloperoxidase (MPO) rabbit polyclonal antibody (A0398, Dako, Glostrup, Denmark) at 1:5000 on a BenchMark XT (Ventana Medical System, Tucson, AZ, USA) as described (96). All immunohistochemical stains were conducted and reviewed by an expert hepatopancreaticobiliary pathologist (Haeryoung Kim).

RESULTS

1. Cyst fluid sample characteristics

The demographics and clinical information of the study sample are described in Table 1. The histological composition of the 30 pancreatic cyst fluid samples consisted of LGD (n=10), HGD (n=5), invasive IPMN (n=5), MCN (n=5), and SCN (n=5). Among the 5 PCLs, there was no significant difference in composition, with the exception of cyst size and serum CEA and CA 19-9 levels, as measured by chemiluminescent microparticle immunoassay. Of the PCLs, the invasive IPMN patient group had the highest average concentrations of CEA and CA19-9 at 5.48 ± 6.82 mg/L and 90.28 ± 129.71 mg/L, respectively. The average CEA concentrations were similar between all groups, except for invasive IPMN, and increased gradually with the progression of IPMN dysplasia. Serum CA19-9 levels were generally higher in the more severe forms of IPMN, as were CEA levels. The average CEA level was approximately 4 times higher in invasive IPMN than in LGD and HGD. In addition, the average concentration of CA19-9 in invasive IPMN was approximately 8 times that of LGD and 4 times that of HGD. A baseline of 3 cm was used to classify cyst sizes (97). The MCN group had the largest average cyst size (7.50 ± 2.18 cm), followed by invasive IPMN (5.74 ± 3.69 cm).

2. In-depth quantitative proteomics of pancreatic cyst fluid

A mass spectrometry-based method, based on our previous study, was used to analyze a cohort of cyst fluid samples to measure the changes in protein expression with respect to the progression of IPMN (90). The overall procedure for discovering markers of IPMN progression, from sample preparation to the LC-MS/MS analysis, is depicted in Figure 1. The discovery cohort included 30 pancreatic cyst fluid samples from 3 types of IPMN (LGD, HGD, and invasive IPMN) and other PCLs (MCN and SCN). The pooled samples were fractionated and analyzed in parallel to generate a peptide library, which was used to expand the coverage of identified proteins for individual samples. Each fractionated sample was analyzed once, whereas all individual samples were analyzed in triplicate on a Q Exactive mass spectrometer.

Raw MS data were processed in MaxQuant (version 1.6.0.16), and the statistical analysis was performed with Perseus (version 1.6.1.1). The MaxQuant analysis identified 1,314,934 spectral matches, 56,583 peptides, and 5834 protein groups, 5774 of which were quantifiable. For label-free quantification, 5578 and 3249 proteins were identified in the peptide library and 30 individual cyst fluid samples, respectively. A total of 2993 proteins (92.1%) that were identified in individual samples overlapped with the peptide library (Figure 2A). The quantified proteins in individual cyst fluids accounted for 86.5% of the 3249 identified proteins (Figure 2B). Notably, the 3 IPMN groups had approximately twice the number of quantified proteins (2220–2500) than MCN (1218) and SCN (1346) (Figure 2C). Overall, the number of quantified proteins varied significantly, even within histological groups. Specifically, the identified and quantified proteins in each

sample ranged from a minimum of 657 identified (298 quantified) in LGD 10 to a maximum of 2587 identified (2014 quantified) in invasive IPMN 1 (Figure 2D).

To improve the proteome coverage, the “match between runs” feature in MaxQuant was utilized to align the retention times and MS/MS spectra of the individual sample against the peptide library (31). In total, an additional 773 and 420 proteins were identified and quantified, respectively, across all individual samples. LGD 6 showed the largest increase in the number of identified and quantified proteins—by 457 and 235, respectively. On average, 100 more peptides were identified in each sample (Figure 2C). This result demonstrates that the overall proteome coverage of individual cyst fluid samples rose, thereby enlarging the pool of potential biomarker candidates.

The dynamic range of protein expression levels spanned over 7 orders of magnitude, but most proteins (95%) were expressed within 4 orders of magnitude (Figure 3). Of these proteins, the levels of pancreatic cancer-associated proteins, such as MUC5AC, MUC2, and CEA, were high. Five proteins (PNLIP, CPA1, CPB1, PRSS1, and PRSS2) in a smaller dynamic range, as shown in Figure 3, are known to be significantly expressed in the pancreas compared with other organs (98). In addition, with the exception of PRSS1, these proteins, denoted in blue, are generally exclusive to the pancreas, per Wilhelm (98). This result confirms the presence of pancreas-specific proteins in our proteome data.

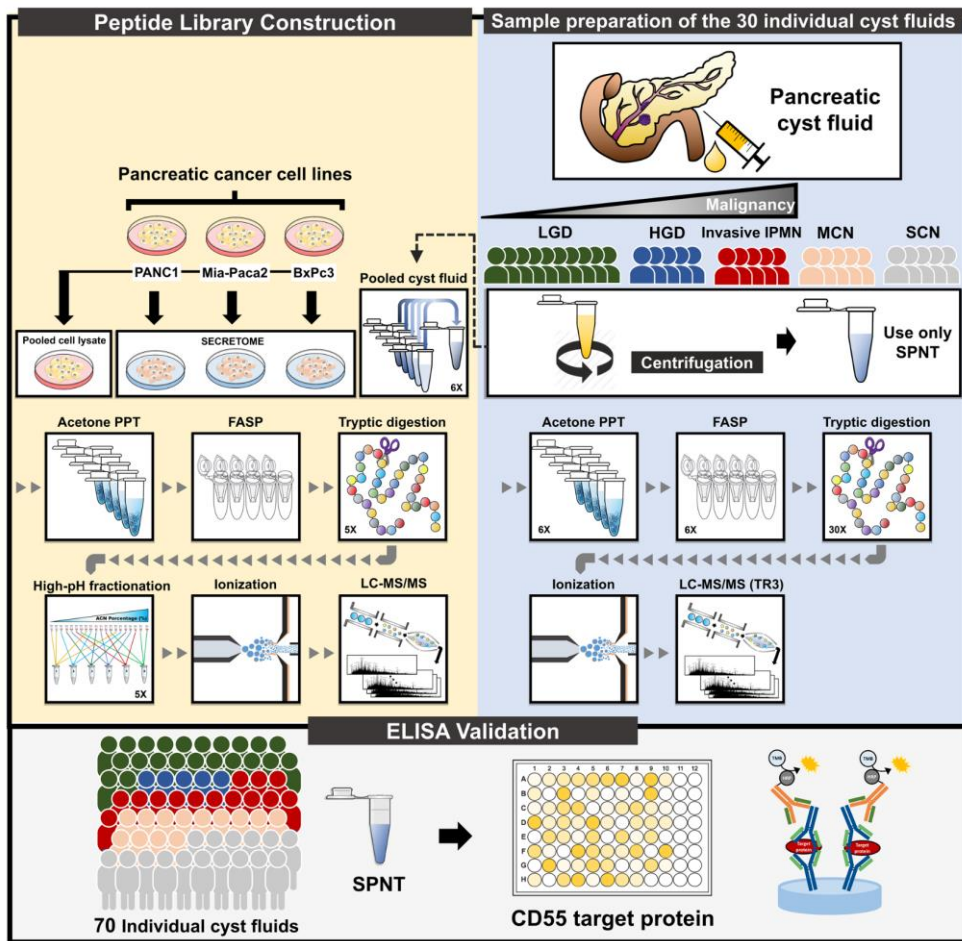


Figure 1. Experimental workflow.

The overall experimental workflow comprises 3 sections: (1) Preparation of 30 individual samples, (2) peptide library construction, and (3) validation by ELISA. LGD, low-grade dysplasia; HGD, high-grade dysplasia; MCN, mucinous cystic neoplasm; SCN, serous cystic neoplasm.

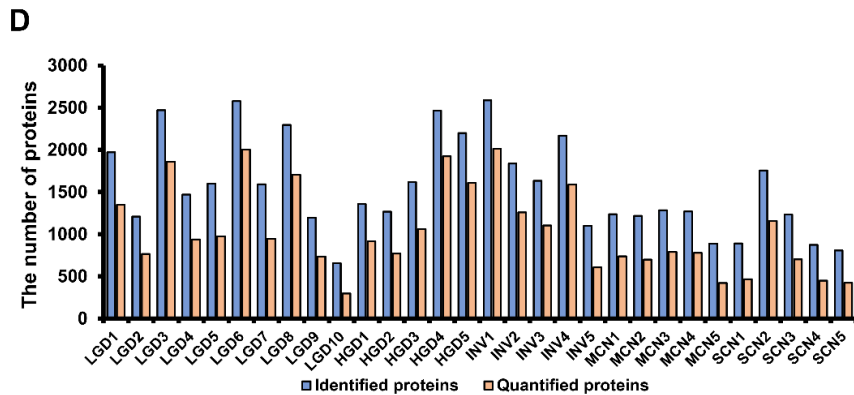
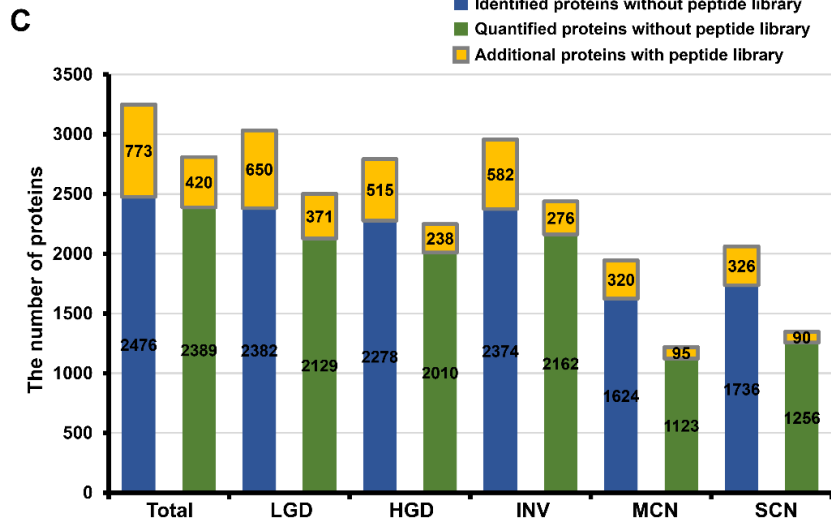
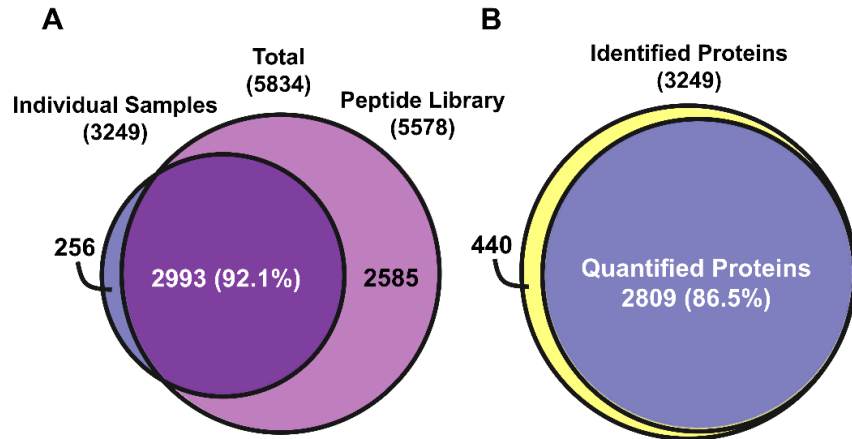


Figure 2. Comparison of protein identification and quantification in total dataset and in sample groups.

(A) Among the 5834 proteins in the dataset, 5578 and 3249 were identified in the peptide library and individual samples, respectively, and 92.1% of the proteins overlapped. (B) Identified and quantified proteins in 30 individual cyst fluids; 86.5% of proteins were quantifiable. (C) Bar graph of the extra coverage enabled by the peptide library in the identified and quantified proteins of each sample group. (D) The number of identified and quantified proteins in each individual sample is indicated in a single bar graph.

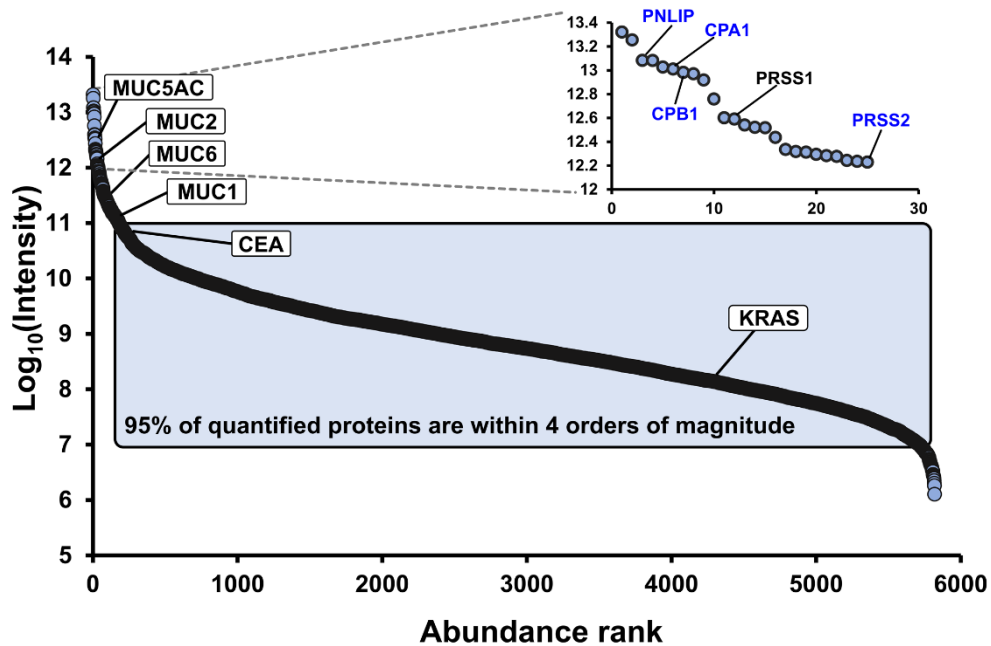


Figure 3. Dynamic range of quantified proteins.

The expression levels of quantified proteins exceeded 7 orders of magnitude. Several markers of pancreatic cancer (MUC5AC, MUC2, and CEA) were high in abundance. Five pancreas-specific proteins (PNLIP, CPA1, CPB1, PRSS1, and PRSS2) were ranked in the top 25.

3. Reproducibility of label-free quantification data

All CV values for the sums of log₂-transformed LFQ intensities across technical replicates of individual samples were less than 20% (0.959% to 13.279%). These data support that the variance that is attributed to the sample injection and quantification was low. The median CVs of log₂-transformed LFQ intensities between technical triplicates of each sample ranged from 0.555% to 3.564%, indicating that the label-free quantification of cyst fluid samples had high reproducibility (Figure 4A). In addition, the average Pearson correlation coefficients of the technical triplicates ranged from 0.871 to 0.954 (Figure 4B–F). The high Pearson correlation coefficients between technical replicates of individual samples indicate that the reproducibility between replicates is high and that the data are suitable for statistical analysis.

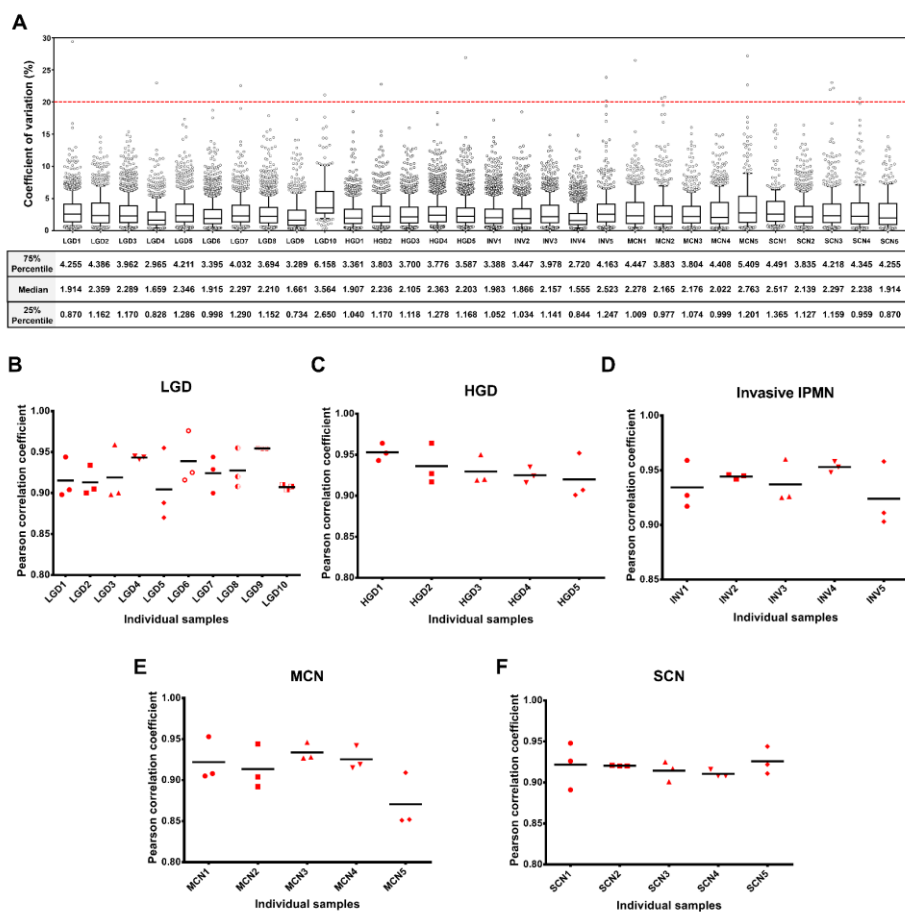


Figure 4. Coefficient of variation (CV%) values and Pearson correlation coefficients between technical triplicates in each individual sample.

(A) The median coefficient of variation (CV%) of log2-transformed LFQ intensity values between the technical triplicates in each individual is represented as box plots. Pearson correlation coefficients of technical replicates (TRs) in LGD (B), HGD (C), invasive IPMN (D), MCN (E), and SCN (F). The red markers represent Pearson correlation coefficient values of each pair of 3 technical replicates. The horizontal line represents the average of 3 Pearson correlation coefficient values in each individual sample.

4. Comparison with proteome databases and other researches

To examine the composition of pancreatic cyst fluid proteins, our data were compared with various proteome data from databases and past studies. The proteins that were identified in individual cyst fluid samples were compared with 3 public databases to screen for secreted proteins. As a result, 1578, 668, and 432 proteins were identified in SecretomeP, SignalP, and TMHMM, respectively (Figure 5A) (93-95). Secreted proteins accounted for 58.1% (1889 proteins) of the 3249 proteins that were identified in individual cyst samples (Figure 5B).

All protein accession numbers were mapped to gene symbols to compare them with 3 proteome databases: (1) the Human Plasma Protein Database, (2) the Human Protein Atlas (<http://www.proteinatlas.org>, June 11, 2018), and (3) the “core” proteome in Wilhelm et al (98). A total of 3039 proteins were listed with their corresponding gene symbols after redundant genes were removed. When compared with the Human Plasma Proteome Database, 79.8% (2424) of the 3039 identified proteins were confirmed to be expressed in plasma or serum (Figure 5C) (29). Our dataset was compared with the Human Protein Atlas to estimate the percentage of pancreatic tissue-specific proteins. As a result, 2937 (96.6%) genes had evidence of corresponding mRNA entries, and 2665 (87.7%) genes had evidence of corresponding protein entries in the pancreas (Figure 5D). Our data were compared with the core proteome in Wilhelm et al., which compiled 5 of the largest proteomic databases and extracted 11,578 human proteins that were ubiquitously expressed in all databases. The comparison found that 2714 (89.3%) of the identified proteins overlapped with those of the core proteome (Figure 5E) (98, 99).

Our previous study identified 2992 proteins in pancreatic cyst fluid samples of IPMN patients by LC-MS/MS analysis. This study identified 5834 proteins, surpassing the earlier report in terms of proteome coverage (Figure 5F) (90). In addition, the expression patterns of the marker candidates (AKR1B10, TFF1, SERPINA5, SERPINA4, MUC5AC, MUC2, TLN1, and TYMP) from our previous study were replicated here (Figure 6).

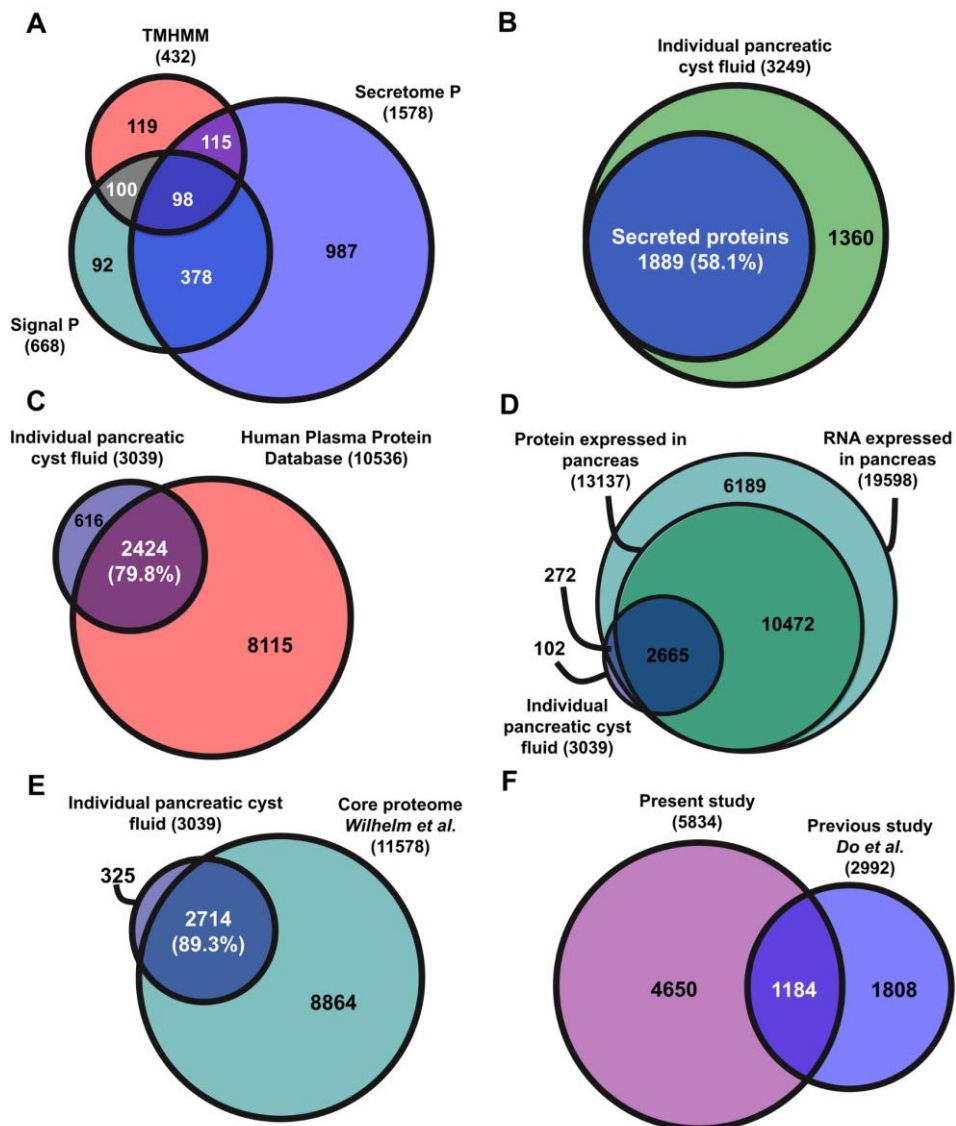


Figure 5. Comparative analysis with various proteome databases and other proteomic studies.

(A) Of the 1889 secreted proteins, 1578, 668, and 432 were predicted to be secreted by SecretomeP, SignalP, and TMHMM, respectively. (B) In total, secreted proteins accounted for 58.1% (1889 proteins) of the 3249 proteins identified in cyst fluid. (C) In comparison with the Human Plasma Proteome Database, 2424 (79.8%) proteins were observed in plasma or serum. (D) Compared with the Human Protein Atlas, 2937 (96.6%) and 2665 (87.7%) proteins had evidence of corresponding mRNA and protein entries, respectively, in the pancreas. (E) In a comparative analysis with identified proteins in individual cyst samples and the core proteome in Wilhelm et al., 2714 (89.3%) proteins were found to be core proteins. (F) A total of 5834 proteins were identified in our dataset, which was approximately twice that of our previous study (Do et al.).

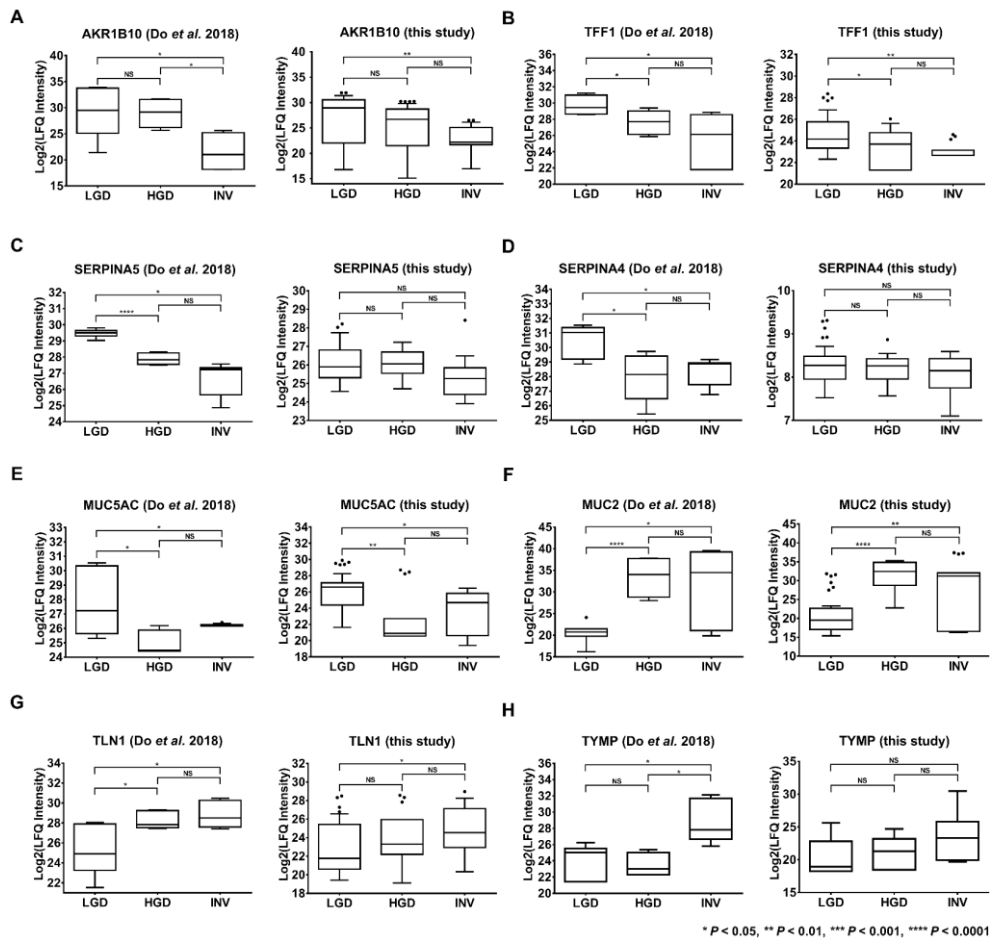


Figure 6. Comparison of the expression patterns of the final marker candidates between our previous and present studies.

In comparison with our previous study, the protein expression patterns in IPMN dysplasia of this dataset were consistent with the 8 final marker candidates (AKR1B10, TFF1, SERPINA5, SERPINA4, MUC5AC, MUC2, TLN1, and TYMP) from our earlier study. LGD, low-grade dysplasia; HGD, high-grade dysplasia; INV, invasive IPMN; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; NS, not significant.

5. Differentially expressed proteins between IPMN dysplasia

The diagram in Figure 7 details the discovery of potential markers of IPMN dysplasia. Of the 5834 identified proteins, 2809 were quantifiable in individual cyst fluid samples and had LFQ intensity values in at least 2 technical replicates in 1 biological replicate. Of the 2809 quantified proteins, 1019 had more than 70% measurable LFQ intensities in at least 1 histological group and were deemed usable for the statistical analysis. This criterion was established to ensure that a putative marker candidate represented at least 1 histological group.

To identify DEPs, student's t-test ($P < 0.05$) was performed for each comparative pair (comparisons 1 to 7). In comparisons 1 (LGD versus HGD), 2 (HGD versus invasive IPMN), and 3 (LGD versus invasive IPMN), 216, 84, and 247 proteins were differentially expressed, respectively—of which 164, 61, and 192 were upregulated.

The variance in expression between comparisons 1 to 3 was depicted in volcano plots. The highlighted final marker candidates including the validation target, CD55, underwent significantly large fold-changes (Figure 8). In the statistical analysis of comparisons 1 to 3, 364 DEPs remained after overlapping proteins were removed from each comparative group. Of the 364 DEPs, 261 were exclusively upregulated, and 80 proteins were exclusively downregulated. The remaining 23 proteins did not have consistent expression patterns across the 3 comparisons.

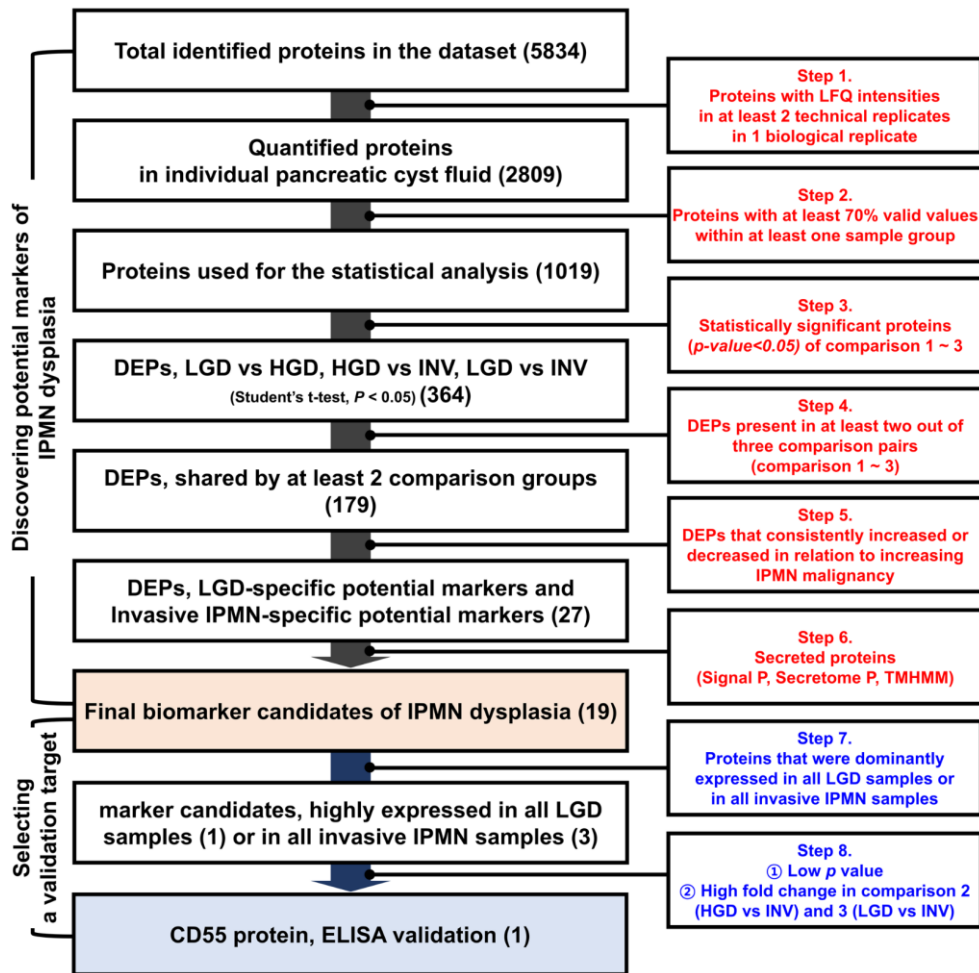


Figure 7. Flowchart of the discovery of potential markers of IPMN dysplasia.

Potential markers of the histological grades of IPMN were discovered following this logically sound step-by-step procedure. The procedure is composed of 6 steps for discovering potential markers of IPMN dysplasia and 2 steps for selecting a target for validation. The 70% valid value criterion in the step 2 was applied to eliminate proteins that failed to represent any histological group.

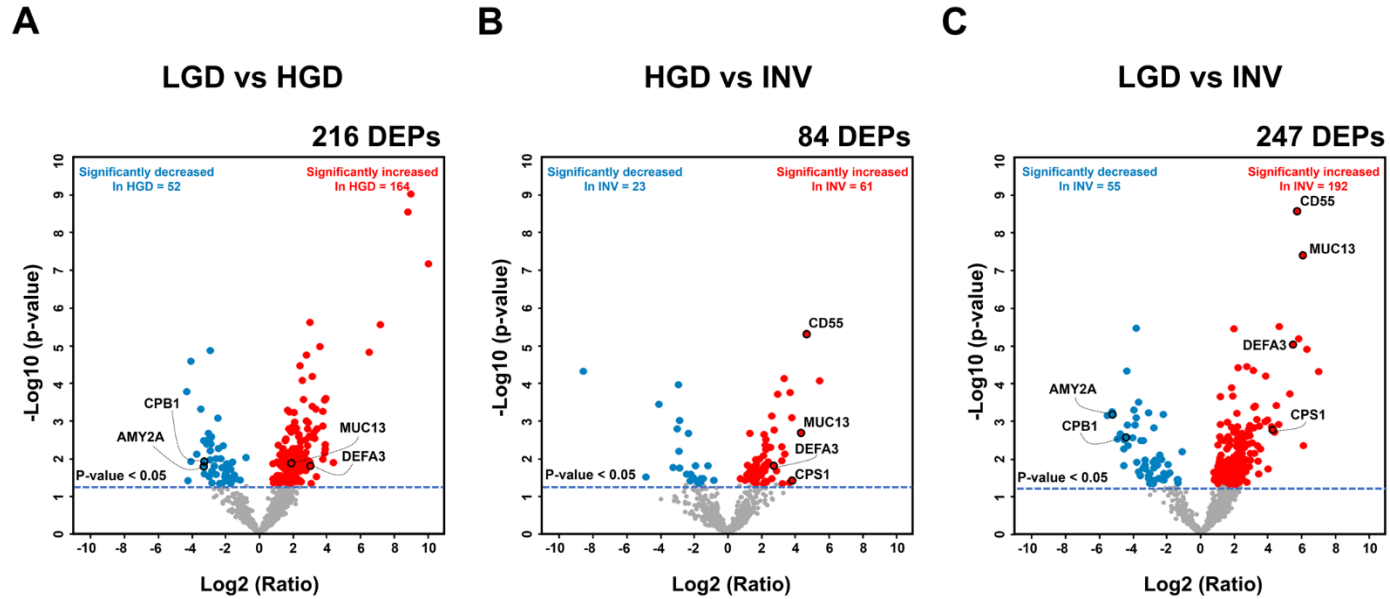


Figure 8. Volcano plots of differentially expressed proteins in three comparison groups.

Student's t-test ($P < 0.05$) was conducted for comparisons 1 (LGD versus HGD) (A), 2 (HGD versus invasive IPMN) (B), and 3 (LGD versus invasive IPMN) (C) to discover differentially expressed proteins (DEPs). The DEPs that were significantly expressed in each comparison group are indicated as colored dots (red: upregulated DEPs, blue: downregulated DEPs). Several marker candidates, including the validation target CD55, are highlighted in each comparison.

6. Molecular characterization of DEPs of IPMN dysplasia

Gene ontology (GO), KEGG pathway analysis, and Ingenuity Pathway Analysis (IPA) were used to characterize the 364 DEPs obtained from the statistical analysis of IPMN dysplasia. A total of 364 proteins were subsequently subjected to gene ontology (GO) analysis and filtered using a $P < 0.05$. The top 8 biological process (BP) terms were significantly associated with molecular transport and malignancy, such as “vesicle-mediated transport,” “secretion,” “exocytosis,” “cell death,” and “cell motility.” In addition, cellular components (CCs) of the DEPs were primarily related to the extracellular compartment, indicating that the proteins in pancreatic cyst fluid are secreted mainly from the surrounding cells and originally reside in the extracellular matrix. The terms that were associated with peptidase activity were present at higher proportions in molecular function (MF), demonstrating that pancreatic cyst fluid contains many digestive enzymes generated from the pancreas (Figure 9) (9). According to the KEGG pathway enrichment ($P < 0.05$), all DEPs of IPMN dysplasia belonged to the “pancreatic secretion pathway” and “glycosylation/gluconeogenesis pathway” (Figure 9).

IPA was conducted to better understand the association of DEPs with malignancy and their nature regarding pancreatic cyst fluid. A total of 216 and 247 DEPs from comparisons 1 (LGD vs HGD) and 3 (LGD vs invasive IPMN) were analyzed, respectively. To evaluate the biological functions that were associated with the DEPs of comparisons 1 and 3 and their activation levels, core analysis was conducted by using the protein accession numbers and their fold-change values (Figure 10A, B). As expected, the biological functions that were related to

malignancy and molecular secretion had a higher rank among the biological functions that were associated with the DEPs of comparisons 1 and 3. The representative functions included “cellular movement,” “cancer,” “organismal injury and abnormalities,” “organismal development,” and “molecular transport,” which included such subcategories as “cell spreading,” “angiogenesis,” “secretion of molecules,” and “secretion of proteins.”

Heat maps of the comparative analysis were used to visualize the diseases and biological function terms across analyses simultaneously to detect trends and significant clusters (Figure 10C, D), allowing us to identify the diseases and biological functions that are predicted to increase or decrease similarly across comparisons 1 and 3. Consistent with our expectations, the biological functions that were related to malignancy (cell spreading, vasculogenesis, and cancer) and molecular secretion (secretion of molecules and secretion of proteins) were highly expressed in comparisons 1 and 3. In addition, the expression levels of malignancy-related terms and molecular secretion-related terms were higher in comparison 3 than in comparison 1, and “cell death of pancreatic cancer cell lines” decreased in comparison 3 (Figure 10C). Pancreas-specific diseases (chronic pancreatitis and poorly differentiated malignant pancreatic tumor) and cancer were significantly associated with the DEPs of comparison 3 (Figure 10D).

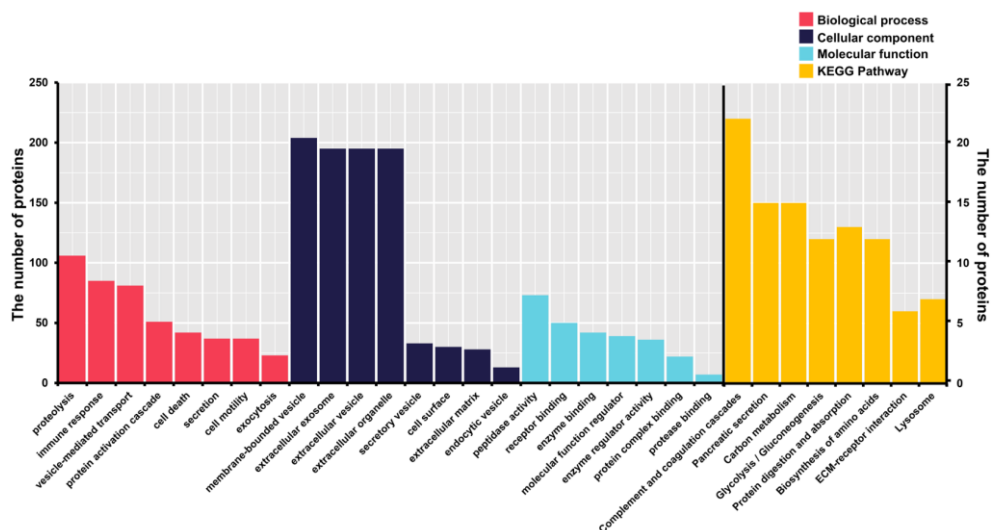


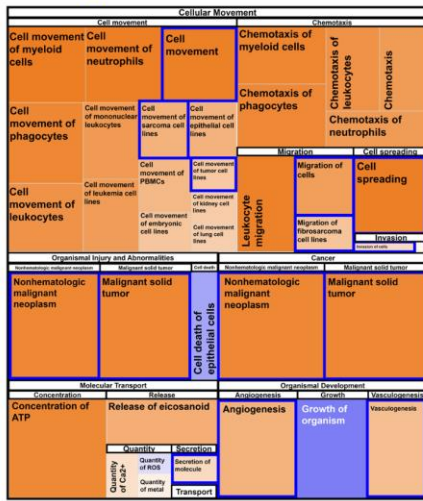
Figure 9. Results of Gene Ontology and KEGG pathway analyses.

GO and KEGG pathway analyses were conducted using the DAVID bioinformatics tool. A total of 364 DEPs that originated from the statistical analysis between IPMN dysplasia were subjected to GO and KEGG pathway analyses. Each colored bar graph indicates the enriched terms in biological process (BP), cellular component (CC), molecular function (MF), and KEGG pathway. The number of participating proteins is shown on the left y-axis for GO terms and the right y-axis for KEGG pathway terms.

A

-2.621 2.957

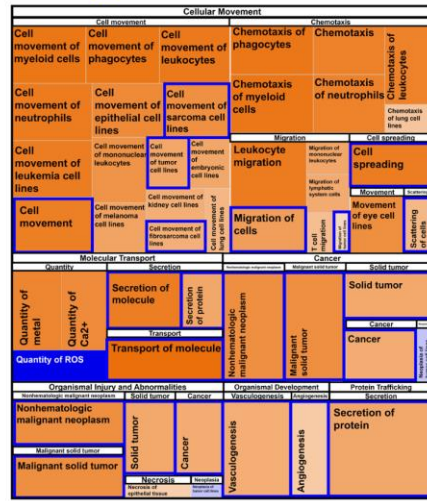
Biological terms associated with malignancy OR molecular secretion



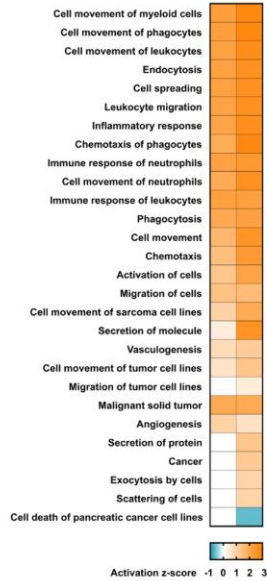
B

-1.876 3.327

Biological terms associated with malignancy OR molecular secretion



C

LGD vs HGD
LGD vs INV

D

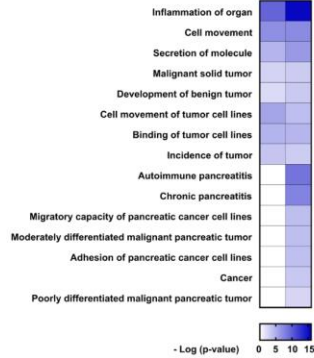
LGD vs HGD
LGD vs INV

Figure 10. Ingenuity Pathway Analysis.

Core analysis of IPA was conducted to evaluate the biological functions associated with the 216 DEPs of comparison 1 (LGD versus HGD) (A) and the 247 DEPs in comparison 3 (LGD versus invasive IPMN) (B). The heat maps indicate the upregulated and downregulated biological functions. The z-score of each biological function is represented by the sizes and colors of the boxes. Larger boxes signify higher z-scores. Orange indicates positive z-scores, and blue denotes negative z-scores. (C, D) Comparative analysis of IPA, visualizing the diseases and biological function terms across multiple analyses (comparisons 1 and 3) simultaneously. (C) The disease and biological function terms related to malignancy and molecular secretion were highly expressed in both comparisons. (D) Pancreas-specific diseases and cancer were significantly associated with the DEPs of comparison 3.

7. Biomarker candidates of IPMN dysplasia

A total of 364 DEPs passed the statistical analysis. Following the rationale that proteins with significant differences in expression in more comparisons are more likely to be biomarkers, 179 DEPs were designated as initial marker candidates, because they were present in at least 2 of 3 comparative pairs (1 to 3) (31, 90). Subsequently, 27 DEPs had expression patterns that consistently increased or decreased with greater IPMN malignancy. Of them, 13 DEPs that were preferentially expressed in invasive IPMN were statistically significant in comparisons 6 (SCN versus invasive IPMN) and 7 (MCN versus invasive IPMN). Similarly, the remaining 14 DEPs were expressed in LGD and showed significant differences in comparisons 4 (SCN versus LGD) and 5 (MCN versus LGD). Based on the rationale that tumor-associated proteins are generally secreted from surrounding tumor cells, the 19 DEPs predicted to be secreted by SecretomeP, SignalP, and TMHMM were selected as the final marker candidates (100, 101).

The heat map in Figure 11 provides an overview of the expression of the 19 final marker candidates of IPMN dysplasia: 7 invasive IPMN-specific marker candidates and 12 LGD-specific marker candidates based on their expression patterns. DEFA3, MUC13, CD55, CPS1, RAB11B, HEXA, and SOD2 were highly expressed in invasive IPMN and expressed at statistically lower levels in other PCLs. In contrast, LEFTY1, AMY2A, KLK1, RNASE1, CELA2A, CELA3A, CPA1, CPB1, CEL, AMY2B, GP2, and CTRC were highly expressed in LGD and but significantly lower in other cystic lesions (Figure 12).

Table 2 details the results of the statistical analysis of the 19 potential markers (7 upregulated proteins and 12 downregulated proteins), including statistical significance, p-values, and fold-changes for 3 comparative groups (1 to 3). A total of 16 proteins, with the exception of RAB11B, KLK1, and CELA2A, were pancreatic tissue-specific proteins, according to the Human Protein Atlas. In addition, 15 proteins, with the exception of DEFA3, MUC13, RAB11B, and LEFTY1, were observed in plasma or serum, per the Plasma Proteome Database (PPD).

The fold-changes of the 19 final marker candidates were assessed in relation to the general distribution of the 1019 proteins that were used for the statistical analysis. To this end, potential markers from comparisons 1 and 3 were displayed in a dynamic range and ordered, based on their fold-change. (Figure 13A, B). Excluding CD55, RAB11B, and CPS1 in comparison 1, 16 proteins had p-values below 0.05 and lay generally near the 2 extremes of the dynamic range. Similarly, in comparison 3, 19 candidates were statistically significant ($P < 0.05$) and located near the upper and lower extremes of the dynamic range. All potential markers had higher fold-changes than CEA, the most well-established pancreatic cancer-associated marker.

Upstream regulator analysis in IPA was conducted to predict the upstream proteins of the final candidates and their biological functions. This analysis predicted the top 20 likely upstream regulators that modulate the 5 final candidates (MUC13, CD55, CPS1, SOD2, and LEFTY1). A total of 11 regulator proteins were associated with SOD2, 4 proteins correlated with CD55, 4 proteins were linked to LEFTY1, and 1 was associated with CPS1 and MUC13 (Figure 13C, Table 3). The upstream

regulators were the following molecular types: Kinase, enzyme, g-protein-coupled receptor, transcription regulator, transporter, and ion channel. The upstream regulators were tumor suppressors, pancreatic mitogens, and other key factors of cancer progression, according to several publications, supporting the credibility of the marker candidates (102-107).

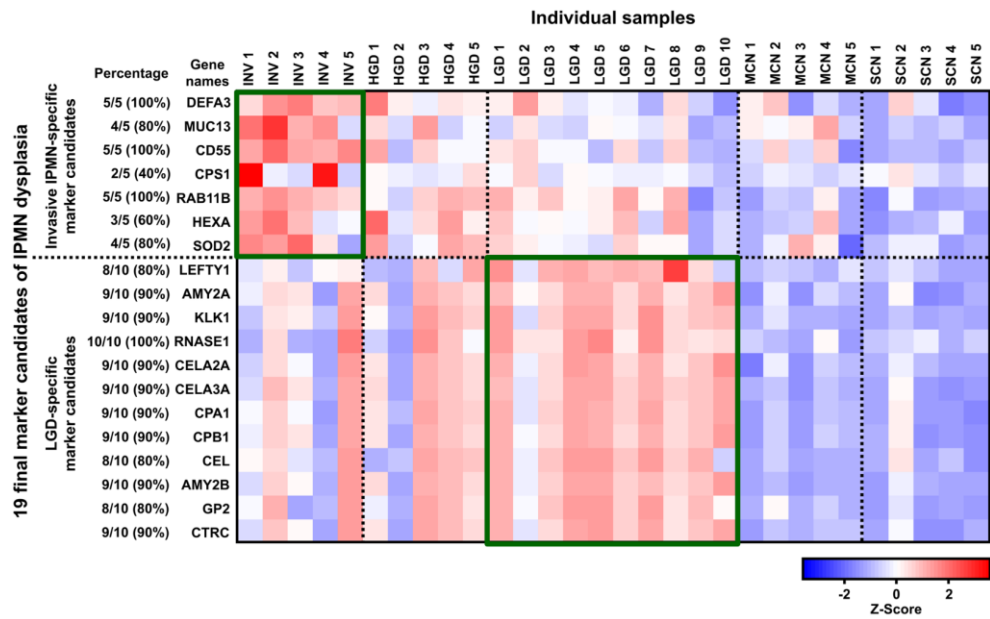


Figure 11. Overview of protein expression of the 19 final marker candidates of IPMN dysplasia.

The heat map represents the expression patterns of the 19 final potential markers of IPMN dysplasia, based on z-score. Z-scores were calculated by averaging the LFQ intensities of the 3 technical replicates of each biological replicate. Red and blue reflect positive and negative values, respectively. Seven invasive IPMN-specific marker candidates were highly expressed in invasive IPMN. In contrast, 12 LGD-specific marker candidates were predominantly expressed in LGD. The percentage reflects the proportion of the 5 invasive IPMN samples showed upregulation of the given protein. Similarly, the percentage of LGD-specific marker candidates represents the number of LGD samples that showed upregulation for the given protein.

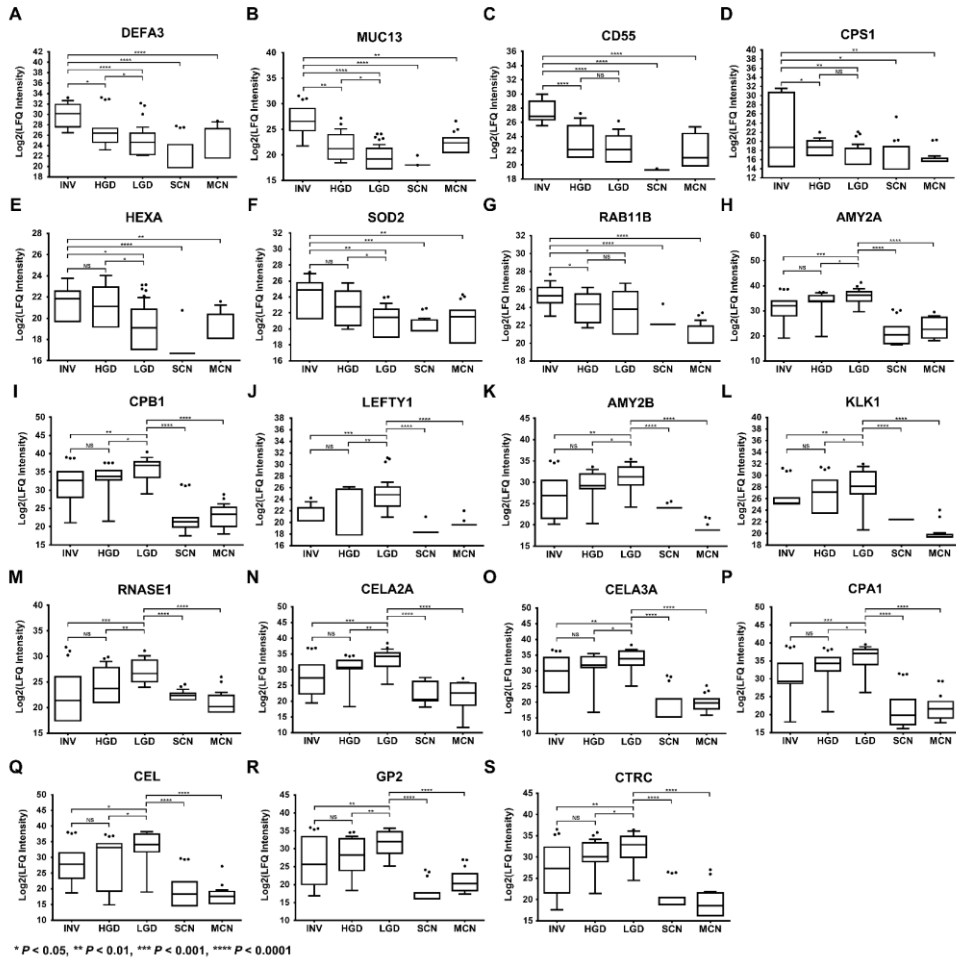


Figure 12. 19 potential markers that were differentially expressed in accordance with the degree of IPMN malignancy.

DEFA3 (A), MUC13 (B), CD55 (C), CPS1 (D), HEXA (E), SOD2 (F), RAB11B (G), AMY2A (H), (I) CPB1, (J) LEFTY1, (K) AMY2B, (L) KLK1, (M) RNASE1, (N) CELA2A, (O) CELA3A, (P) CPA1, (Q) CEL, (R) GP2, and (S) CTSC were differentially expressed, in accordance with the histological grades of IPMN. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; NS, not significant.

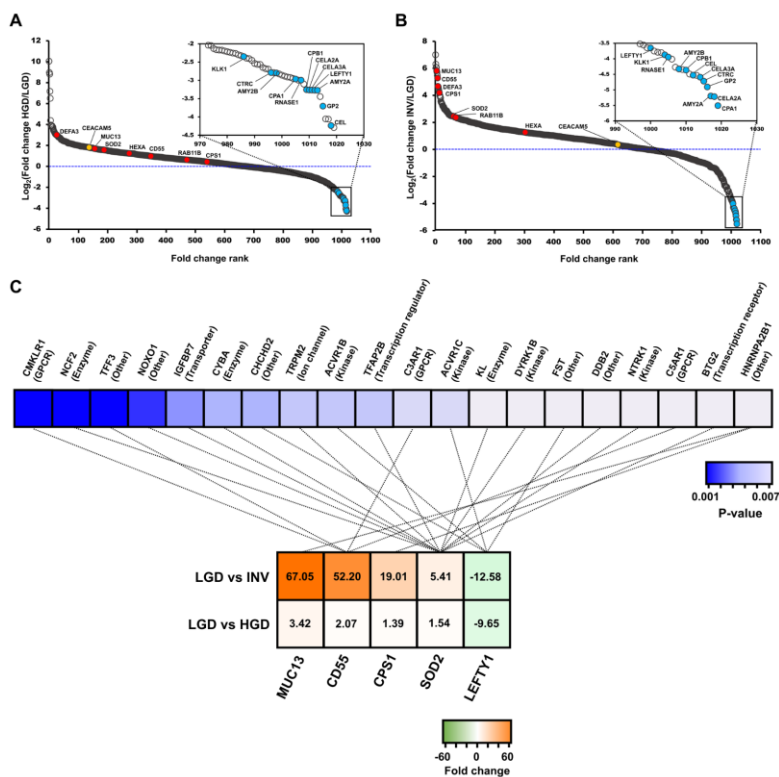


Figure 13. The dynamic range of protein fold-changes in comparisons 1 and 3 and the results of upstream regulator analysis in Ingenuity Pathway Analysis.

The dynamic range with marked fold-changes of the 19 final marker candidates in comparisons 1 (LGD versus HGD) (A) and 3 (LGD versus invasive IPMN) (B). The red and blue dots indicate the log2-transformed fold-changes of the 7 upregulated and 12 downregulated proteins. The yellow dot represents the log2-transformed fold-change of CEA protein. (C) Five potential markers (MUC13, CD55, CPS1, SOD2, and LEFTY1) and their upstream regulators are connected by dotted lines. The molecular types of each upstream regulator are shown in parentheses. Each marker candidate is listed in order of decreasing fold-change, whereas the upstream regulators are listed in order of increasing p-values.

Table 2. Detailed statistical analysis of 19 final marker candidates.

		LGD vs HGD			HGD vs INV			LGD vs INV		
	Gene name	<i>T</i> test Significance	adjusted <i>P</i> value	Log ₂ (Fold change)	<i>T</i> test Significance	adjusted <i>P</i> value	Log ₂ (Fold change)	<i>T</i> test Significance	adjusted <i>P</i> value	Log ₂ (Fold change)
Up	DEFA3	+	0.022	2.772	+	0.016	2.695	+	0.000	5.467
	MUC13	+	0.021	1.776	+	0.002	4.291	+	0.000	6.067
	CD55		0.234	1.054	+	0.000	4.652	+	0.000	5.706
	CPS1		0.351	0.479	+	0.038	3.770	+	0.002	4.249
	HEXA	+	0.032	1.237		0.952	0.043	+	0.036	1.280
	SOD2	+	0.024	1.585		0.402	0.878	+	0.003	2.463
	RAB11B		0.541	0.625	+	0.030	1.811	+	0.011	2.436
Down	AMY2A	+	0.018	-3.272		0.416	-1.925	+	0.001	-5.197
	CPB1	+	0.013	-3.255		0.611	-1.109	+	0.003	-4.364
	LEFTY1	+	0.003	-3.271		0.719	-0.382	+	0.000	-3.653
	AMY2B	+	0.026	-2.798		0.440	-1.530	+	0.001	-4.327
	KLK1	+	0.050	-2.349		0.375	-1.523	+	0.001	-3.872
	RNASE1	+	0.002	-2.995		0.567	-0.950	+	0.001	-3.945
	CELA2A	+	0.013	-3.263		0.353	-1.955	+	0.001	-5.219
	CELA3A	+	0.028	-3.265		0.570	-1.327	+	0.003	-4.592
	CPA1	+	0.034	-2.963		0.267	-2.544	+	0.001	-5.507
	CEL	+	0.042	-4.237		0.915	-0.290	+	0.016	-4.527
	GP2	+	0.009	-3.704		0.615	-1.202	+	0.003	-4.906
	CTRC	+	0.049	-2.788		0.394	-1.933	+	0.002	-4.721

LGD, low-grade dysplasia; HGD, high-grade dysplasia.

Table 3. Results of upstream regulator analysis in Ingenuity Pathway Analysis.

No.	Target molecules in dataset	Upstream regulator	Protein IDs	Protein names	Molecule type	p-value of overlap
1	SOD2	NCF2	P19878	Neutrophil cytosol factor 2	enzyme	0.001
		NOXO1	Q8NFA2	NADPH oxidase organizer 1	other	0.002
		IGFBP7	Q16270	Insulin-like growth factor-binding protein 7	transporter	0.004
		CYBA	P13498	Cytochrome b-245 light chain	enzyme	0.005
		TRPM2	O94759	Transient receptor potential cation channel subfamily M member 2	ion channel	0.006
		TFAP2B	Q92481	Transcription factor AP-2-beta	transcription regulator	0.006
		KL	Q9UEF7	Klotho	enzyme	0.008
		DYRK1B	Q9Y463	Dual specificity tyrosine-phosphorylation-regulated kinase 1B	kinase	0.008
		DDB2	Q92466	DNA damage-binding protein 2	other	0.008
		NTRK1	P04629	High affinity nerve growth factor receptor	kinase	0.008
2	CD55	BTG2	P78543	Protein BTG2	transcription regulator	0.008
		CMKLR1	Q99788	Chemokine-like receptor 1	g-protein coupled receptor	0.001
		TFF3	Q07654	Trefoil factor 3	other	0.001
		C3AR1	Q16581	C3a anaphylatoxin chemotactic receptor	g-protein coupled receptor	0.007
3	LEFTY1	C5AR1	P21730	C5a anaphylatoxin chemotactic receptor 1	g-protein coupled receptor	0.008
		CHCHD2	Q9Y6H1	Coiled-coil-helix-coiled-coil-helix domain-containing protein 2	other	0.005
		ACVR1B	P36896	Activin receptor type-1B	kinase	0.006
		ACVR1C (ALK7)	Q8NER5	Activin receptor type-1C	kinase	0.007
		FST	P19883	Follistatin	other	0.008
4	CPS1, MUC13	HNRNPA2B1	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	other	0.008

8. Determination of CD55 by ELISA

The selection of the validation target was based on a protein's predominance in LGD or invasive IPMN and its statistical significance (Figure 7). Accordingly, CD55 was chosen as a validation target of IPMN dysplasia for 3 main reasons: (1) It was highly expressed in all (100%, 5/5) invasive IPMN samples (Figure 11), (2) it had the smallest p-value of all potential markers from all comparative groups that involved invasive IPMN (Figure 12, Table 2), and (3) it had the highest fold-change between comparisons 2 (HGD vs. invasive IPMN) and 3 (LGD vs. invasive IPMN) (Figure 8).

CD55 was validated in 70 cyst fluid samples (22 LGD, 5 HGD, 14 invasive IPMN, 13 MCN, and 16 SCN) by ELISA. CD55 concentrations in individual cyst fluid samples were calculated and demonstrated in 2 types of IPMN classification (Figure 14A, B). The concentration of CD55 was the highest in invasive IPMN, and its expression patterns generally correlated with LFQ intensity values. CD55 concentrations in invasive IPMN (mean: 1.354 ng/mL, STDEV: 1.532 ng/mL) were significantly higher than in LGD (mean: 0.598 ng/mL, STDEV: 1.045) ($P < 0.05$). In addition, CD55 concentrations in high-risk IPMN (mean: 1.219 ng/mL, STDEV: 1.567) were significantly higher versus low-risk IPMN (mean: 0.598 ng/mL, STDEV: 1.045). The expression levels of CD55 in invasive IPMN compared with SCN and MCN were statistically significant ($P < 0.05$).

The intraplate and interplate repeatability of the ELISA were calculated to evaluate the precision of the CD55 ELISA. The range of CV values of each control sample is summarized in Figure 14C. The OD₄₅₀ values of 3 replicates with 10

positive and 11 negative control samples were measured in a single ELISA plate to evaluate the intraplate repeatability. The OD₄₅₀ values had CVs of between 0.363% and 12.663%, with an average CV value of 3.645%. The OD₄₅₀ value of each control sample was measured at different times to evaluate the interplate repeatability. The analysis yielded an average CV value of 7.130%, and individual CVs ranged from 1.378% to 19.846%. The CV values for the intra-assay and inter-assay comparisons were less than 20%, indicating that the CD55 ELISA is highly reproducible and stable.

9. Western blot of CD55

To reconfirm CD55 expression between the 5 cystic lesions, Western blot was conducted using 30 cyst fluid samples (8 LGD, 4 HGD, 8 invasive IPMN, 5 MCN, and 5 SCN). Ponceau S staining was included as a loading control to confirm that comparable amounts of individual samples were loaded onto each gel. The resulting CV value was 12.53% (Figure 15A). The signal intensity of CD55 was the highest in invasive IPMN, and its expression patterns correlated with the MS analysis findings. The signal intensities in invasive IPMN were significantly higher than in LGD ($P < 0.001$), MCN ($P < 0.01$), and SCN ($P < 0.01$) (Figure 15B). In addition, the signal intensities in high-risk IPMN were significantly higher versus low-risk IPMN ($P < 0.001$) (Figure 15C).

10. Immunohistochemistry of CD55 and Myeloperoxidase

Immunohistochemical stains for CD55 and MPO were performed on formalin-fixed paraffin-embedded (FFPE) tissue sections from SCN, LGD, HGD, and invasive IPMN (Figure 16). CD55 expression was observed predominantly in the apical border of the tumor epithelial cells, showing increased distribution and intensity, in accordance with the histological grades of IPMN, whereas strong membranous staining was observed in invasive IPMN (Figure 16A–D). Also, we examined MPO, a neutrophil marker, to identify neutrophil infiltration in invasive IPMN, based on a previous study that reported that CD55 is responsible for transepithelial migration of neutrophils (108). As expected, neutrophil infiltration increased from LGD to HGD and invasive IPMN, which had the highest neutrophil counts (Figure 16E–H). CD55 expression and neutrophil infiltration were not observed in SCN.

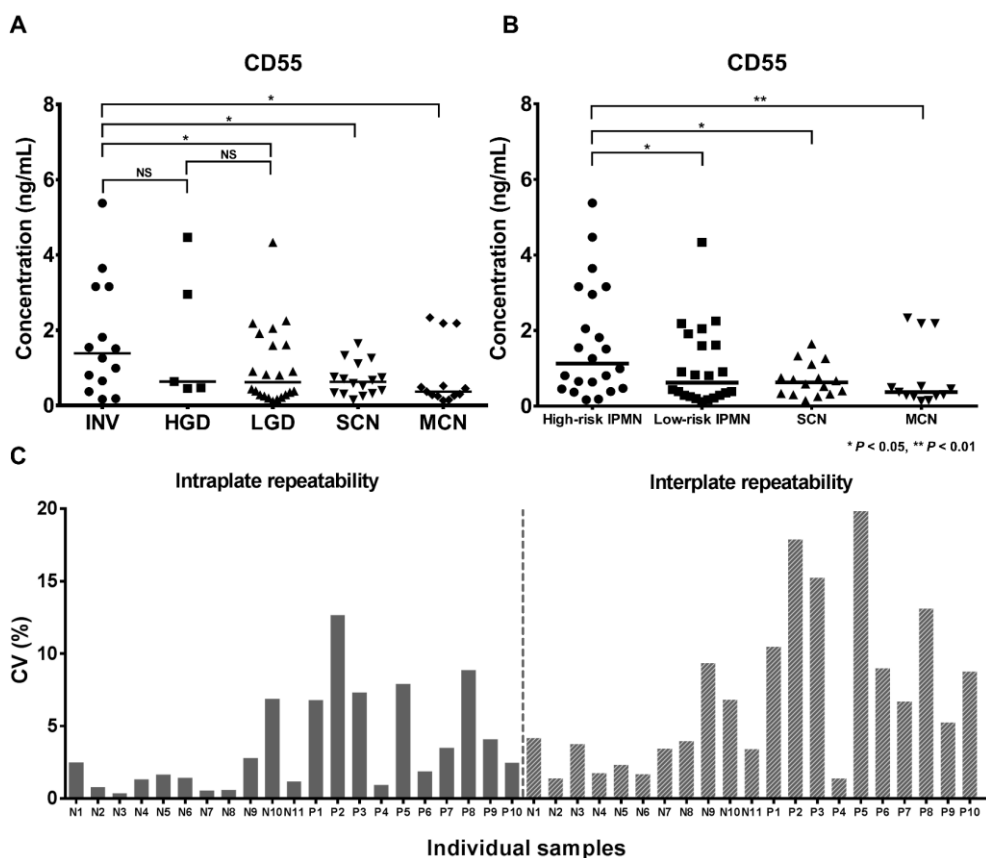


Figure 14. Validation of CD55 as a potential biomarker target by ELISA.

The CD55 concentrations are indicated according to two types of IPMN classification. The median CD55 concentrations were highest in invasive IPMN (A) and high-risk IPMN (B). (C) The coefficient of variation values of 3 replicates of negative controls (N1–N11) and positive controls (P1–P10) in intraplate and interplate repeatability of CD55 by ELISA. Intraplate repeatability was measured by using 3 replicates for each of the 21 control samples on a single plate. Interplate repeatability was evaluated by operating 3 plates, including 21 control samples at different times. *, $P < 0.05$; **, $P < 0.01$; NS, not significant.

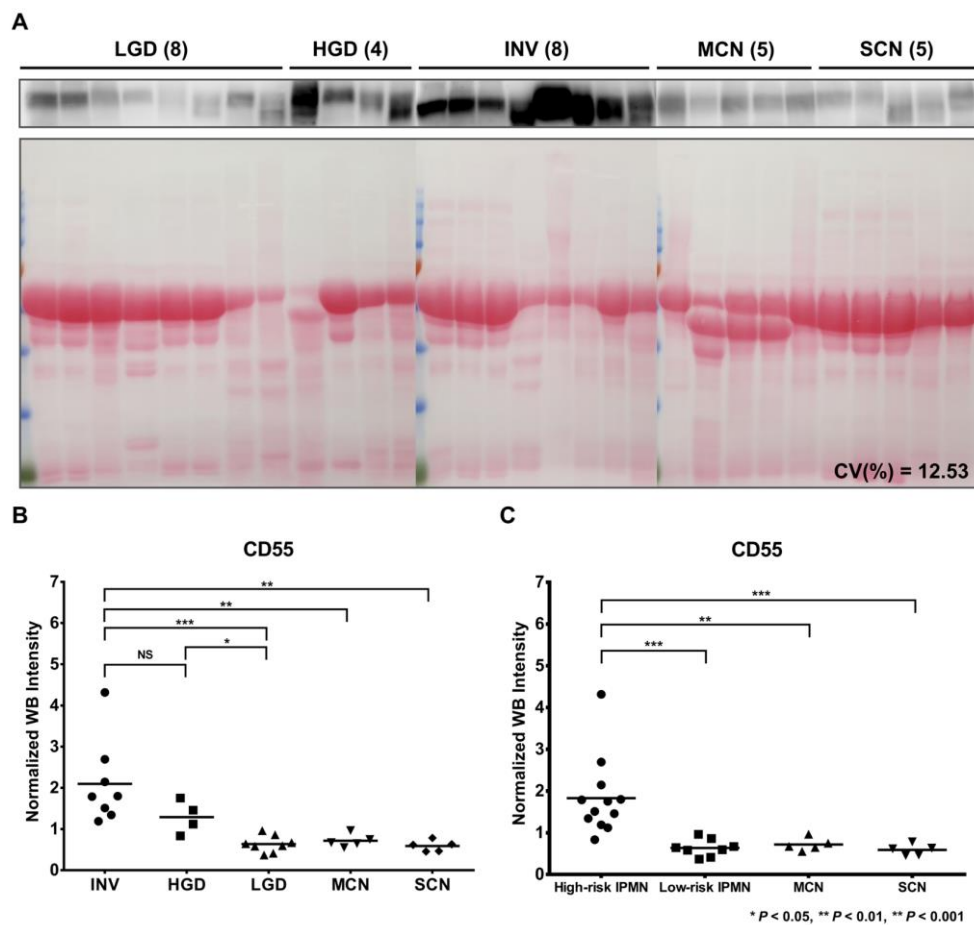
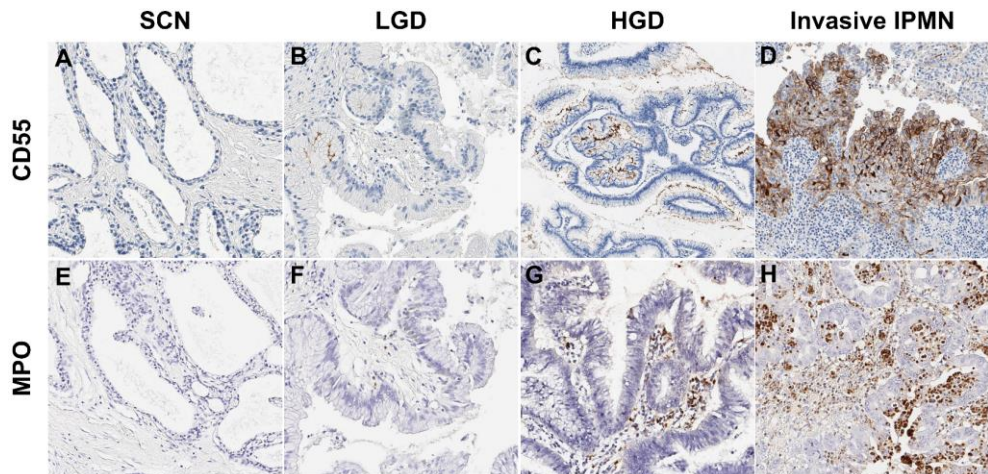


Figure 15. Validation of CD55 as a potential marker by Western blot.

(A) Western blot band of CD55, with Ponceau S staining used as a loading control.

The scatter dot plots of signal intensities are indicated according to two types of IPMN classification (B, C). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; NS, not significant.



(Haeryoung Kim, Cancers 12(9):2383)

Figure 16. Immunohistochemical staining of CD55 and MPO.

CD55 expression in serous cystic neoplasm (A), low-grade dysplasia (B), high-grade dysplasia (C), and invasive IPMN (D). Neutrophil infiltration in serous cystic neoplasm (E), low-grade dysplasia (F), high-grade dysplasia (G), and invasive IPMN (H). A–D: CD55 immunohistochemistry, E–H: myeloperoxidase (MPO) immunohistochemistry, original magnification x400.

DISCUSSION

In this study, we discovered reliable marker candidates of IPMN dysplasia using cyst fluid from IPMN, MCN, and SCN patients by LC-MS/MS and investigated their molecular characterization, based on the advantages of pancreatic cyst fluid and MS-based proteomic approaches (86-88). The current diagnostic screens cannot accurately determine the IPMN-associated grade of dysplasia (109), leading to unnecessary surgical resections for low-risk IPMN patients (3). In addition, most studies have focused on discovering diagnostic markers that differentiate IPMNs from other PCLs, rather than the grade of dysplasia in IPMN (110). Our report is the first study to discover potential markers of IPMN dysplasia using cyst fluid from 3 major types of PCL by LC-MS/MS and validated them by orthogonal method.

Our proteome data have 3 notable aspects: (1) increased depth of proteome coverage through the use of a peptide library, (2) high reproducibility, and (3) an abundance of pancreas-associated proteins. We hypothesized that a high proportion of proteins are coexpressed in cyst fluid and pancreatic cancer cell lines. Thus, we expected that the “match between runs” feature in MaxQuant would help identify proteins in cyst fluid that are normally unidentifiable without a peptide library (92). Consistent with our expectation, approximately twice as many proteins were identified in this dataset than in our previous study (Figure 5F), constituting the largest proteomic dataset of pancreatic cyst fluid (16, 17, 90). Consequently, we discovered potential markers of the histological grades of IPMN in a larger pool of

proteins. Further, the low median CV values and high Pearson correlation coefficients of LFQ intensities between technical replicates indicated that the individual samples were injected into the mass spectrometer without significant variance and that the technical replicates were analyzed reproducibly. (Figure 4). In previous studies, it was concluded that pancreatic cyst fluid contains secreted proteins from surrounding tumor cells (100, 101) and plasma proteins that penetrate into the cyst epithelium due to tissue injury or the enhanced permeability and retention (EPR) effect of the surrounding blood vessels (111). Tumor-promoting mediators, such as immunosuppressive cytokines, that are released from cancer-associated fibroblasts and mast cells in the tumor microenvironment can promote the neoplastic evolution of IPMN (112-114). They induce an aggressive phenotype and drug resistance in premalignant pancreatic lesions. One possible mechanism of the malignant evolution of IPMN, in light of our findings and existing literature, is that the expression of CD55, promoted by the immunosuppressive cytokine, interleukin-4, prevents complement-dependent cytotoxicity in cancer cells, which consequently accelerates the malignant transformation of IPMN dysplasia (115, 116). The results of the comparative analyses with the 3 databases for the secretome analysis and the Human Plasma Protein Database (Figure 5A–C) support these findings (100, 101, 110). In addition, 15 of the final marker candidates were detected in plasma and serum, supporting their viability in a blood-based assay (29). A high proportion (approximately 90%) of pancreas-associated proteins were identified in our dataset (Figure 5D, E) and 5 pancreas tissue-specific proteins, as defined by Wilhelm et al.

(98), were in the top 25 proteins (Figure 3), demonstrating that our proteome has sufficient coverage of pancreas-specific proteins.

GO and KEGG pathway analyses were performed to identify key processes of the 364 DEPs in IPMN dysplasia. The results indicated the enrichment of terms that pertained to tumorigenesis: pancreatic secretion, enzymatic activity, and malignancy (Figure 9). GO terms that were related to “pancreatic secretion” and “molecular transport” were highly ranked in the GO and KEGG analyses, suggesting that the DEPs in IPMN dysplasia are generally secreted from the surrounding tumor cells. One of the most highly enriched GO terms, “proteolysis,” is the most fundamental feature of malignancy (117). Proteolytic degradation of ECM constituents accelerates tumor cell growth, migration, and angiogenesis. The most highly enriched KEGG pathway, “complement and coagulation cascades,” is associated with tumor growth and metastasis (118, 119). Complement activation promotes an immunosuppressive microenvironment and thus induces angiogenesis, activating cancer-related signaling pathways. In addition, several studies have reported increases in complement activity in biological fluids from cancer patients (120, 121). Coagulation cascades can be activated directly by cancer procoagulants, which are released by tumor cells.

The criteria for discovering marker candidates of IPMN dysplasia that comprised 8 steps (Figure 7) and the association of the final marker candidates with the pancreas-related disease (IPMN and PDAC) and malignancy support the credibility of the potential markers. For instance, MUC13 has been studied extensively. According to 2 previous studies with similar goals as ours, MUC13

increased with histological grade of IPMN (122, 123). In addition, several studies have indicated that MUC13 is highly upregulated in PDAC tissue but not in adjacent normal tissue and is related to PDAC progression (124, 125). CD55 is involved in the dedifferentiation, invasiveness, migration, and metastasis of tumors and association with pancreatic cancer (126, 127). Further, Iacobuzio-Donahue confirmed that CD55 is highly expressed in pancreatic cancer when measured by microarrays (128). Two previous studies that aimed to discover protein markers of mucinous and nonmucinous cysts selected AMY2A as a biomarker, consistent with the expression patterns in our study (17, 85). In addition, AMY2A was expressed at higher levels in nontumor versus PDAC tissues (129). According to the label-free quantification data of another group, CPB1, a member of the carboxypeptidase family, was confirmed to be downregulated in PDAC tissue (130). The biological functions and expression patterns of our potential markers are consistent with previous studies.

Our data also showed that most of the 20 upstream regulators that modulate 5 of the potential markers were tumor suppressors, pancreatic mitogens, and angiogenesis-related molecules, which are significantly related to pancreatic cancer according to previous publications (Figure 13C, Table 3) (102-107). The association between upstream proteins of the potential markers and cancer progression increases the credibility of our candidates. For instance, it is plausible that the downregulation of 3 upstream regulators of SOD2 (IGFBP7, KL, BTG2), which are potential tumor suppressors, leads to an increase in SOD2 when the malignancy of IPMN worsens.

Of the final marker candidates, CD55 was validated using 70 individual cyst fluid samples by ELISA as it evidently differed in expression between the histological groups of IPMN. In specific, CD55 had the lowest p-values between all comparative groups and the highest fold-changes in comparisons 2 (HGD versus invasive IPMN) and 3 (LGD versus invasive IPMN) (Figure 8, Figure 12, Table 2). Although previous studies have concluded that CD55 is associated with the dedifferentiation and invasiveness of tumors (126, 127), this study is the first to report CD55 as a marker of IPMN dysplasia. The statistical significance in the ELISA analysis was lower compared with the MS analysis. Further, the CD55 concentrations by ELISA were generally lower in all sample groups (Figure 8A, B). However, the low statistical power and concentrations do not diminish the value of this potential marker because the expression patterns of CD55 by ELISA were consistent with our MS data. In addition, the Western blot and IHC results for CD55 were consistent with the MS expression data, further supporting CD55 as a potential marker of IPMN dysplasia (Figure 15, Figure 16).

Several studies have examined CD55 as a potential biomarker and therapeutic target, establishing that CD55 is frequently upregulated in various cancer types and can serve as an indicator of cancer progression (131). In a preclinical study, Saygin et al. demonstrated that CD55 maintains self-renewal and cisplatin resistance in endometrioid tumors to accelerate tumor development (132). Other preclinical studies concluded that silencing CD55 enhances the therapeutic efficacy of rituximab (133) and the anti-HER2 monoclonal antibodies trastuzumab and pertuzumab (134). In this context, it is evident that CD55 has a significant function

in tumor progression, based on past studies and our findings. When considering the aforementioned preclinical studies regarding CD55, this marker has a significant potential as a reflection of the neoplastic evolution of IPMN.

We demonstrated that the CD55 expression is responsible for neutrophil infiltration by using immunohistochemical stains. This result is consistent with previous studies that represent an association between tumor-associated neutrophils and advanced IPMN lesions (*135, 136*). In Sadot et al., MPO, a neutrophil marker, is highly expressed in IPMN with a higher grade of dysplasia and represented that the neutrophil is responsible for the malignant progression of IPMN dysplasia (*136*). These results raise the possibility of CD55 protein as a marker in association with the IPMN malignancy.

In our previous study, which applied a similar MS-based approach, we reported potential markers that can differentiate the histological grades of IPMN using only cyst fluid from IPMN patients (90). However, the exclusion of other PCLs, such as MCNs and SCNs, is not applicable to an actual clinical environment. To compensate for this limitation, cyst fluid from other PCLs (MCN and SCN) as well as IPMN were included to increase the likelihood of discovering more clinically relevant marker candidates. Further, we increased the size of our validation cohort by 4-fold and measured protein levels by ELISA instead of Western blot, due to the higher sensitivity and specificity of the former. In contrast to Western blot, which cannot distinguish proteins with similar molecular weights, ELISA is highly specific to its target epitope and consequently generates credible quantitative concentrations (*137*).

However, several challenges remain to be addressed. Our study was limited by its simplistic design (single-center). Using a larger cohort from multiple centers would decrease the potential bias that might have been unique to the cohort in this study. Thus, it would be necessary to examine the clinical potential of CD55 through randomized, controlled, multicenter validation in a large cohort. Although 5 marker candidates, including CD55, were statistically significant ($P < 0.05$) in the univariate analysis for predicting the neoplastic evolution of IPMN, no significant covariates were identified in the multivariate analysis. One explanation is the relatively small sample size used in the study, which comprised 10 low-risk IPMN and 10 high-risk IPMN samples. Thus, further validation with more samples is necessary for obtaining more reliable results. Another limitation is that all of the cyst fluid in this study was collected from the tissue of patients who were undergoing resection rather than EUS-guided aspiration, which is a closer representation of clinical practice. Thus, the evaluation of CD55 using preoperative EUS-guided cyst fluid is a natural next step in validating the diagnostic value of CD55 as a marker for IPMN dysplasia.

In summary, we have generated the largest proteomic dataset of pancreatic cyst fluid to date and discovered potential markers of IPMN dysplasia using cyst fluid from 3 major types of PCLs (IPMN, MCN, and SCN) by LC-MS/MS. We significantly increased the protein coverage of each sample with a peptide library and discovered markers from a larger pool of candidates. By bioinformatics analyses, the DEPs were associated with biological functions that were related to pancreatic cancer, malignancy, and molecular secretion. Our process for discovering potential markers of IPMN dysplasia was logically sound. The agreement in the expression

pattern of CD55 between the MS and ELISA data demonstrates that we have discovered reliable marker candidates of IPMN dysplasia. The development of cyst fluid markers can facilitate an accurate assessment of the degree of IPMN dysplasia and effectively guide surgical decision-making. Ultimately, if the developed marker is implemented in clinical practice, the accurate assessment of IPMN dysplasia will prevent unnecessary surgical resection for low-risk IPMN patients.

CHAPTER III

Clinical Application of Multiple Reaction Monitoring-Mass Spectrometry to Human Epidermal Growth Factor Receptor 2 Measurements as a Potential Diagnostic Tool for Breast Cancer Therapy

INTRODUCTION

Human epidermal growth factor receptor 2 (HER2) is a transmembrane protein that can promote the differentiation, development, and survival of cancer cells (*138, 139*). It is often overexpressed in breast cancer and correlates with a worse prognosis (*138, 140*). In targeted therapy, anti-HER2 therapy is administered to patients who overexpress HER2 in cancer cells, inhibiting its downstream signaling pathways (*141*). Thus, accurate detection of HER2 using optimal techniques is crucial for providing the appropriate care to patients.

Immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH), approved by the Food and Drug Administration, are the most widely used methods for assessing HER2 status (*142*). Although these techniques have been the gold standard for HER2 evaluation, they have limitations (*143*). The semiquantitative nature and subjectivity of IHC contribute to its high variation and cost, in association with false positive and negative results (*144, 145*). In addition, FISH, which should be used to verify equivocal HER2 cases (*146, 147*), has several disadvantages: automated slide stainers are expensive and not always readily available in routine pathology laboratories, and the additional staining is time-consuming and costly (*148, 149*). Thus, a novel technique that can accurately evaluate whether patients could benefit from HER2-targeted therapy than IHC is needed to improve the throughput and economic viability of the overall workflow (*150*).

Targeted mass spectrometry (MS)-based approaches are emerging as

alternatives to IHC, based on their high reproducibility and quantitative nature (151, 152). Multiple reaction monitoring (MRM) has been widely applied to various clinical samples (153-155), including formalin-fixed paraffin-embedded (FFPE) tissue, which has several advantages: plentiful access to vast archives of pathologically characterized clinical samples, and ability to be stored for extended periods without requiring expensive equipment (156).

Three notable studies have previously aimed to quantify HER2 expression levels using FFPE tissues and targeted MS-based approaches (153, 157, 158) to stratify HER2 status more accurately and improve the selection of patients for HER2-targeted therapy compared with conventional methods. However, their complex and time-consuming sample preparation procedures make the MRM-MS assay less practical as a clinical assay.

Comparing the expression levels of cellular target proteins by MS-based analysis is predicated on defining the number of cells analyzed per sample. Currently, the total protein or peptide concentrations in each processed sample are used for convenience (21, 159) when i) normalizing the amount of specific target proteins, ii) reducing the potential analytical variability that might originate when dramatically different protein amounts are processed, and iii) ensuring an adequate amount of analyte for reliable detection. However, current total protein and peptide measures are poor representations of cell counts, because only a portion of the protein originates from the cells of interest. In addition, the accompanying assays lengthen the overall process and generate inter-experimenter variation. Thus, we designed an alternative protocol to facilitate the clinical application of HER2 quantification by

MRM-MS assay. Because HER2 expression is exclusive to the surface of epithelial cells (160), normalizing HER2 using a factor exclusive to epithelial cells in a breast tumor can result in a more accurate stratification of HER2 status. Thus, we applied the quantitative data of an epithelial cell-specific protein as a new normalization factor for calculating HER2 expression levels in an MRM-MS assay.

However, applying the novel protocol is not feasible without addressing the discrepancy between the tumor size and the corresponding tumor content across individual FFPE tissue specimens. To address this problem, we devised a two-part solution: (1) determining the number of FFPE tissue slides, based on the cell count from a single slide, to ensure adequate amounts of protein per analysis; and (2) using the expression level of an epithelial cell-specific protein as a normalization factor when measuring HER2 expression levels.

We aimed to establish a novel MRM-MS assay to determine HER2 status, especially for ambiguous IHC results in FFPE breast cancer samples, by determining an adequate number of FFPE slides per sample to perform a reliable MS analysis and using the expression levels of an epithelial cell-specific protein as a normalization factor.

MATERIALS AND METHODS

1. Patients and tissue samples

Two hundred ten patients who underwent surgical resection after being diagnosed with invasive ductal carcinoma at Seoul National University Hospital from January 2010 to December 2017 were selected. Those who had received neoadjuvant chemotherapy were excluded. The final cohort was composed of HER2 0 (n=30), HER2 1+ (n=30), HER2 2+/FISH-negative (n=61), HER2 2+/FISH-positive (n=59), and HER2 3+ cases (n=30). Pathological and clinical data were reviewed thoroughly and obtained from the electronic medical records system. A summary of the patients and the characteristics of their breast cancer tissues is presented in Table 1. All contents of this study were approved by our institutional review board (Institutional Review Board No. 1709-037-883), and all participants provided written informed consent.

Table 1. Demographic and clinical characteristics of the study population in the MRM-MS analysis (N = 210).

Group	HER2 status				
	HER2 0 (N = 30)	HER2 1+ (N = 30)	HER2 2+/ FISH-negative (N = 61)	HER2 2+/ FISH-positive (N = 59)	HER2 3+ (N = 30)
Age (years)					
mean \pm SD	53.80 \pm 9.57	53.60 \pm 11.39	48.62 \pm 9.86	54.22 \pm 10.95	50.57 \pm 11.66
FISH status					
Negative	0	2	61	0	0
Positive	0	0	0	59	20
NA	30	28	0	0	10
Estrogen Receptor					
Negative	18	7	4	20	20
Positive	12	23	57	39	10
Progesterone Receptor					
Negative	21	7	11	30	24
Positive	9	23	50	29	6
Subtype					
HER2	0	0	0	21	20
Luminal A	10	23	53	0	0
Luminal B	3	0	4	38	10
TNBC	17	7	4	0	0
Nuclear grade					
1	0	1	0	1	0
2	5	15	38	13	5
3	25	14	22	45	25
NA	0	0	1	0	0
Histological grade					
I	1	3	6	1	0
II	4	16	37	25	8
III	25	11	17	33	22
NA	0	0	1	0	0
Tumor size					
< 2.0 cm	7	12	28	30	19
2.0 - 4.9 cm	21	17	32	27	10
\geq 5.0 cm	2	1	1	2	1

FISH, fluorescence in situ hybridization.

2. Assessment of HER2 status in breast cancer

The diagnostic algorithm for scoring HER2 was to perform IHC in all cases, supplemented by FISH in equivocal IHC cases (HER2 2+), per the 2007 and 2013 American Society of Clinical Oncology/College of American Pathologists guidelines (*142, 147*). IHC and FISH were performed on 4 μ m tissue sections. HER2 protein expression was determined in FFPE sections on a Benchmark Ultra automatic immunostaining device (Ventana Medical System, Tucson, AZ, USA) with anti-HER2/neu (4B5) rabbit monoclonal antibody (Ventana Medical System, Tucson, AZ, USA) as described (*161*). All immunohistochemical stains were reviewed by 2 expert breast pathologists (I. A. Park and H. S. Ryu). ETV6 rearrangements were identified by interphase FISH with ETV6 (TEL) split-apart dual-color probes by PathVysion assay (Abbott Molecular, Downers Grove, IL, USA), following the manufacturer's instructions. At least 100 interphase nuclei were analyzed to interpret ETV6 rearrangements (*162*).

3. Protein extraction and digestion of FFPE tissue samples

FFPE tissue sections (10 μ m thick) were cut from whole-mount breast cancer tissue FFPE blocks and placed onto glass slides. Paraffin was removed in 6 min washes with xylene (Merck KGaA, Darmstadt, Germany), followed by rehydration steps in a series of ethanol (Merck KGaA, Darmstadt, Germany) for 3 min at 100%, 100%, 85%, 70%, 50%, and 0%. After the remaining solution was blotted, the tumor tissue was manually scraped off from the glass slide and transferred into Eppendorf tubes. After the sample was incubated for 30 min at 95°C for antigen retrieval, the tissue

was homogenized and sonicated for 80 seconds (Sonics & Materials, Newtown, CT, USA) in 100 μ L lysis buffer, containing 30 mM Tris HCl, pH 8.5 and 0.5% RapiGest (Waters, Manchester, UK). The sample was incubated again for 3 h at 95°C, followed by sonication. The sample was centrifuged (16,602 g, 20°C, 30 min) to separate supernatant from debris and other solid contents, and the pellet was discarded. Protein concentration was measured by bicinchoninic acid assay (159).

Eighty microliters of each sample was incubated with 20 μ L 100 mM dithiothreitol for 1 h at 60°C. After the incubation, the sample was alkylated with 20 μ L 240 mM iodoacetamide and incubated in the dark for 1 h at room temperature. Then, the sample was digested with 60 μ L trypsin solution (a fixed amount of trypsin, 4 μ g) (Promega, Madison, WI, USA) for 4 h at 37°C (Figure 1). The digestion was completed by adding 20 μ L 30% formic acid. The sample was centrifuged (16,602 g, 1 h, 4°C) to remove insoluble chemicals, such as RapiGest byproducts.

One hundred microliters of the supernatant were transferred to a new tube and spiked with 1000 fmol of unpurified stable isotope-labeled internal standard (SIS) peptides as an internal standard, which contained isotopically labeled (^{13}C and ^{15}N) arginine or lysine (JPT, Berlin, Germany) (30% to 70% purity, according to the manufacturer) for MRM-MS assay (Figure 2). Subsequently, the remaining supernatant was transferred to another tube for tryptophan fluorescence assay (21). All incubation steps were performed on a Thermomixer C (Eppendorf, Westbury, NY, USA).

The use of purified synthetic peptides limited the practicality of developing assays for all confirmed candidate peptides. Thus, only purified synthetic peptides

for HER2 and JAM1 were used. Purified synthetic peptides (light, unlabeled forms) for HER2 and JAM1 were obtained from 21st Century Biochemicals, Inc. (Marlborough, England), to calculate the purity of the unpurified synthetic peptides (heavy, labeled forms). After synthesis, the peptides were purified by high-performance liquid chromatography (HPLC), with subsequent quantification by amino acid analysis (purity > 95%).

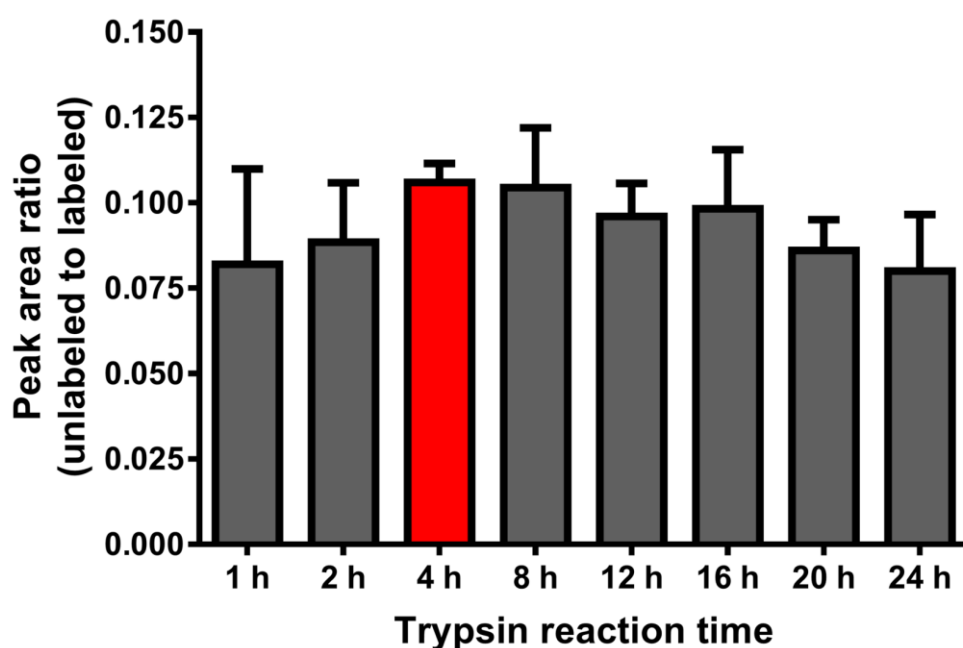


Figure 1. Determination of optimal incubation time for tryptic digestion.

With regard to reaction time, various times (1, 2, 4, 8, 12, 16, 20, and 24 h) were tested with a fixed amount of trypsin (4 μ g). Four micrograms of trypsin and a 4-h reaction time were selected as the optimal parameters for tryptic digestion. Error bars represent the standard deviation of the 3 measurements.

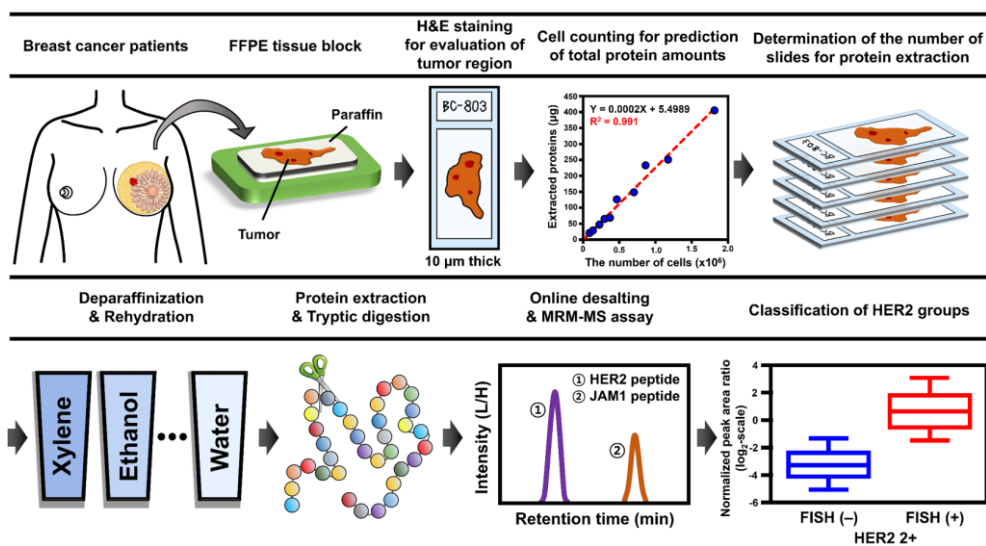


Figure 2. Schematic of overall procedure, from sample preparation to MRM-MS assay, for determining HER2 status.

Four- and 10-μm-thick formalin-fixed paraffin-embedded (FFPE) tissue sections per sample were used for hematoxylin and eosin staining and protein extraction, respectively. Hematoxylin and eosin-stained images were reviewed by pathologists to identify and assess tumor regions. After determining the number of FFPE slides based on the tumor cell counts in a single slide to ensure an adequate amount of protein, FFPE tissue sections were deparaffinized and rehydrated. The resulting tryptic peptides were analyzed by MRM-MS assay for measuring HER2 expression levels.

4. Parameters for representing tumor content

The light-to-heavy peptide peak area ratios (PARs) for HER2 surrogate peptides were used to estimate the amount of HER2 protein. Five categories of normalization factors that represent tumor content were acquired and evaluated to normalize the amount of HER2 protein, as follows: (1) tumor area (μm^2), (2) total cell count, (3) total protein amount (μg), (4) total peptide amount (μg), and (5) light-to-heavy peptide PARs for surrogate peptides from 10 epithelial cell-specific proteins and 20 housekeeping proteins, as assessed by MRM-MS analysis.

The 2 types of normalization factors—tumor area and total cell count—were estimated from hematoxylin and eosin-stained tissue sections by Aperio ImageScope, ver. 12.3.0.5056 (Leica Biosystems, Buffalo Grove, IL, USA) (163, 164). After the tumor area was annotated using a pen tool, the area and total cell count of the annotated region were measured with the nuclear counting algorithm. Morphological operators (dilation, erosion, and connected component labeling) were used to identify individual nuclei. In addition, nuclei were further discriminated from other components, based on size and shape (compactness, roundness, and elongation)—essential for automatically ignoring connective tissue, adipose tissue, and other undesired components. The third normalization factor was the total amount of proteins that were extracted from the tumor region of 10 μm thick FFPE tissue sections, the concentrations of which were determined by bicinchoninic acid assay. After tryptic digestion, the digests from FFPE tissues were estimated by tryptophan fluorescence assay to measure the total peptide amount, which was the fourth normalization factor. The fifth normalization factor was the light-to-heavy peptide

peak area ratios (PARs) for surrogate peptides of 10 epithelial cell-specific proteins and 20 housekeeping proteins, which were monitored in parallel with the 6 HER2 surrogate peptides in the final MRM-MS method. The final protein list and surrogate peptide sequences that were analyzed in the MRM-MS assay are detailed in Table 2. The following formula was used to normalize the light-to-heavy peptide PARs for HER2 surrogate peptides.

$$\frac{\text{Peak area ratio for HER2 surrogate peptide}}{\text{Normalization factor}}$$

Table 2. The final protein list and peptide sequences analyzed in the MRM-MS assay.

No.	Uniprot ID	Peptide Sequence	Precursor ion. Light (m/z)	Precursor ion. Heavy (m/z)	Precursor ion charge	Product ion. Light (m/z)	Product ion. Heavy (m/z)	Product ion charge	Product ion Type
1	ERBB2	VLQGLPR	391.7	396.8	2	570.3	580.3	1	y5
2	ERBB2	GIWIPDGENVK	614.3	618.3	2	758.4	766.4	1	y7
3	ERBB2	LLDIDETEHADGGK	559.3	561.9	3	725.3	729.3	2	y13
4	ERBB2	ELVSEFSR	483.7	488.8	2	625.3	635.3	1	y5
5	ERBB2	FVVIQNEDLGASPLDSTFYR	790.1	793.4	3	998.5	1008.5	1	y8
6	ERBB2	GLQSLPTHDPSPQR	549.3	552.6	3	574.3	579.3	2	y10
7	CALR	FVLSSGK	369.2	373.2	2	491.3	499.3	1	y5
8	CALR	FYALSASFEPFSNK	804.4	808.4	2	1113.5	1121.5	1	y10
9	COF1	NIILEEGK	458.3	462.3	2	575.3	583.3	1	y5
10	CTND1	GYELLFQPEVVR	725.4	730.4	2	874.5	884.5	1	y7
11	DDX3X	SFLDLLNATGK	646.4	650.4	2	831.5	839.5	1	y8
12	DESP	AELIVQPELK	570.3	574.3	2	713.4	721.4	1	y6
13	JAM1	VTFLPTGITFK	612.4	616.4	2	763.4	771.4	1	y7
14	G3P	GALQNIIPASTGAAK	706.4	710.4	2	815.5	823.5	1	y9
15	G3P	VPTANVSVDLTCR	510.9	514.3	3	664.3	674.3	1	y5
16	GELS	HVVPNEVVVQR	425.9	429.2	3	501.3	511.3	1	y4
17	GELS	TGAQELLR	444.3	449.3	2	530.3	540.3	1	y4

18	K1C18	STFSTNYR	488.2	493.2	2	640.3	650.3	1	y5
19	K1C18	VIDDTNITR	523.8	528.8	2	834.4	844.4	1	y7
20	K1C18	ASLENSLR	445.2	450.2	2	618.3	628.3	1	y5
21	K1C19	ILGATIENSR	537.3	542.3	2	847.4	857.4	1	y8
22	K1C19	AALEDTLAETEAR	695.3	700.4	2	1005.5	1015.5	1	y9
23	K2C7	SAYGGPVGAGIR	552.8	557.8	2	783.4	793.5	1	y9
24	K2C7	EVTINQSLLAPLR	727.4	732.4	2	1011.6	1021.6	1	y9
25	K2C7	LPDIFEAQIAGLR	481.6	484.9	3	529.3	539.4	1	y5
26	K2C8	ISSSSFRR	435.7	440.7	2	670.3	680.3	1	y6
27	K2C8	ASLEAAIADAEQR	448.9	452.2	3	618.3	628.3	1	y5
28	LDHA	VTLTSEEEAR	567.8	572.8	2	934.4	944.5	1	y8
29	NECT4	YEEELTLTR	577.3	582.3	2	861.5	871.5	1	y7
30	PDIA4	FDVSGYPTLK	563.8	567.8	2	765.4	773.4	1	y7
31	PDIA6	TGEAIVDAALSALR	693.9	698.9	2	915.5	925.5	1	y9
32	PDIA6	ELSFGR	354.7	359.7	2	477.2	477.2	1	b4
33	PPIB	IGDEDVGR	430.7	435.7	2	747.3	757.3	1	y7
34	RAB7A	VILGDSGVGK	529.3	533.3	2	619.3	627.3	1	y7
35	RAB7A	EAINVEQAFQTIAR	530.6	534.0	3	588.3	598.4	1	y5
36	RAN	FNVWDTAGQEK	647.8	651.8	2	934.4	942.4	1	y8
37	RL22	AGNLGGGVVTIER	621.8	626.8	2	887.5	897.5	1	y9
38	RL22	ITVTSEVPFSK	604.3	608.3	2	894.5	902.5	1	y8
39	RL30	SLESINSR	453.2	458.2	2	705.4	715.4	1	y6

40	RL30	LVILANNCPALR	677.4	682.4	2	894.5	902.5	1	y8
41	RL32	AAQLAIR	371.7	376.7	2	472.3	482.3	1	y4
42	RL9	TILSNQTVDIPENVDTLK	705.1	707.7	3	1028.6	1036.6	1	y9
43	RL9	GVTLGFR	375.2	380.2	2	379.2	389.2	1	y3
44	RS10	IAIYELLFK	555.3	559.3	2	925.5	933.6	1	y7
45	RS13	GLTPSQIGVILR	627.4	632.4	2	982.6	992.6	1	y9
46	RS13	LILIESR	422.3	427.3	2	617.4	627.4	1	y5
47	RS16	GGGHVAQIYAIR	414.6	417.9	3	522.3	532.3	1	y4
48	RS17	VCEEIAIPSK	629.8	633.8	2	999.6	1007.6	1	y9
49	TBB5	ISVYYNEATGGK	651.3	655.3	2	1002.5	1010.5	1	y9
50	TBB5	ALTVPELTQQVFDAK	554.0	556.6	3	638.3	642.3	2	y11
51	TERA	WALSQSNPSALR	665.3	670.4	2	959.5	969.5	1	y9
52	VINC	AIPDLTAPVAAVQAASNLVR	692.7	696.1	3	758.5	768.5	1	y7
53	VINC	AQQVSQGLDVLTAKE	486.6	489.3	3	646.4	654.4	1	y6
54	VINC	SLGEISALTSK	553.3	557.3	2	905.5	913.5	1	y9
55	VINC	ELTPQVVSAAR	585.8	590.8	2	827.5	837.5	1	y8

MRM-MS, multiple reaction monitoring-mass spectrometry; EC protein, epithelial-cell specific protein; HK protein, housekeeping protein; CE, collision energy; AuDIT, the automated detection of inaccurate and imprecise transitions; P, p-value; CV, coefficient of variation.

a Product ions in bold are the best single transition chosen for quantification.

5. Capillary liquid chromatography

Chromatographic separation was performed on a fully automated online 1260 Capillary LC system (Agilent Technologies, Santa Clara, CA, USA) prior to MRM-MS analysis. The sample vials of the autosampler were maintained at 4°C. Forty microliters of the digested sample was injected into a guard column (2.1×15.0 mm, $1.8 \mu\text{m}$, 80 \AA) (Agilent Technologies, Santa Clara, CA, USA), and online desalting was performed with the effluent directed to waste, at $40 \mu\text{L}/\text{min}$ for 10 min in 3% solvent B [0.1% formic acid/acetonitrile (v/v)] (Thermo Fisher Scientific, Waltham, MA, USA) at 40°C. After the valve was switched, the sample was transferred from the guard column to analytical column (0.5×35.0 mm, $3.5 \mu\text{m}$, 80 \AA) (Agilent Technologies, Santa Clara, CA, USA) in 3% solvent B, running at a flow rate of $40 \mu\text{L}/\text{min}$ for 5 min. Subsequently, chromatographic peptide separation was performed on the analytical column at 40°C. Bound peptides were then separated on a binary gradient from 3% to 60% solvent B at a flow rate of $40 \mu\text{L}/\text{min}$ for 40 min. The columns were washed with 60% solvent B at $40 \mu\text{L}/\text{min}$ for 2 min and then equilibrated with 3% solvent B for 1 min. The injector needle and tubing were washed with aqueous 50% acetonitrile solution after each sample was injected.

6. Mass spectrometry

MRM-MS assay was performed on an Agilent 6490 triple quadrupole (QQQ) mass spectrometer with a Jetstream electrospray source (Agilent Technologies, Santa Clara, CA, USA). The acquisition parameters were as follows: positive ion mode, gas temperature 250°C, gas flow 15 L/min, nebulizer 30 psi, sheath gas temperature

350°C, and sheath gas flow 12 L/min. The fragment voltage was adjusted to 380 V, the delta electron multiplier voltage (EMV) was set to 200 V, and the cell accelerator voltage was 5 V. The dwell time and cycle time of the mass transitions were 20 milliseconds (ms) and 2000 ms, respectively. The resolution parameters of the Q1 quadrupole and Q3 quadrupole were set to “unit”: 0.7 Da at half height.

7. Data analysis

The raw MRM-MS data files were analyzed in Skyline, ver. 4.2.0.19009 (MacCoss Lab., University of Washington, USA), which aligned the quantitative features (*165*). The relative abundance of transitions was determined by the normalized peak areas of the unpurified SIS peptides. The Savitzky-Golay method was applied to smooth the peak chromatograms (*166*). Statistical analysis (Mann–Whitney U test, area under the receiver operating curve, correlation analysis) and visualizations were performed in MedCalc, ver. 12.7 (MedCalc, Mariakerke, Belgium); IBM SPSS, ver. 19.0 (SPSS, Chicago, IL, USA); and Graph Pad Prism, ver. 6.0 (Graph PAD, San Diego, CA, USA).

All data have been deposited to Panorama Public (*167*) and ProteomeXchange ID (PXD017691) (*168*).

RESULTS

1. Tissue and patient characteristics

A total of 210 breast cancer patients were classified into 5 groups, based on HER2 status, as determined by IHC and FISH (Table 1). All 120 samples that were HER2 2+ (HER2 equivocal) by IHC were tested by FISH to determine whether they were HER2-negative or -positive. With the exception of 7 specimens, the tumor sizes of all samples were less than 5 cm; 96 specimens (45.7%) were less than 2 cm, and 107 specimens (51.0%) were between 2 and 4.9 cm.

2. Optimization of trypsin digestion

To determine the optimal conditions for the trypsin digestion, we examined the relationship between digestion efficiency and reaction time. Protein extracts from 6 HER2 3+ FFPE tissue samples were pooled and divided into 8 aliquots. The aliquots were digested for varying times (1, 2, 4, 8, 12, 16, 20, and 24 h), each with 4 μ g trypsin at 37°C. All digested samples were spiked with 1000 fmol of 6 SIS peptides of HER2 (VLQGLPR, GIWIPDGENVK, LLDIDETEHADGGK, ELVSEFSR, FVVIQNEDLGASPLDSTFYR, and GLQSLPTHDPSPQR) as an internal standard. After selecting the top 4 transitions with reproducible PARs, the average PAR was plotted versus trypsin reaction time, reflecting the efficiency of the

digestion (Figure 1). The average PAR was the highest when the trypsin digestion time was set to 4 h, which was subsequently selected as the optimal digestion time.

3. Equation for predicting protein amounts extracted from individual samples

The amounts of proteins extracted from a single tissue section unpredictably differ between individual samples because the tumor area and protein content, even in cases with the same tumor area, vary between individuals. This variation can eventually translate into errors in the quantitative MRM-MS assay. In addition, the extracted protein amounts may be insufficient for the MRM-MS assay because the number of slides required to obtain adequate initial protein amounts can be difficult to predict. Conversely, the excessive use of FFPE tissue slides can result in the unnecessary waste of samples. Thus, a guideline was needed for how many FFPE tissue sections from each sample were required to obtain the required amounts of extracted proteins prior to the sample preparation procedures.

A linear regression equation of the calibration curve was obtained by plotting the expected extracted protein amounts against the cell counts in the tumor area. The obtained equation was used as a standard guideline for estimating the protein amounts extracted from each FFPE section, based on the cell counts calculated by the Aperio software (*163, 164*). Based on the assumption that the amount of extracted protein would be proportional to the number of cells, proteins were extracted and measured from 30 FFPE slides, which comprised 3 consecutive tissue sections, 10 μm apart, from 10 individual samples. Total cell counts for each

tumor area were estimated in hematoxylin and eosin-stained tissue sections, using the nuclear counting algorithm in Aperio (163, 164). The curve of extracted proteins versus tumor cell counts was linear (Figure 3), with a regression coefficient of 0.991, showing that the quality of this calibration curve was sufficient to yield reliable results between cell counts and extracted proteins. In addition, all coefficient of variation values for the amounts of proteins that were extracted from the 3 consecutive FFPE tissue sections were < 20% in each individual, confirming that the relation between estimated protein amounts and cell counts in a tumor area was proportional.

To assess how representative the 30 slides (from 10 samples) that were used to derive a linear regression equation, shown in Figure 3, were of the entire study (579 slides from 210 samples), the sampling error was calculated and found to be 17.44% at the 95% confidence level. In addition, the major histopathological characteristics were compared between the two sample groups. A total of six histopathological characteristics [tumor area (μm^2), cell counts, fat content within the tumor (%), necrosis area (%), tumor content (%), and tumor-infiltrating lymphocytes (%)] were statistically analyzed by the Mann–Whitney U test. As a result, no significant differences were found between the entire 210-sample cohort and the 10 samples used to develop the linear regression equation, except for the necrosis area (%), which was marginally significantly different between the two sample groups ($P = 0.045$) (Figure 4). This result demonstrated that the 10 individual samples were representative of the 210 samples used in this study.

After counting the total number of cells in tumor areas of 210 individual samples using Aperio (163), we converted the estimated cell counts into protein amounts per the equation ($Y = 0.0002X + 5.4989$) generated by the calibration curve. The number of FFPE tissue sections necessary to obtain a minimum of 150 μg of proteins was determined to prepare samples for MRM-MS assay, and the required amount of protein was obtained from 87.6% of samples.

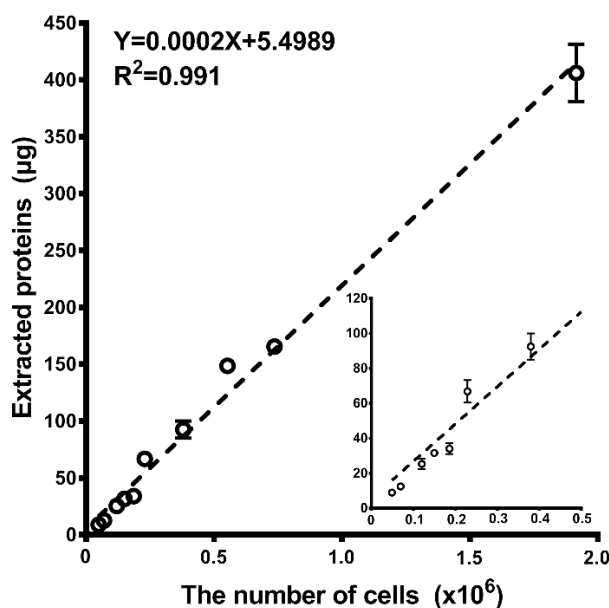


Figure 3. Calibration curve, constructed from plotting the extracted protein amount against the total cell count in the tumor area.

A total of 30 formalin-fixed paraffin-embedded (FFPE) slides from 10 individual samples were used to construct the calibration curve. Error bars represent SD values for the extracted proteins from the 3 consecutive tissue sections. The smaller curve in the inset is a magnification of the low-concentration region.

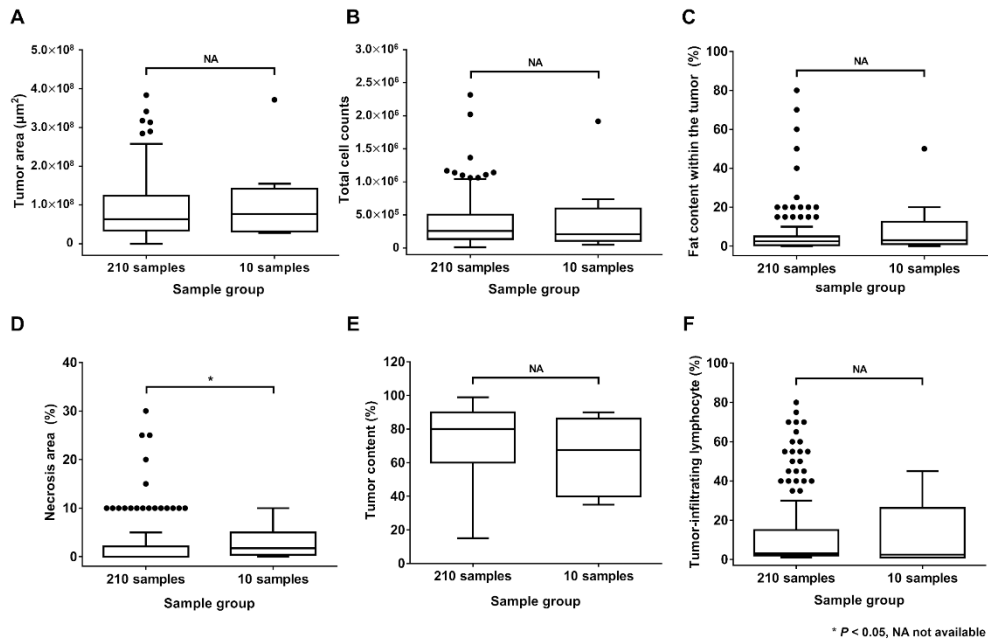


Figure 4. Comparison of six histopathological characteristics between two sample groups.

The comparison of six histopathological characteristics—tumor area (μm^2) (A), total cell counts (B), fat content within the tumor (%) (C), necrosis area (%) (D), tumor content (%) (E), and tumor-infiltrating lymphocytes (%) (F)—between the 10-sample set used to derive the linear regression equation in Fig. 2 and the 210-sample set used in the study. Significant differences were not found for any histopathological characteristics, with the exception of the percentage of necrosis ($P = 0.045$). Top bars: Mann–Whitney U test, *, $P < 0.05$, NA, Not available.

4. Target candidate selection and interference screen of peptides

A novel normalization factor was needed to measure HER2 expression levels without measuring the amounts of proteins and peptides, as done in conventional sample preparation methods. Several epithelial cell-specific proteins were selected as candidate normalization factors for HER2 expression levels because HER2 is expressed exclusively in epithelial cells (*160, 169*). Housekeeping proteins that are expressed at similar levels in all cells were also evaluated as normalization factors (*170*). In addition to the HER2 proteins, 23 housekeeping proteins and 23 epithelial cell-specific proteins were data-mined from previous reports and public databases as candidates for normalizing the amount of HER2 protein.

Including HER2, 47 proteins (123 tryptic peptides) with well-matched MS/MS libraries from the National Institute of Standards and Technology (NIST) were analyzed by mass spectrometry (*171*). To select peptides that were detected reproducibly, semiquantitative MRM-MS was performed on FFPE tissue samples that were pooled from 10 patients with HER2 3+ cases. This sample was analyzed in triplicate to select MS-detectable peptides. As a result, 37 proteins and 62 tryptic peptides were selected as detectable peptides for further analysis.

In total, 62 unpurified SIS peptides that corresponded to the endogenous peptides of 37 proteins were synthesized with stable isotope-labeled lysine or arginine residues at their C-termini (SIS peptides). The SIS peptide mixture was spiked into a pool of the fully digested FFPE samples (10 HER2 3+), and the resulting peptide mixture was analyzed in triplicate. The automated detection of

inaccurate and imprecise transitions (AuDIT) algorithm was used to eliminate falsely detected ions to select 3 transitions that had high and reproducible intensity for each tryptic peptide (172). Transitions with $P > 0.05$, between the product ion intensities of endogenous peptides and SIS peptides, and constant peak area without regard to the repeated analysis (coefficient of variation [CV] < 0.2), were selected as final targets. Among the 62 tryptic peptides that were surveyed, 55 surrogate peptides, corresponding to 31 proteins, were confirmed as interference-free transitions in a pool of the fully digested FFPE samples of HER2-positive breast cancers (Table 2).

5. MRM-MS assay development

Reverse calibration curves were generated for an FFPE tissue sample that was pooled from 18 patients with HER2 3+ cases to confirm the suitability of the peptide for the MRM-MS assay and to determine the limit of detection (LOD), the limit of quantification (LOQ), the lower limit of quantification (LLOQ), and the upper limit of quantification (ULOQ) for each MRM-MS assay. The LOD and LOQ were determined by adding 3 and 10 times the standard deviation (SD), respectively, to the mean values of the zero samples. The LLOQ and ULOQ were defined as the lowest and highest concentrations that met the precision (CV $< 20\%$) and signal-to-noise (S/N > 5) criteria, respectively.

The concentration points of the reverse calibration curves were established to cover 3 orders of magnitude by making 2-fold serial dilutions of the unpurified SIS peptide (heavy form) mixture. Each calibration point was spiked into the matrix,

which was composed of digested FFPE tissue samples. All calibration points were analyzed in technical triplicates, which demonstrated high precision in the assay with 3 transitions (Figure 5).

Overall, the linear responses spanned 3 orders of magnitude, from 0.647 (LLOQ) to 1325 (ULOQ) fmol and from 2.587 (LLOQ) to 1325 (ULOQ) fmol, for the HER2 and junctional adhesion molecule A (JAM1) surrogate peptides, respectively. The calibration points in Figure 5 were adjusted based on the purity of the unpurified SIS peptides (Figure 6). The 3 replicates demonstrated high precision in the assay with the 3 transitions when measuring both HER2 and JAM1 surrogate peptides, with minimal interference from endogenous peptides (Figure 5A, B). The PARs of the SIS peptide and the endogenous peptide were plotted against their concentrations to generate a calibration curve (Figure 5C, D). For the HER2 MRM-MS assay, the LOD was 0.399 fmol and the LOQ was 0.985 fmol, with a linear regression value of $R^2 = 0.999$ ($Y = 1.029X - 4.732$) for the assay calibration curve. CVs were in the range of 4.02% to 19.21%, as analyzed in triplicate. The calibration curve for the JAM1 surrogate peptide had the following parameters: LOD of 2.212 fmol, LOQ of 5.043 fmol, and linear regression value of $R^2 = 0.998$ ($Y = 1.113X - 3.772$). CVs ranged from 4.91% to 23.45% in technical triplicates of each concentration point.

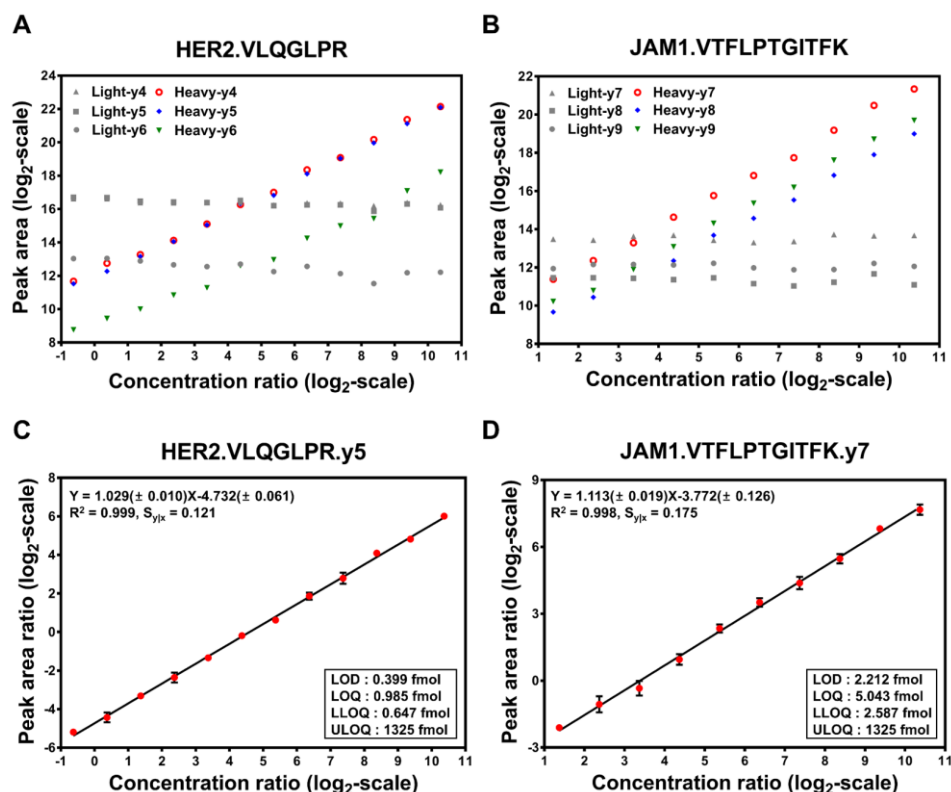


Figure 5. Evaluation of calibration curves.

The linearity of the MRM-MS assay was determined by spiking stable isotope-labeled internal standard peptides into FFPE tissue lysates at various concentrations (0.647–1,325 fmol for HER2, 2.587–1,325 fmol for JAM1). Logarithmic plots of the peak areas, calculated from the average of three replicates, for all product ions of the HER2 peptide (VLQGLPR) (A) and JAM1 peptide (VTFLPTGITFK) (B). Reverse calibration curves for the HER2 peptide (VLQGLPR) (C) and JAM1 peptide (VTFLPTGITFK) (D) were constructed with 12 and 10 concentration points, respectively. LOD, LOQ, LLOQ, and ULOQ are the limit of detection, the limit of quantification, the lower limit of quantification, and the upper limit of quantification, respectively.

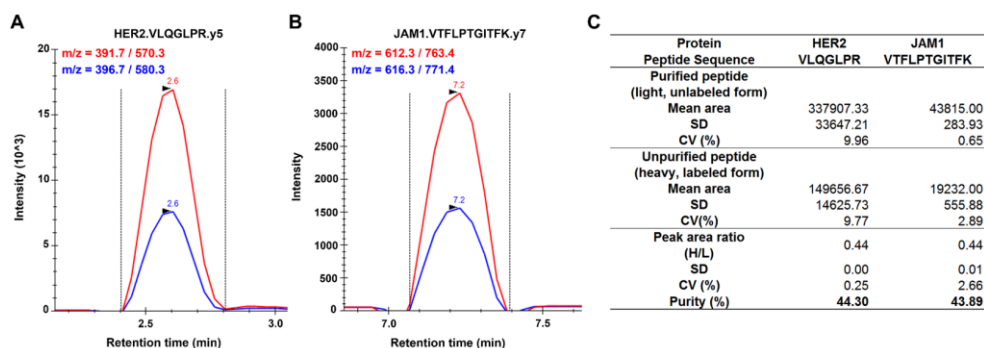


Figure 6. Measurement of purity of unpurified synthetic peptide.

Extracted ion chromatograms (XICs) for unpurified (blue line) and purified (red line) synthetic HER2 (A) and JAM1 (B) peptides, assessed by the MRM-MS assay. The spiking concentration used for synthetic peptides (unpurified and purified) was 200 fmol, and peptides were analyzed 3 times by MRM-MS assay. MRM-MS traces show the coelution of unpurified and purified synthetic peptides. The peak area ratio (unpurified/purified) allowed the purity of unpurified synthetic HER2 and JAM1 peptides to be calculated (C).

6. Calculating the purity of the unpurified synthetic peptides

Two types of synthetic peptides were used to calculate the purity of the unpurified HER2 and JAM1 synthetic peptides: (1) unpurified synthetic peptides (heavy, labeled forms) and (2) purified synthetic peptides (light, unlabeled forms). We calculated the purity of the unpurified synthetic peptides by synthesizing purified synthetic HER2 and JAM1 peptides. The purity of the purified synthetic peptides exceeded 95%, according to the amino acid analysis (AAA) conducted by the manufacturer. The purified and unpurified synthetic peptides were then injected in equal amounts (200 fmol) and analyzed 3 times using the MRM-MS assay, to confirm the purity of the unpurified synthetic peptides, using the PARs of unpurified and purified synthetic peptides.

Figure 6 shows the coelution of the transitions, from the unpurified and purified synthetic peptides. The PAR (unpurified:purified) was calculated to determine the purity of the unpurified synthetic peptide. The purity of the unpurified synthetic peptide VLQGLPR (isotope-labeled form), derived from HER2, was 44.30%. The purity of the unpurified synthetic peptide VTFLPTGITFK, derived from JAM1, was 43.89%.

7. Evaluating the stability and reproducibility of the assay

The stability and reproducibility characteristics of HER2 and JAM1 surrogate peptides were determined according to Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines (173) (Figure 7); information regarding these

transitions is listed in Table 2. The stability evaluates the storage conditions of target peptides in the fully digested FFPE samples: (1) 0, 6, and 24 h at 4°C; (2) 1 freeze-thaw cycle; (3) 2 freeze-thaw cycles; and (4) 4 weeks at -80°C. The reproducibility measures the reproducibility of a quantified endogenous analyte in the overall process to evaluate the reproducibility of the entire procedure for the MRM-MS assay.

Stability studies were conducted with 2 types of quality control (QC) samples, under various storage conditions. A pool of the fully digested FFPE samples was used as a matrix to prepare 12 aliquots for each of the low- and the medium-QC samples. To assess short-term storage stability, three aliquots of each QC sample (the 1st batch) were stored in an autosampler at 4°C and analyzed after 0, 6, and 24 h. Long-term storage stability was assessed by analyzing each QC sample after storing in a freezer at -80°C. Six of each QC sample were thawed, and 3 were immediately analyzed (the 2nd batch, 1 freeze-thaw cycle). The remaining 3 thawed aliquots were analyzed after an additional freeze-thaw cycle (the 3rd batch, 2 freeze-thaw cycles). The last 3 aliquots of each QC sample were thawed after 4 weeks of storage at -80°C and analyzed (the 4th batch). All analyses were repeated in duplicates.

The variability among the PARs for HER2 and JAM1 peptides in measurements of low- and medium-QC samples was similar under the 6 conditions (Figure 7A). The mean CV values for the HER2 and JAM1 peptide PARs across all batches were less than 7% in both low- and medium-QC samples. The PARs for

HER2 and JAM1 peptides at time zero (1st batch, 0 h) were compared with those of the other 5 conditions. The deviations from baseline values were within the acceptable limit ($< \pm 20\%$) in both low- and medium-QC samples (Figure 7B).

The reproducibility of the entire MRM-MS assay workflow was evaluated by preparing and analyzing individual FFPE slides from the same sample on different days. Each FFPE specimen was analyzed in duplicate on each of 5 days (10 analyses in total) by taking a fresh FFPE slide each day throughout the entire workflow.

The CV for each peptide was calculated, based on the mean PAR of duplicates of each day across the 5-d period. The mean CVs of the two surrogate peptides over the 5-d period were less than 13% for all samples. In addition, the mean CV values for the light-to-heavy peptide PARs for the HER2 surrogate peptide (VLQGLPR), normalized by those for the JAM1 surrogate peptide (VTFLPTGITFK), over the 5-d period were within 11% for all samples. This result demonstrated that the overall MRM-MS assay workflow was stable when using FFPE tissue specimens over several days (Figure 7C–E).

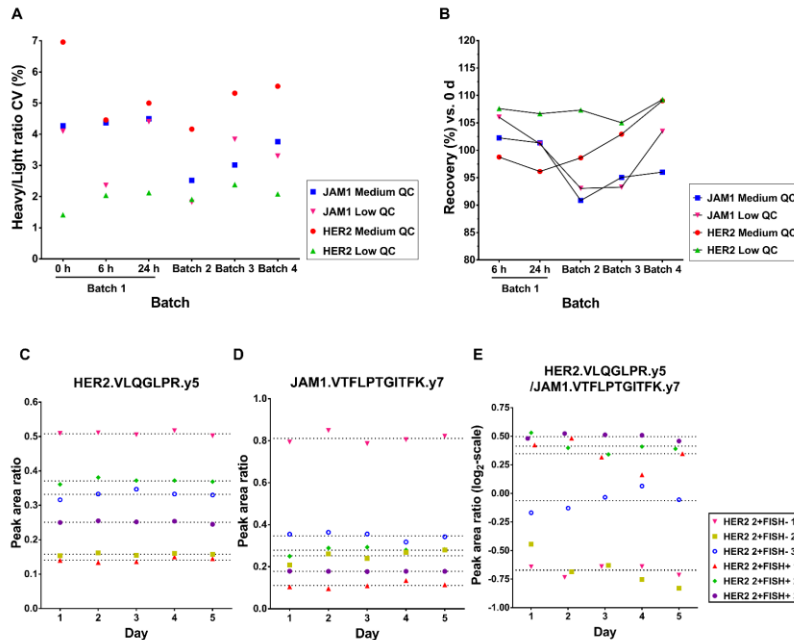


Figure 7. Evaluation of stability and reproducibility of HER2 and JAM1 surrogate peptides.

The stability of the HER2 (VLQGLPR) and JAM1 (VTFLPTGITFK) surrogate peptides were determined under six different storage conditions. The mean coefficient of variation (CV) values for the PARs for HER2 and JAM1 surrogate peptides in low- and medium-QC samples, for all batches, are shown ($CV < 7\%$) (A). The variability of PARs for the two target protein surrogate peptides, compared with the baseline values of time zero (B). The reproducibility of HER2 (C), JAM1 (D) surrogate peptides, and the light-to-heavy peptide peak area ratio for the HER2 surrogate peptide, normalized by that for the JAM1 surrogate peptide (E), was evaluated using 6 FFPE tissue specimens (3 HER2 2+/FISH- and 3 HER2 2+/FISH+). Injections were performed 2 times per day. The black dashed line signifies the mean value over 5-d period.

8. Measurement of surrogate peptides for estimation of HER2 expression levels

The workflow for determining HER2 status by the MRM-MS assay is depicted in Figure 2. The reviewed hematoxylin and eosin images, which are marked with bold lines that delineate the tumor area, were used to estimate the total cell count for the tumor area using the nuclear counting algorithm in Aperio (163). The number of slides necessary to extract a minimum of 150 μg of proteins per sample was calculated from the calibration curve in Figure 3. A total of 55 surrogate peptides—6 HER2 peptides and 49 candidate peptides for normalization (19 surrogate peptides from 10 epithelial cell-specific proteins and 30 surrogate peptides from 20 housekeeping proteins)—were quantified by MRM-MS analysis to determine HER2 status. The entire list of proteins and surrogate peptide sequences for the MRM-MS assay is detailed in Table 2. By Spearman rank correlation analysis, all 6 HER2 surrogate peptides correlated positively and significantly with each other (Figure 8).

9. Comparison of normalization factors for determining HER2 status

Five categories of normalization factors for the normalization of the light-to-heavy peptide PARs for HER2 surrogate peptides were measured and compared: tumor area (μm^2), total cell count, total protein amount (μg), total peptide amount (μg), and light-to-heavy peptide PARs for the surrogate peptides of 10 epithelial cell-specific proteins and 20 housekeeping proteins from the MRM-MS assay. This generated 5 types of normalized quantitative data for HER2 surrogate peptides, which were then compared with IHC and FISH data on 210 individual samples.

To select the normalization factor that best represents the number of tumor cells, which in turn determines HER2 expression levels used to discriminate between equivocal HER2 subgroups, area under the receiver operating curve (AUROC) values were calculated using each normalized value of 120 HER2 2+ samples. A total of 318 AUROC values were generated when considering 53 normalization factors and 6 HER2 surrogate peptides. When the 318 combinations were arranged according to decreasing AUROC values, the combination between the light-to-heavy peptide PAR for the HER2 surrogate peptide (VLQGLPR) and that for the JAM1 surrogate peptide (VTFLPTGITFK) showed the highest AUROC value of 0.908 (95% confidence interval [CI], 0.842–0.953), followed by AUROC values of 0.810 (95% CI, 0.728–0.875) for total cell count, 0.802 (95% CI, 0.719–0.869) for total peptide amount (μg), 0.777 (95% CI, 0.692–0.848) for total protein amount (μg), and 0.771 (95% CI, 0.685–0.842) for tumor area (μm^2) (Figure 9A).

Additionally, a single HER2 surrogate peptide (VLQGLPR) was found to be superior to the average of HER2 surrogate peptides for the discrimination of equivocal HER2 subgroups. The light-to-heavy peptide PAR for the superior HER2 surrogate peptide (VLQGLPR), normalized by that for JAM1 (VTFLPTGITFK), had an area under the receiver operating curve (AUROC) value of 0.908, whereas the average light-to-heavy peptide PAR for the pair of HER2 surrogate peptides with the highest Spearman correlation (VLQGLPR and GLQSLPTHDP LQR, Spearman correlation coefficient, $\rho=0.872$) (Figure 8), normalized by that for JAM1 (VTFLPTGITFK), had an AUROC value of 0.858 (95% CI, 0.788–0.928), and the average light-to-heavy peptide PAR for all 6 HER2 surrogate peptides, normalized

by that for JAM1 (VTFLPTGITFK), had an AUROC value of 0.827 (95% CI, 0.752–0.903) (Figure 9B).

When using the light-to-heavy peptide PAR for the HER2 surrogate peptide (VLQGLPR), normalized by that for the JAM1 surrogate peptide (VTFLPTGITFK), 54 of 61 HER2 2+/FISH-negative samples (88.52%) were classified as HER2-negative by the MRM-MS assay, whereas 52 of 59 HER2 2+/FISH-positive samples (88.14%) were categorized as HER2-positive by the MRM-MS assay. In total, 106 (88.33%) of 120 samples were correctly classified by the MRM-MS assay (Table 3).

Table 3. The discriminatory power of the normalized HER2 expression levels for 120 equivocal HER2 cases.

		Gold standard (IHC/FISH)	
		Positive	Negative
MRM-MS	Positive	52	7
	Negative	7	54

Percent of cases correctly classified: 88.33%

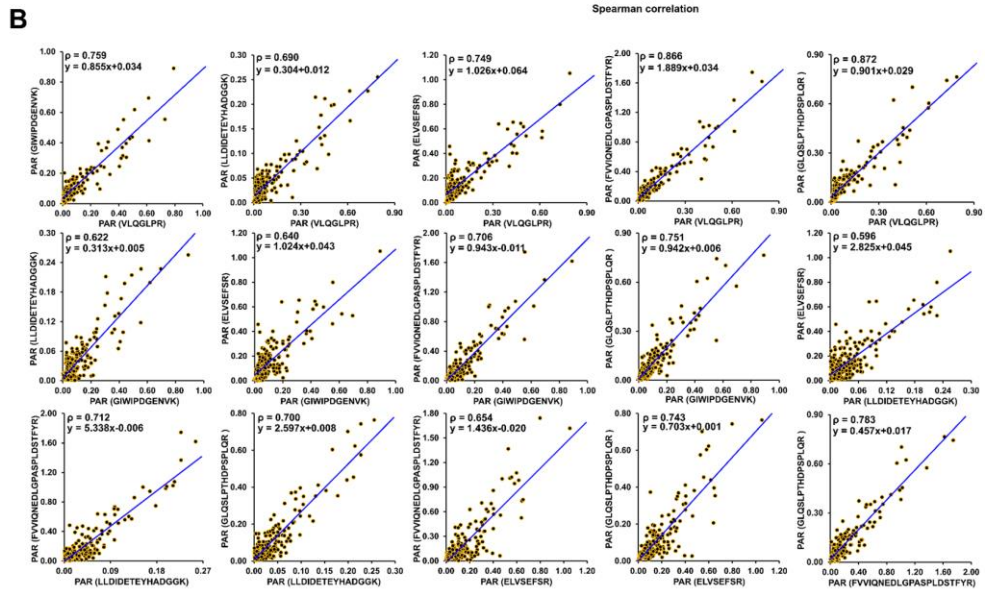
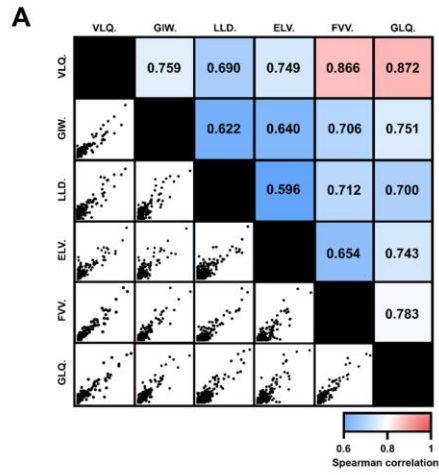


Figure 8. Scatterplots of Spearman rank correlation coefficients between six HER2 peptides.

Scatterplots of Spearman rank correlation coefficients (ρ) between six HER2 peptides, showing that all six HER2 peptides correlated positively and significantly with each other. (A) Heat maps representing the scatterplots and Spearman rank correlation coefficients between each HER2 peptide. Pink denotes greater association ($0.8 < \text{Spearman rank correlation coefficients} < 1$) between two peptides, in contrast to blue, which indicates less association ($0.5 < \text{Spearman rank correlation coefficients} < 0.8$). Amino acid sequences designations were determined by the first three amino acid sequences. (B) A total of 15 scatterplots between HER2 peptides, representing the equation of the linear regression and Spearman rank correlation coefficients. VLQ., VLQGLPR; GIW., GIWIPDGENVK; LLD., LLDIDETEHADGGK; ELV., ELVSEFSR; FVV., FVVIQNEDLGASPLDSTFYR; GLQ., GLQSLPTHDPSPQR.

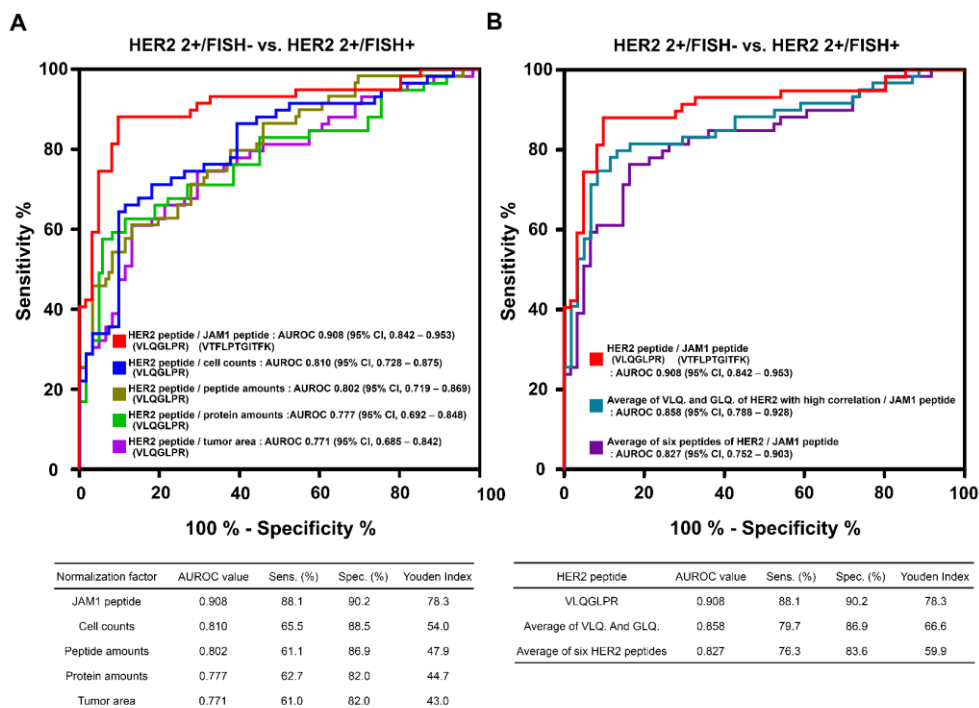


Figure 9. Area under the receiver operating curve (AUROC) with respect to normalized quantified data on HER2 peptide in HER2 2+/FISH-negative versus HER2 2+/FISH-positive.

(A) Five types of normalized quantitative data for HER2 surrogate peptides were used to generate the AUROC curve. (B) The AUROC values were calculated using the normalized values in samples from the equivocal HER2 subgroups (61 HER2 2+/FISH-negative, 59 HER2 2+/FISH-positive). A single HER2 surrogate peptide (VLQGLPR) was superior to the average of the pair of HER2 surrogate peptides with the highest Spearman correlation and to the average of all HER2 surrogate peptides. VLQ., VLQGLPR; GLQ., GLQSLPTHDPSPQLR.

10. Diagnostic performances of single- and multi-marker panels

The diagnostic performances of normalized HER2 surrogate peptide values were evaluated using classical statistical analysis. A total of 60 samples (30 HER2 2+/FISH-negative and 30 HER2 2+/FISH-positive) in the training set were randomly selected from the 120 equivocal HER2 cases. Single- and multi-marker analyses were performed by logistic regression to determine the best predictive model (174). AUROC values were generated by 5-fold cross-validation in the training set. The best cutoff value was defined as the value with the maximal sum of sensitivity and specificity (175).

When selecting marker candidates, only the AUROC value was considered when evaluating the performance of the marker. The sensitivities and specificities of the final single and multiple markers were assessed to confirm the performance of the marker and the accuracy of the diagnostic method.

In the single-marker analysis, the light-to-heavy peptide PAR for the HER2 surrogate peptide (VLQGLPR), normalized by that for the JAM1 surrogate peptide (VTFLPTGITFK), demonstrated the best diagnostic performance for both the training set (AUROC = 0.950, sensitivity = 93.3%, specificity = 93.3%) and the test set (AUROC = 0.891, sensitivity = 82.8%, specificity = 90.3%) (Table 4).

In the multi-marker analysis, the combination of three peptide pairs, i) the light-to-heavy peptide PAR for the HER2 surrogate peptide (VLQGLPR), normalized by that for the JAM1 surrogate peptide (VTFLPTGITFK), ii) the light-to-heavy peptide PAR for the HER2 surrogate peptide (ELVSEFSR), normalized by

that for the RS16 surrogate peptide (GGGHVAQIYAIR), and iii) the light-to-heavy peptide PAR for the HER2 surrogate peptide (VLQGLPR), normalized by that for the VINC surrogate peptide (SLGEISALTSK), exhibited the best diagnostic performance for both the training set (AUROC = 0.969, sensitivity = 93.3%, specificity = 93.3%) and the test set (AUROC = 0.899, sensitivity = 76.7%, specificity = 96.7%) (Table 4). To determine whether the AUROC values differed between single- and multi-marker analyses, the DeLong test was conducted in the test set. The DeLong test concluded that the single-marker analysis did not significantly differ from the multi-marker analysis ($P = 0.8432$).

Table 4. Cross-validation results of the single- and multi-marker analyses, as measured by MRM-MS assay, between the training and test sets.

Group		Single-marker analysis		Multi-marker analysis	
		Training set	Test set	Training set	Test set
HER2 2+FISH- vs. HER2 2+FISH+	AUROC	0.950	0.891	0.969	0.899
	Sensitivity	0.933	0.828	0.933	0.767
	Specificity	0.933	0.903	0.933	0.967

11. Agreement between MRM-MS data and IHC/FISH data

To determine whether the data that were generated by MRM-MS matched well with the IHC and FISH scores, the light-to-heavy peptide PARs for the HER2 surrogate peptide (VLQGLPR) that was normalized by those for the JAM1 surrogate peptide (VTFLPTGITFK) in 210 samples were examined. By Mann–Whitney U test, significant differences were found in 5 HER2 groups, and particularly the MRM-MS data distinguished between HER2 2+/FISH-negative and HER2 2+/FISH-positive groups ($P < 0.001$), which could not be differentiated by IHC (Figure 10A). The MRM-MS assay also distinguished HER2-negative from HER2-positive breast cancer, which would be expected to benefit from HER2-targeted therapy ($P < 0.001$) (Figure 10B). An AUROC analysis was conducted to further assess the ability of the MRM-MS assay to distinguish between HER2-negative and HER2-positive breast cancer. An optimal cutoff value of 0.2635 (\log_2 -scaled normalized PAR = -1.9241) was defined as the value that provided the highest levels of clinical sensitivity and specificity, as evidenced by the proximity of this value to the top left corner of the curve, correlating with an AUROC value of 0.960 (95% CI, 0.933–0.987) (Figure 10C).

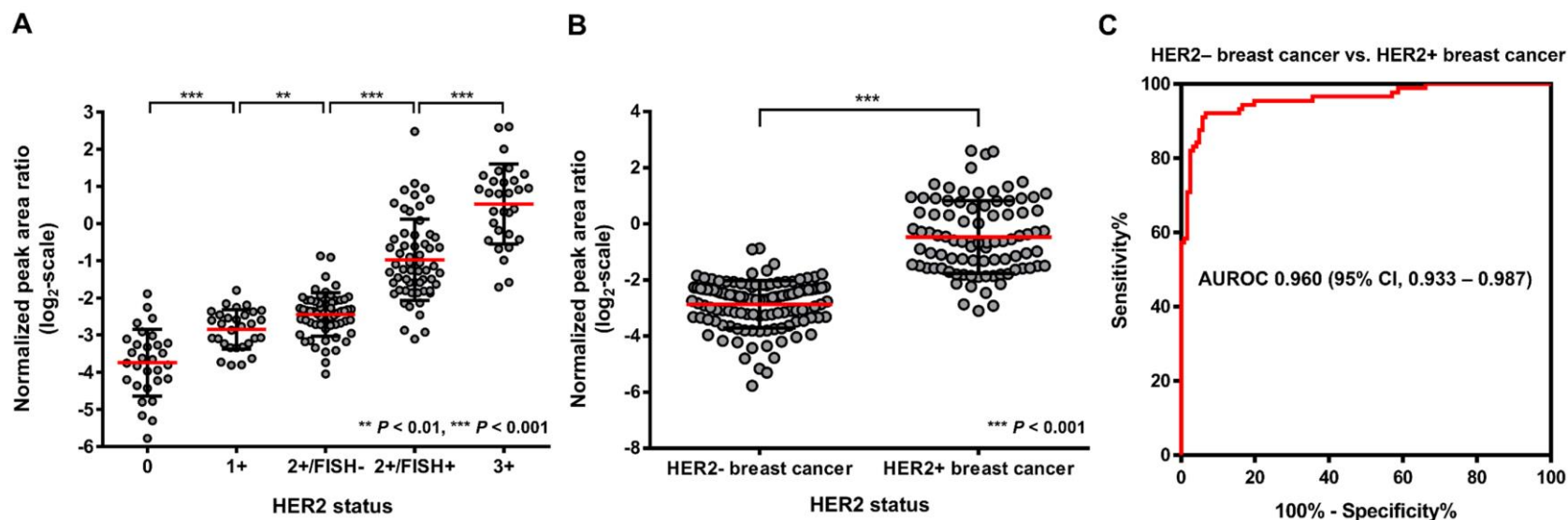


Figure 10. Scatter dot plots and an area under the receiver operating curve (AUROC) of the light-to-heavy peptide peak area ratios for a HER2 surrogate peptide normalized by those for a JAM1 surrogate peptide plotted against IHC/FISH score.

In scatter dot plots (A, B) and an AUROC curve (C), the light-to-heavy peptide peak area ratios for the HER2 surrogate peptide (VLQGLPR), normalized by those for the JAM1 surrogate peptide (VTFLPTGITFK), with the highest AUROC value used to determine whether the data generated by MRM-MS assay followed the tendencies in the IHC and FISH data. Top bars: Mann–Whitney U test; **, $P < 0.01$; ***, $P < 0.001$.

DISCUSSION

In this study, we adopted MS-based targeted proteomics, which has become the preferred method for biomarker studies of various human samples due to its high analytical sensitivity, reproducibility, accuracy, and precision (*176*), by complementing the limitations of conventional techniques for determining HER2 status: semiquantitative scores, high interobserver variability, and extra labor required by additional staining to arbitrate equivocal HER2 cases (*177, 178*). This report details the development of a clinical MRM-MS assay with FFPE breast cancer tissues that can overcome the aforementioned limitations of conventional methods by stratifying HER2 status more simply and precisely.

Two previous representative studies aimed to determine the HER2 status more accurately with FFPE tissues using a targeted MS-based approach (*158, 179*), and our study complements these studies. A more recent study analyzed 40 individual samples, which was a substantially smaller cohort than ours. Notably, the creation of an aptamer-peptide probe entailed in the previous study is complex and laborious (*179*). The complexity of the sample preparation procedures described in both studies renders them impractical for use as clinical assays. In contrast, rather than controlling the total protein and peptide amounts within the workflow by directly measuring these components, we simplified the workflow by estimating the number of slides upfront that would be required to ensure a sufficient amount of extracted protein for each analysis and by using the expression levels of an epithelial

cell-specific protein as a normalization factor for measuring HER2 expression levels. As such, our method has the potential to save time and costs as part of an overall workflow in determining HER2 status, which may accelerate the clinical adoption of similar MRM-MS assays.

We compared 5 categories of normalization factors to select the most suitable alternative to conventional normalization methods in sample preparation (21, 159). Consistent with our hypothesis, the light-to-heavy peptide PAR for a surrogate peptide of JAM1, a breast epithelial cell protein, was found to be the most suitable normalization factor for measuring HER2 expression levels, followed by total cell count, total peptide amount (μg), total protein amount (μg), and tumor area (μm^2) (Figure 9A). Tumor area had the worst performance because it merely provided indirect values of tumor content. Specifically, larger tumor sizes do not necessarily represent greater tumor content, because tumor cell densities differ substantially between samples. Similarly, aside from JAM1, all examined normalization factors reflect both tumor cells and nontumor cells (macrophages, fibroblasts, and lymphocytes), which likely lower their normalization performances.

Two notable aspects of our study differentiate it from earlier efforts to quantify HER2 expression levels by mass spectrometry: (1) simplifying the overall workflow by predicting the protein amounts that can be extracted from each FFPE specimen and by using an epithelial cell-specific protein as a normalization factor for quantifying HER2 expression levels (160); and (2) potentially reducing the number of equivocal HER2 cases, which account for 18% of all newly diagnosed breast cancers (180). As a result of the superior accuracy of MRM-MS assay relative

to IHC reducing the number of equivocal cases requiring FISH assessment, we estimate that our MRM-MS/FISH workflow would reduce the mean analytical time by 1 hour and reduce the mean cost per analysis 3.5-fold relative to an IHC/FISH workflow (180-182). When using normalized HER2 quantitative data, 106 (88.33%) of 120 equivocal HER2 cases were correctly classified (Table 3).

However, several challenges remain to be addressed. To implement our developed MRM-MS assay in clinical practice, an absolute cutoff value for normalized HER2 quantitative data should be established by using purified SIS peptides in a future study. In addition, variable cell counts and the poor correlation observed between cell counts and extracted protein amounts across all 210 samples were attributed to variations in tumor sizes and cell densities in FFPE tissues and the imprecision of the manual scraping procedure, respectively. Thus, compared with the calibration curve in Figure 3, the correlation between the cell counts and extracted protein amounts was lower in 210 samples because the proteins were extracted from multiple slides, which further compounded variations in extracted protein amounts. However, we do not believe that this poor correlation adversely affected the sample preparation procedure. The required amount of protein ($> 150 \mu\text{g}$) was obtained from most of the samples, using the linear regression equation in Figure 3. In addition, variations in the extracted protein amounts across samples were likely reduced by determining the number of FFPE slides to use for each sample, based on the equation in Figure 3, which provides a clear guideline for sample use, as opposed to previous subjective decisions. A possible solution is to increase the precision in excising the tumor area using laser-capture microdissection, which

would likely improve the correlation between the extracted protein amounts and the cell counts by limiting the obtained tissue to the exact tumor area and also reduce the required labor costs compared with manually scraping the FFPE slides with a scalpel (183). In addition, several post-translational modifications (PTMs) induced by formaldehyde fixation and the sample preparation may be considered to hinder the exact quantification of the target peptides (184). However, PTMs are generally low in abundance and even the most frequent modification, such as lysine methylation, accounts for only around 2% of all peptides (185, 186). Therefore, PTMs due to FFPE sample preparation has little effect on the quantitative accuracy of target peptides by MRM-MS.

In summary, our MRM-MS assay, which distinguishes between equivocal HER2 subgroups, can potentially decrease the time and costs required for the diagnosis of breast cancer patients by reducing the number of cases that require ancillary FISH tests (Figure 11). In addition, the simplified assay procedure can reduce the barriers to entry for the clinical application of the MRM-MS assay. The proposed protocol would provide clinicians with valuable diagnostic information and facilitate the proper treatment for breast cancer patients.

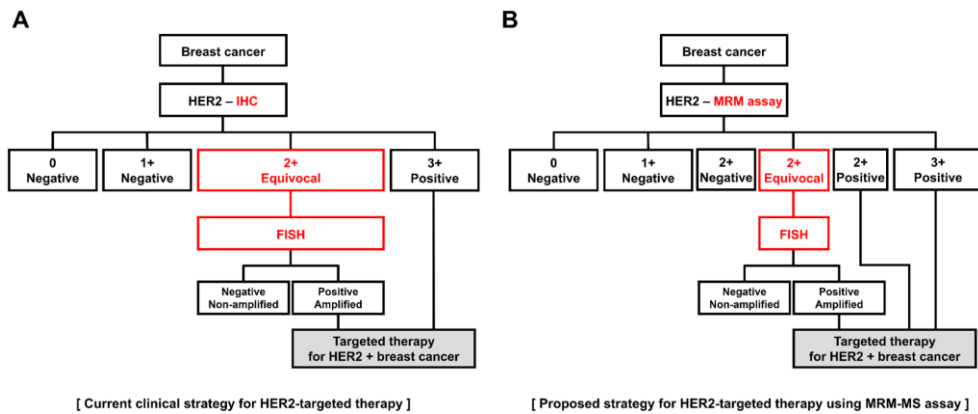


Figure 11. Conventional diagnostic strategy using immunohistochemistry (IHC) and the proposed strategy using MRM-MS assay for HER2-targeted therapy.

(A) The conventional diagnostic strategy for HER2 scoring entails performing IHC in all cases, supplemented by fluorescence in situ hybridization (FISH) in equivocal IHC cases (HER2 2+). (B) The novel application of MRM-MS assay to breast cancer patients discriminates between equivocal HER2 subgroups (HER2 2+/FISH-negative and HER2 2+/FISH-positive), reducing the number of cases that require ancillary FISH tests.

GENERAL CONCLUSION

MS-based proteomic analysis has evolved continuously and emerged as a prominent tool in the field of biomarker discovery. In addition, targeted mass spectrometry analysis is increasingly considered as an effective tool for complementing the limitations of conventional methods, such as immunoassay, for detecting protein biomarkers due to its high analytical sensitivity, reproducibility, accuracy, and precision. However, there are issues that must be solved in order for MS-based proteomics to be applied more widely in the field of biomarker discovery and clinical practice: (1) preparation methods for various pathological specimens in biomarker discovery have not been well established; (2) the sample preparation process is so cumbersome that it is difficult to be applied to clinical practice. Therefore, I intended to solve these issues in these MS-based proteomic studies. Specifically, I have established a novel preparation method for rare clinical specimen, pancreatic cyst fluid, in biomarker discovery and simplified the sample preparation procedures for MRM-MS assay.

In chapter I and II, pancreatic cyst fluid was used to discover potential biomarkers for IPMN dysplasia. Cyst fluid is composed of secreted proteins from surrounding tumor cells, the concentration of potential biomarkers in cyst fluid is higher than in blood. Therefore, pancreatic cyst fluid has several advantages over serum and plasma with regard to the discovery of markers for IPMN. However, biomarker discovery using cyst fluid has not been studied much because relatively low protein concentration and high mucus make cyst fluid difficult to handle. To

overcome the inherent characteristics of cyst fluid and discover potential biomarkers in it, I have removed the mucus by sonicating viscous cyst fluid and increased the depth of proteome coverage by generating a peptide library through high-pH fractionation. In this series of the sample preparation process, the largest number of cyst fluid proteins have been identified to date and potential markers of the degree of IPMN malignancy have been discovered. Especially, in chapter II, more clinically relevant biomarker candidates of IPMN malignancy were discovered by adding pancreatic cyst fluid from other pancreatic cystic lesions (MCN and SCN) as well as IPMN.

In chapter III, I devised a simplified assay procedure to yield more accurate HER2 expression levels relative to immunohistochemistry and reduce the barriers to entry for the clinical application of the MRM-MS assay. I applied the quantitative data of an epithelial cell-specific protein as a new normalization factor for calculating HER2 expression levels instead of total protein or peptide amounts measured by BCA assay and tryptophan fluorescence assay. I could shorten the overall process and decrease inter-experimental variation by removing accompanying assays in the newly devised procedure. As a result, a novel MRM-MS assay distinguished between equivocal HER2 subgroups that could not be differentiated by IHC. This can potentially decrease the time and costs required for the diagnosis of breast cancer patients by reducing the number of cases that require ancillary FISH tests.

In these studies, two approaches to biomarker studies through MS-based proteomics were conducted: (1) discovering novel biomarkers of IPMN malignancy

using cyst fluid with high-resolution mass spectrometry, (2) establishing a novel MRM-MS assay to determine HER2 status using FFPE tissue specimens of breast cancer patients. I have generated the largest proteomic dataset of pancreatic cyst fluid to date and discovered potential markers of the degree of IPMN dysplasia. The development of cyst fluid markers can facilitate an accurate assessment of the IPMN malignancy and effectively guide surgical decision-making. In addition, the newly established MRM-MS assay yields more accurate HER2 expression levels relative to immunohistochemistry and should help to guide clinicians toward the proper treatment for breast cancer patients, based on their HER2 expression. However, several challenges remain to be addressed. It would be necessary to examine the clinical potential of IPMN markers through randomized, controlled, multicenter validation in a large cohort. In addition, more simple and automatic sample preparation procedures should be established by excising the tumor area using laser-capture microdissection to implement our developed MRM-MS assay in clinical practice.

Despite these difficulties, a series of proteomic analyses demonstrated that MS-based proteomics remains one of the most powerful tools to discover potential markers of specific diseases and assess known biomarkers, based on their high reproducibility and quantitative nature.

REFERENCES

1. Larghi A, Panic N, Capurso G, Leoncini E, Arzani D, Salvia R, et al. Prevalence and risk factors of extrapancreatic malignancies in a large cohort of patients with intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Ann Oncol* 2013;24:1907-11.
2. Tanaka M, Fernandez-del Castillo C, Adsay V, Chari S, Falconi M, Jang JY, et al. International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. *Pancreatology* 2012;12:183-97.
3. Ivry SL, Sharib JM, Dominguez DA, Roy N, Hatcher SE, Yip-Schneider MT, et al. Global protease activity profiling provides differential diagnosis of pancreatic cysts. *Clin Cancer Res* 2017;23:4865-74.
4. Carr RA, Yip-Schneider MT, Dolejs S, Hancock BA, Wu H, Radovich M, Schmidt CM. Pancreatic cyst fluid vascular endothelial growth factor a and carcinoembryonic antigen: A highly accurate test for the diagnosis of serous cystic neoplasm. *J Am Coll Surg* 2017.
5. Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR, et al. Recurrent gnas mutations define an unexpected pathway for pancreatic cyst development. *Sci Transl Med* 2011;3:92ra66.
6. Park WG, Mascarenhas R, Palaez-Luna M, Smyrk TC, O'Kane D, Clain JE, et al. Diagnostic performance of cyst fluid carcinoembryonic antigen and amylase in histologically confirmed pancreatic cysts. *Pancreas* 2011;40:42-5.
7. Kehagias D, Smyrniotis V, Kalovidouris A, Gouliamos A, Kostopanagiotou E, Vassiliou J, Vlahos L. Cystic tumors of the pancreas: Preoperative imaging, diagnosis, and treatment. *Int Surg* 2002;87:171-4.
8. Maker AV, Carrara S, Jamieson NB, Pelaez-Luna M, Lennon AM, Dal Molin M, et al. Cyst fluid biomarkers for intraductal papillary mucinous neoplasms of the pancreas: A critical review from the international expert meeting on pancreatic branch-duct-intraductal papillary mucinous neoplasms. *J Am*

- Coll Surg 2015;220:243-53.
9. Snozek CL, Mascarenhas RC, O'Kane DJ. Use of cyst fluid cea, ca19-9, and amylase for evaluation of pancreatic lesions. Clin Biochem 2009;42:1585-8.
 10. Frossard JL, Amouyal P, Amouyal G, Palazzo L, Amaris J, Soldan M, et al. Performance of endosonography-guided fine needle aspiration and biopsy in the diagnosis of pancreatic cystic lesions. Am J Gastroenterol 2003;98:1516-24.
 11. Furukawa T, Kuboki Y, Tanji E, Yoshida S, Hatori T, Yamamoto M, et al. Whole-exome sequencing uncovers frequent gnas mutations in intraductal papillary mucinous neoplasms of the pancreas. Sci Rep 2011;1:161.
 12. Yip-Schneider MT, Wu H, Dumas RP, Hancock BA, Agaram N, Radovich M, Schmidt CM. Vascular endothelial growth factor, a novel and highly accurate pancreatic fluid biomarker for serous pancreatic cysts. J Am Coll Surg 2014;218:608-17.
 13. Poersch A, Sousa LO, Greene LJ, Faça VM, Reis FJCD. Proteomic analysis of ovarian cancer tumor fluid is a rich source of potential biomarkers. J Proteome Bioinform 2014;S5.
 14. Hata T, Dal Molin M, Hong SM, Tamura K, Suenaga M, Yu J, et al. Predicting the grade of dysplasia of pancreatic cystic neoplasms using cyst fluid DNA methylation markers. Clin Cancer Res 2017.
 15. Das KK, Xiao H, Geng X, Fernandez-Del-Castillo C, Morales-Oyarvide V, Daglilar E, et al. Mab das-1 is specific for high-risk and malignant intraductal papillary mucinous neoplasm (IPMN). Gut 2014;63:1626-34.
 16. Cuoghi A, Farina A, Z'Graggen K, Dumonceau JM, Tomasi A, Hochstrasser DF, et al. Role of proteomics to differentiate between benign and potentially malignant pancreatic cysts. J Proteome Res 2011;10:2664-70.
 17. Gbormittah FO, Haab BB, Partyka K, Garcia-Ott C, Hancapic M, Hancock WS. Characterization of glycoproteins in pancreatic cyst fluid using a high-performance multiple lectin affinity chromatography platform. J Proteome Res 2014;13:289-99.
 18. Park J, Han D, Do M, Woo J, Wang JI, Han Y, et al. Proteome

- characterization of human pancreatic cyst fluid from intraductal papillary mucinous neoplasm by liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2017;31:1761-72.
19. Wisniewski JR, Mann M. Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal Chem* 2012;84:2631-7.
 20. Han D, Moon S, Kim Y, Min H, Kim Y. Characterization of the membrane proteome and N-glycoproteome in BV2 mouse microglia by liquid chromatography-tandem mass spectrometry. *BMC Genomics* 2014;15:95.
 21. Wisniewski JR, Gaugaz FZ. Fast and sensitive total protein and peptide assays for proteomic analysis. *Anal Chem* 2015;87:4110-6.
 22. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stagetips. *Nat Protoc* 2007;2:1896-906.
 23. Han D, Jin J, Woo J, Min H, Kim Y. Proteomic analysis of mouse astrocytes and their secretome by a combination of FASP and stagetip-based, high ph, reversed-phase fractionation. *Proteomics* 2014;14:1604-9.
 24. Cox J, Mann M. Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008;26:1367-72.
 25. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007;4:207-14.
 26. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75:4646-58.
 27. Vizcaino JA, Cote RG, Csordas A, Dienes JA, Fabregat A, Foster JM, et al. The proteomics identifications (PRIDE) database and associated tools: Status in 2013. *Nucleic Acids Res* 2013;41:D1063-9.
 28. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and

- maximal peptide ratio extraction, termed maxlfq. *Mol Cell Proteomics* 2014;13:2513-26.
29. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et al. Plasma proteome database as a resource for proteomics research: 2014 update. *Nucleic Acids Res* 2014;42:D959-65.
 30. Tanaka M, Chari S, Adsay V, Fernandez-del Castillo C, Falconi M, Shimizu M, et al. International consensus guidelines for management of intraductal papillary mucinous neoplasms and mucinous cystic neoplasms of the pancreas. *Pancreatology* 2006;6:17-32.
 31. Oh JH, Wong HP, Wang X, Deasy JO. A bioinformatics filtering strategy for identifying radiation response biomarker candidates. *PLoS One* 2012;7:e38870.
 32. Sharma K, Schmitt S, Bergner CG, Tyanova S, Kannaiyan N, Manrique-Hoyos N, et al. Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci* 2015;18:1819-31.
 33. Azimifar SB, Nagaraj N, Cox J, Mann M. Cell-type-resolved quantitative proteomics of murine liver. *Cell Metab* 2014;20:1076-87.
 34. Aasebo E, Opsahl JA, Bjorlykke Y, Myhr KM, Kroksveen AC, Berven FS. Effects of blood contamination and the rostro-caudal gradient on the human cerebrospinal fluid proteome. *PLoS One* 2014;9:e90429.
 35. Chung YT, Matkowskyj KA, Li H, Bai H, Zhang W, Tsao MS, et al. Overexpression and oncogenic function of aldo-keto reductase family 1b10 (AKR1B10) in pancreatic carcinoma. *Mod Pathol* 2012;25:758-66.
 36. Zhu J, Thakolwiboon S, Liu X, Zhang M, Lubman DM. Overexpression of cd90 (THY-1) in pancreatic adenocarcinoma present in the tumor microenvironment. *PLoS One* 2014;9:e115507.
 37. Hosokawa M, Kashiwaya K, Eguchi H, Ohigashi H, Ishikawa O, Furihata M, et al. Over-expression of cysteine proteinase inhibitor cystatin 6 promotes pancreatic cancer growth. *Cancer Sci* 2008;99:1626-32.
 38. Jinfeng M, Kimura W, Hirai I, Sakurai F, Moriya T, Mizutani M. Expression of MUC5AC and MUC6 in invasive ductal carcinoma of the pancreas and

- relationship with prognosis. *Int J Gastrointest Cancer* 2003;34:9-18.
39. Michiko Horinouchi KN, Akiko Nakamura, Masamichi Goto, Sonshin Takao, Michiie Sakamoto, Noriyoshi Fukushima, Atsuo Miwa, Tatsuro Irimura, Kohzoh Imai, Eiichi Sato, Suguru Yonezawa. Expression of different glycoforms of membrane mucin (MUC1) and secretory mucin (MUC2, MUC5AC and MUC6) in pancreatic neoplasms. *Acta Histochem Cytochem* 2003;36:443-53.
 40. Arumugam T, Brandt W, Ramachandran V, Moore TT, Wang H, May FE, et al. Trefoil factor 1 stimulates both pancreatic cancer and stellate cells and increases metastasis. *Pancreas* 2011;40:815-22.
 41. Huang T, Jiang SW, Qin L, Senkowski C, Lyle C, Terry K, et al. Expression and diagnostic value of HE4 in pancreatic adenocarcinoma. *Int J Mol Sci* 2015;16:2956-70.
 42. Sanada Y, Hirose Y, Osada S, Tanaka Y, Takahashi T, Yamaguchi K, Yoshida K. Immunohistochemical study of claudin 18 involvement in intestinal differentiation during the progression of intraductal papillary mucinous neoplasm. *Anticancer Res* 2010;30:2995-3003.
 43. Shinkai K, Nakano K, Cui L, Mizuuchi Y, Onishi H, Oda Y, et al. Nuclear expression of Y-box binding protein-1 is associated with poor prognosis in patients with pancreatic cancer and its knockdown inhibits tumor growth and metastasis in mice tumor models. *Int J Cancer* 2016;139:433-45.
 44. Takao S, Takebayashi Y, Che X, Shinchi H, Natsugoe S, Miyadera K, et al. Expression of thymidine phosphorylase is associated with a poor prognosis in patients with ductal adenocarcinoma of the pancreas. *Clin Cancer Res* 1998;4:1619-24.
 45. Tanaka M, Shibahara J, Fukushima N, Shinozaki A, Umeda M, Ishikawa S, et al. Claudin-18 is an early-stage marker of pancreatic carcinogenesis. *J Histochem Cytochem* 2011;59:942-52.
 46. Douziech N, Calvo E, Coulombe Z, Muradia G, Bastien J, Aubin RA, et al. Inhibitory and stimulatory effects of somatostatin on two human pancreatic cancer cell lines: A primary role for tyrosine phosphatase SHP-1.

Endocrinology 1999;140:765-77.

47. Yonezawa S, Higashi M, Yamada N, Goto M. Precursor lesions of pancreatic cancer. *Gut Liver* 2008;2:137-54.
48. Sarmiento N, Sanchez-Bernal C, Ayra M, Perez N, Hernandez-Hernandez A, Calvo JJ, Sanchez-Yague J. Changes in the expression and dynamics of SHP-1 and SHP-2 during cerulein-induced acute pancreatitis in rats. *Biochim Biophys Acta* 2008;1782:271-9.
49. Ferreira AR, Bettencourt M, Alho I, Costa AL, Sousa AR, Mansinho A, et al. Serum YB-1 (Y-box binding protein 1) as a biomarker of bone disease progression in patients with breast cancer and bone metastases. *J Bone Oncol* 2017;6:16-21.
50. O'Brien TS, Fox SB, Dickinson AJ, Turley H, Westwood M, Moghaddam A, et al. Expression of the angiogenic factor thymidine phosphorylase/platelet-derived endothelial cell growth factor in primary bladder cancers. *Cancer Res* 1996;56:4799-804.
51. Sopha SC, Gopal P, Merchant NB, Revetta FL, Gold DV, Washington K, Shi C. Diagnostic and therapeutic implications of a novel immunohistochemical panel detecting duodenal mucosal invasion by pancreatic ductal adenocarcinoma. *Int J Clin Exp Pathol* 2013;6:2476-86.
52. Radon TP, Massat NJ, Jones R, Alrawashdeh W, Dumartin L, Ennis D, et al. Identification of a three-biomarker panel in urine for early detection of pancreatic adenocarcinoma. *Clin Cancer Res* 2015;21:3512-21.
53. Li J, Chen H, Curcuro JR, Patel S, Johns TO, Patel D, et al. Serum HE4 level as a biomarker to predict the recurrence of gynecologic cancers. *Curr Drug Targets* 2017;18:1158-64.
54. Hellstrom I, Raycraft J, Hayden-Ledbetter M, Ledbetter JA, Schummer M, McIntosh M, et al. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res* 2003;63:3695-700.
55. Ishibashi Y, Ohtsu H, Ikemura M, Kikuchi Y, Niwa T, Nishioka K, et al. Serum TFF1 and TFF 3 but not TFF 2 are higher in women with breast cancer than in women without breast cancer. *Sci Rep* 2017;7:4846.

56. Zhang DH, Yang ZL, Zhou EX, Miao XY, Zou Q, Li JH, et al. Overexpression of THY1 and ITGA6 is associated with invasion, metastasis and poor prognosis in human gallbladder carcinoma. *Oncol Lett* 2016;12:5136-44.
57. Yoshitake H, Takahashi M, Ishikawa H, Nojima M, Iwanari H, Watanabe A, et al. Aldo-keto reductase family 1, member B10 in uterine carcinomas: A potential risk factor of recurrence after surgical therapy in cervical cancer. *Int J Gynecol Cancer* 2007;17:1300-6.
58. Blackberg M, Berling R, Ohlsson K. Tissue kallikrein in severe acute pancreatitis in patients treated with high-dose intraperitoneal aprotinin. *Pancreas* 1999;19:325-34.
59. Sun X, Zhang Q, Chen W, Hu Q, Lou Y, Fu QH, et al. HOOK1 inhibits malignancy and epithelial-mesenchymal transition in hepatocellular carcinoma. *Tumour Biol* 2017;39:1010428317711098.
60. Wakita T, Hayashi T, Nishioka J, Tamaru H, Akita N, Asanuma K, et al. Regulation of carcinoma cell invasion by protein C inhibitor whose expression is decreased in renal cell carcinoma. *Int J Cancer* 2004;108:516-23.
61. Xu YF, Ren XY, Li YQ, He QM, Tang XR, Sun Y, et al. High expression of Talin-1 is associated with poor prognosis in patients with nasopharyngeal carcinoma. *BMC Cancer* 2015;15:332.
62. Jing Y, Jia D, Wong CM, Oi-Lin Ng I, Zhang Z, Liu L, et al. SERPINA5 inhibits tumor cell migration by modulating the fibronectin-integrin beta1 signaling pathway in hepatocellular carcinoma. *Mol Oncol* 2014;8:366-77.
63. Wang Z, Liu Y, Lu L, Yang L, Yin S, Wang Y, et al. Fibrillin-1, induced by Aurora-A but inhibited by BRCA2, promotes ovarian cancer metastasis. *Oncotarget* 2015;6:6670-83.
64. He X, Zhu Z, Johnson C, Stoops J, Eaker AE, Bowen W, DeFrances MC. PIK3IP1, a negative regulator of PI3K, suppresses the development of hepatocellular carcinoma. *Cancer Res* 2008;68:5591-8.
65. Dinets A, Pernemalm M, Kjellin H, Sviatoha V, Sofiadis A, Juhlin CC, et al.

- Differential protein expression profiles of cyst fluid from papillary thyroid carcinoma and benign thyroid lesions. *PLoS One* 2015;10:e0126472.
66. Romero-Calvo I, Ocon B, Martinez-Moya P, Suarez MD, Zarzuelo A, Martinez-Augustin O, de Medina FS. Reversible ponceau staining as a loading control alternative to actin in western blots. *Anal Biochem* 2010;401:318-20.
 67. Aebersold R, Burlingame AL, Bradshaw RA. Western blots versus selected reaction monitoring assays: Time to turn the tables? *Mol Cell Proteomics* 2013;12:2381-2.
 68. Yang T, Xu F, Xu J, Fang D, Yu Y, Chen Y. Comparison of liquid chromatography-tandem mass spectrometry-based targeted proteomics and conventional analytical methods for the determination of P-glycoprotein in human breast cancer cells. *J Chromatogr B Analyt Technol Biomed Life Sci* 2013;936:18-24.
 69. Klibansky DA, Reid-Lombardo KM, Gordon SR, Gardner TB. The clinical relevance of the increasing incidence of intraductal papillary mucinous neoplasm. *Clin Gastroenterol Hepatol* 2012;10:555-8.
 70. Yoshioka M, Uchinami H, Watanabe G, Sato T, Shibata S, Kume M, et al. F-18 fluorodeoxyglucose positron emission tomography for differential diagnosis of pancreatic tumors. *Springerplus* 2015;4:154.
 71. Chang YR, Park JK, Jang JY, Kwon W, Yoon JH, Kim SW. Incidental pancreatic cystic neoplasms in an asymptomatic healthy population of 21,745 individuals: Large-scale, single-center cohort study. *Medicine (Baltimore)* 2016;95:e5535.
 72. Basturk O, Hong SM, Wood LD, Adsay NV, Albores-Saavedra J, Biankin AV, et al. A revised classification system and recommendations from the baltimore consensus meeting for neoplastic precursor lesions in the pancreas. *Am J Surg Pathol* 2015;39:1730-41.
 73. Scheiman JM, Hwang JH, Moayyedi P. American gastroenterological association technical review on the diagnosis and management of asymptomatic neoplastic pancreatic cysts. *Gastroenterology* 2015;148:824-

48 e22.

74. European Study Group on Cystic Tumours of the P. European evidence-based guidelines on pancreatic cystic neoplasms. *Gut* 2018;67:789-804.
75. Tanaka M, Fernandez-Del Castillo C, Kamisawa T, Jang JY, Levy P, Ohtsuka T, et al. Revisions of international consensus fukuoka guidelines for the management of IPMN of the pancreas. *Pancreatology* 2017;17:738-53.
76. Xu MM, Yin S, Siddiqui AA, Salem RR, Schrope B, Sethi A, et al. Comparison of the diagnostic accuracy of three current guidelines for the evaluation of asymptomatic pancreatic cystic neoplasms. *Medicine (Baltimore)* 2017;96:e7900.
77. Jang JY, Park T, Lee S, Kim Y, Lee SY, Kim SW, et al. Proposed nomogram predicting the individual risk of malignancy in the patients with branch duct type intraductal papillary mucinous neoplasms of the pancreas. *Ann Surg* 2017;266:1062-8.
78. Attiyeh MA, Fernandez-Del Castillo C, Al Efishat M, Eaton AA, Gonen M, Batts R, et al. Development and validation of a multi-institutional preoperative nomogram for predicting grade of dysplasia in intraductal papillary mucinous neoplasms (IPMNs) of the pancreas: A report from the pancreatic surgery consortium. *Ann Surg* 2018;267:157-63.
79. Bassi C, Salvia R, Molinari E, Biasutti C, Falconi M, Pederzoli P. Management of 100 consecutive cases of pancreatic serous cystadenoma: Wait for symptoms and see at imaging or vice versa? *World J Surg* 2003;27:319-23.
80. Buscaglia JM, Giday SA, Kantsevov SV, Jagannath SB, Magno P, Wolfgang CL, et al. Patient- and cyst-related factors for improved prediction of malignancy within cystic lesions of the pancreas. *Pancreatology* 2009;9:631-8.
81. Singhi AD, Nikiforova MN, Fasanella KE, McGrath KM, Pai RK, Ohori NP, et al. Preoperative gnas and kras testing in the diagnosis of pancreatic mucinous cysts. *Clin Cancer Res* 2014;20:4381-9.
82. Thornton GD, McPhail MJ, Nayagam S, Hewitt MJ, Vlavianos P, Monahan

- KJ. Endoscopic ultrasound guided fine needle aspiration for the diagnosis of pancreatic cystic neoplasms: A meta-analysis. *Pancreatology* 2013;13:48-57.
83. Woolf KM, Liang H, Sletten ZJ, Russell DK, Bonfiglio TA, Zhou Z. False-negative rate of endoscopic ultrasound-guided fine-needle aspiration for pancreatic solid and cystic lesions with matched surgical resections as the gold standard: One institution's experience. *Cancer Cytopathol* 2013;121:449-58.
 84. Hata T, Dal Molin M, Suenaga M, Yu J, Pittman M, Weiss M, et al. Cyst fluid telomerase activity predicts the histologic grade of cystic neoplasms of the pancreas. *Clin Cancer Res* 2016;22:5141-51.
 85. Park J, Yun HS, Lee KH, Lee KT, Lee JK, Lee SY. Discovery and validation of biomarkers that distinguish mucinous and nonmucinous pancreatic cysts. *Cancer Research* 2015;75:3227-35.
 86. Kristjansdottir B, Partheen K, Fung ET, Marcickiewicz J, Yip C, Brannstrom M, Sundfeldt K. Ovarian cyst fluid is a rich proteome resource for detection of new tumor biomarkers. *Clin Proteomics* 2012;9:14.
 87. Talebian M, von Bartheld MB, Braun J, Versteegh MI, Dekkers OM, Rabe KF, Annema JT. EUS-FNA in the preoperative staging of non-small cell lung cancer. *Lung Cancer* 2010;69:60-5.
 88. Tanase C, Albulescu R, Neagu M. Proteomic approaches for biomarker panels in cancer. *J Immunoass Immunoch* 2016;37:1-15.
 89. Jabbar KS, Arike L, Verbeke CS, Sadik R, Hansson GC. Highly accurate identification of cystic precursor lesions of pancreatic cancer through targeted mass spectrometry: A phase iic diagnostic study. *J Clin Oncol* 2017;36:367-75.
 90. Do M, Han D, Wang JI, Kim H, Kwon W, Han Y, et al. Quantitative proteomic analysis of pancreatic cyst fluid proteins associated with malignancy in intraductal papillary mucinous neoplasms. *Clin Proteomics* 2018;15:17.
 91. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: A peptide search engine integrated into the maxquant

- environment. *J Proteome Res* 2011;10:1794-805.
92. Tyanova S, Temu T, Cox J. The maxquant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 2016;11:2301-19.
 93. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* 2004;17:349-56.
 94. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J Mol Biol* 2001;305:567-80.
 95. Petersen TN, Brunak S, von Heijne G, Nielsen H. Signalp 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* 2011;8:785-6.
 96. Kim H, An S, Lee K, Ahn S, Park DY, Kim JH, et al. Pancreatic high-grade neuroendocrine neoplasms in the korean population: A multicenter study. *Cancer Res Treat* 2020;52:263-76.
 97. Chebib I, Yaeger K, Mino-Kenudson M, Pitman MB. The role of cytopathology and cyst fluid analysis in the preoperative diagnosis and management of pancreatic cysts >3 cm. *Cancer Cytopathol* 2014;122:804-9.
 98. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. *Nature* 2014;509:582-7.
 99. Schirle M, Heurtier MA, Kuster B. Profiling core proteomes of human cell lines by one-dimensional page and liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* 2003;2:1297-305.
 100. Ke E, Patel BB, Liu T, Li XM, Haluszka O, Hoffman JP, et al. Proteomic analyses of pancreatic cyst fluids. *Pancreas* 2009;38:e33-42.
 101. Schmidt A, Aebersold R. High-accuracy proteome maps of human body fluids. *Genome Biol* 2006;7:242.
 102. Abramovitz L, Rubinek T, Ligumsky H, Bose S, Barshack I, Avivi C, et al. KL1 internal repeat mediates klotho tumor suppressor activities and inhibits bFGF and IGF-I signaling in pancreatic cancer. *Clin Cancer Res*

- 2011;17:4254-66.
103. An W, Ben QW, Chen HT, Zheng JM, Huang L, Li GX, Li ZS. Low expression of IGFBP7 is associated with poor outcome of pancreatic ductal adenocarcinoma. *Ann Surg Oncol* 2012;19:3971-8.
 104. Liu C, Yang Z, Li D, Liu Z, Miao X, Yang L, et al. Overexpression of B2M and loss of ALK7 expression are associated with invasion, metastasis, and poor-prognosis of the pancreatic ductal adenocarcinoma. *Cancer Biomark* 2015;15:735-43.
 105. Murphy KM, Brune KA, Griffin C, Sollenberger JE, Petersen GM, Bansal R, et al. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: Deleterious BRCA2 mutations in 17%. *Cancer Res* 2002;62:3789-93.
 106. Su GH, Bansal R, Murphy KM, Montgomery E, Yeo CJ, Hruban RH, Kern SE. ACVR1B (ALK4, activin receptor type 1B) gene mutations in pancreatic carcinoma. *Proc Natl Acad Sci U S A* 2001;98:3254-7.
 107. Sakamoto H, Kimura H, Sekijima M, Matsumoto K, Arao T, Chikugo T, et al. Plasma concentrations of angiogenesis-related molecules in patients with pancreatic cancer. *Jpn J Clin Oncol* 2012;42:105-12.
 108. Furniss RCD, Low WW, Mavridou DAI, Dagley LF, Webb AI, Tate EW, Clements A. Plasma membrane profiling during enterohemorrhagic e. Coli infection reveals that the metalloprotease stce cleaves CD55 from host epithelial surfaces. *J Biol Chem* 2018;293:17188-99.
 109. Bassi C, Sarr MG, Lillemoe KD, Reber HA. Natural history of intraductal papillary mucinous neoplasms (IPMN): Current evidence and implications for management. *J Gastrointest Surg* 2008;12:645-50.
 110. Thiruvengadam N, Park WG. Systematic review of pancreatic cyst fluid biomarkers: The path forward. *Clin Transl Gastroenterol* 2015;6:e88.
 111. Greish K. Enhanced permeability and retention (EPR) effect for anticancer nanomedicine drug targeting. *Methods Mol Biol* 2010;624:25-37.
 112. Argentiero A, De Summa S, Di Fonte R, Iacobazzi RM, Porcelli L, Da Via M, et al. Gene expression comparison between the lymph node-positive and

- negative reveals a peculiar immune microenvironment signature and a theranostic role for wnt targeting in pancreatic ductal adenocarcinoma: A pilot study. *Cancers* 2019;11.
113. Porcelli L, Iacobazzi RM, Di Fonte R, Serrati S, Intini A, Solimando AG, et al. CAFs and TGF-beta signaling activation by mast cells contribute to resistance to Gemcitabine/Nabpaclitaxel in pancreatic cancer. *Cancers (Basel)* 2019;11.
 114. Hiraoka N, Yamazaki-Itoh R, Ino Y, Mizuguchi Y, Yamada T, Hirohashi S, Kanai Y. CSCL17 and ICAM2 are associated with a potential anti-tumor immune response in early intraepithelial stages of human pancreatic carcinogenesis. *Gastroenterology* 2011;140:310-21.
 115. Nasu J, Mizuno M, Uesu T, Takeuchi K, Inaba T, Ohya S, et al. Cytokine-stimulated release of decay-accelerating factor (DAF;CD55) from HT-29 human intestinal epithelial cells. *Clin Exp Immunol* 1998;113:379-85.
 116. Reis ES, Mastellos DC, Ricklin D, Mantovani A, Lambris JD. Complement in cancer: Untangling an intricate relationship. *Nat Rev Immunol* 2018;18:5-18.
 117. Wolf K, Wu YI, Liu Y, Geiger J, Tam E, Overall C, et al. Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. *Nat Cell Biol* 2007;9:893-904.
 118. Afshar-Kharghan V. The role of the complement system in cancer. *J Clin Invest* 2017;127:780-9.
 119. Falanga A, Marchetti M, Vignoli A. Coagulation and cancer: Biological and clinical aspects. *J Thromb Haemost* 2013;11:223-33.
 120. Bjorge L, Hakulinen J, Vintermyr OK, Jarva H, Jensen TS, Iversen OE, Meri S. Ascitic complement system in ovarian cancer. *Br J Cancer* 2005;92:895-905.
 121. Ytting H, Jensenius JC, Christensen IJ, Thiel S, Nielsen HJ. Increased activity of the mannan-binding lectin complement activation pathway in patients with colorectal cancer. *Scand J Gastroenterol* 2004;39:674-9.
 122. Stiles ZE, Khan S, Patton KT, Jaggi M, Behrman SW, Chauhan SC.

- Transmembrane mucin MUC13 distinguishes intraductal papillary mucinous neoplasms from non-mucinous cysts and is associated with high-risk lesions. *HPB (Oxford)* 2018;21:87-95.
123. Mito K, Saito M, Morita K, Maetani I, Sata N, Mieno M, Fukushima N. Clinicopathological and prognostic significance of MUC13 and AGR2 expression in intraductal papillary mucinous neoplasms of the pancreas. *Pancreatology* 2018;18:407-12.
 124. Kumari S, Khan S, Gupta SC, Kashyap VK, Yallapu MM, Chauhan SC, Jaggi M. MUC13 contributes to rewiring of glucose metabolism in pancreatic cancer. *Oncogenesis* 2018;7:19.
 125. Khan S, Zafar N, Khan SS, Setua S, Behrman SW, Stiles ZE, et al. Clinical significance of MUC13 in pancreatic ductal adenocarcinoma. *HPB (Oxford)* 2018;20:563-72.
 126. Wobus M, Vogel B, Schmucking E, Hamann J, Aust G. N-glycosylation of CD97 within the egf domains is crucial for epitope accessibility in normal and malignant cells as well as cd55 ligand binding. *Int J Cancer* 2004;112:815-22.
 127. Karpus ON, Veninga H, Hoek RM, Flierman D, van Buul JD, Vandenakker CC, et al. Shear stress-dependent downregulation of the adhesion-G protein-coupled receptor CD97 on circulating leukocytes upon contact with its ligand CD55. *J Immunol* 2013;190:3740-8.
 128. Iacobuzio-Donahue CA, Maitra A, Olsen M, Lowe AW, Van Heek NT, Rosty C, et al. Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. *American Journal of Pathology* 2003;162:1151-62.
 129. Lowe AW, Olsen M, Hao Y, Lee SP, Taek Lee K, Chen X, et al. Gene expression patterns in pancreatic tumors, cells and tissues. *PLoS One* 2007;2:e323.
 130. Song Y, Wang Q, Wang D, Junqiang L, Yang J, Li H, et al. Label-free quantitative proteomics unravels carboxypeptidases as the novel biomarker in pancreatic ductal adenocarcinoma. *Transl Oncol* 2018;11:691-9.

131. Mikesch JH, Schier K, Roetger A, Simon R, Buerger H, Brandt B. The expression and action of decay-accelerating factor (CD55) in human malignancies and cancer therapy. *Cell Oncol* 2006;28:223-32.
132. Saygin C, Wiechert A, Rao VS, Alluri R, Connor E, Thiagarajan PS, et al. CD55 regulates self-renewal and cisplatin resistance in endometrioid tumors. *J Exp Med* 2017;214:2715-32.
133. Macor P, Tripodo C, Zorzet S, Piovan E, Bossi F, Marzari R, et al. In vivo targeting of human neutralizing antibodies against CD55 and CD59 to lymphoma cells increases the antitumor activity of rituximab. *Cancer Res* 2007;67:10556-63.
134. Mamidi S, Cinci M, Hasmann M, Fehring V, Kirschfink M. Lipoplex mediated silencing of membrane regulators (CD46, CD55 and CD59) enhances complement-dependent anti-tumor activity of trastuzumab and pertuzumab. *Mol Oncol* 2013;7:580-94.
135. Reid MD, Basturk O, Thirabanasak D, Hruban RH, Klimstra DS, Bagci P, et al. Tumor-infiltrating neutrophils in pancreatic neoplasia. *Mod Pathol* 2011;24:1612-9.
136. Sadot E, Basturk O, Klimstra DS, Gonen M, Lokshin A, Do RK, et al. Tumor-associated neutrophils and malignant progression in intraductal papillary mucinous neoplasms: An opportunity for identification of high-risk disease. *Ann Surg* 2015;262:1102-7.
137. Oh SH, Choi YB, Kim JH, Weihl CC, Ju JS. Comparisons of ELISA and western blot assays for detection of autophagy flux. *Data Brief* 2017;13:696-9.
138. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235:177-82.
139. Natali PG, Nicotra MR, Bigotti A, Venturo I, Slamon DJ, Fendly BM, Ullrich A. Expression of the p185 encoded by HER2 oncogene in normal and transformed human tissues. *Int J Cancer* 1990;45:457-61.
140. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al.

- Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989;244:707-12.
141. Eroglu Z, Tagawa T, Somlo G. Human epidermal growth factor receptor family-targeted therapies in the treatment of HER2-overexpressing breast cancer. *Oncologist* 2014;19:135-50.
 142. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline update. *J Clin Oncol* 2013;31:3997-4013.
 143. Dowsett M, Hanna WM, Kockx M, Penault-Llorca F, Ruschoff J, Gutjahr T, et al. Standardization of HER2 testing: Results of an international proficiency-testing ring study. *Mod Pathol* 2007;20:584-91.
 144. Tsai YF, Tseng LM, Lien PJ, Hsu CY, Lin YS, King KL, et al. HER2 immunohistochemical scores provide prognostic information for patients with HER2-type invasive breast cancer. *Histopathology* 2019;74:578-86.
 145. Garrison LP, Jr., Babigumira JB, Masaquel A, Wang BC, Lalla D, Brammer M. The lifetime economic burden of inaccurate HER2 testing: Estimating the costs of false-positive and false-negative HER2 test results in us patients with early-stage breast cancer. *Value Health* 2015;18:541-6.
 146. Roepman P, Horlings HM, Krijgsman O, Kok M, Bueno-de-Mesquita JM, Bender R, et al. Microarray-based determination of estrogen receptor, progesterone receptor, and HER2 receptor status in breast cancer. *Clin Cancer Res* 2009;15:7003-11.
 147. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American society of clinical oncology/college of american pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007;25:118-45.
 148. Yaziji H, Goldstein LC, Barry TS, Werling R, Hwang H, Ellis GK, et al. HER-2 testing in breast cancer using parallel tissue-based methods. *JAMA* 2004;291:1972-7.

149. Ridolfi RL, Jamehdor MR, Arber JM. Her-2/neu testing in breast carcinoma: A combined immunohistochemical and fluorescence in situ hybridization approach. *Mod Pathol* 2000;13:866-73.
150. Press MF, Seoane JA, Curtis C, Quinaux E, Guzman R, Sauter G, et al. Assessment of ERBB2/HER2 status in HER2-equivocal breast cancers by fish and 2013/2014 asco-cap guidelines. *JAMA Oncol* 2018;5:366-75.
151. Steiner C, Ducret A, Tille JC, Thomas M, McKee TA, Rubbia-Brandt L, et al. Applications of mass spectrometry for quantitative protein analysis in formalin-fixed paraffin-embedded tissues. *Proteomics* 2014;14:441-51.
152. Bateman NW, Conrads TP. Recent advances and opportunities in proteomic analyses of tumour heterogeneity. *J Pathol* 2018;244:628-37.
153. An E, Ock CY, Kim TY, Lee KH, Han SW, Im SA, et al. Quantitative proteomic analysis of HER2 expression in the selection of gastric cancer patients for trastuzumab treatment. *Ann Oncol* 2017;28:110-5.
154. Kim H, Sohn A, Yeo I, Yu SJ, Yoon JH, Kim Y. Clinical assay for AFP-L3 by using multiple reaction monitoring-mass spectrometry for diagnosing hepatocellular carcinoma. *Clin Chem* 2018;64:1230-8.
155. Sohn A, Kim H, Yeo I, Kim Y, Son M, Yu SJ, et al. Fully validated SRM-MS-based method for absolute quantification of PIVKA-II in human serum: Clinical applications for patients with hcc. *J Pharm Biomed Anal* 2018;156:142-6.
156. Fox CH, Johnson FB, Whiting J, Roller PP. Formaldehyde fixation. *J Histochem Cytochem* 1985;33:845-53.
157. Hembrough T, Thyparambil S, Liao WL, Darfler MM, Abdo J, Bengali KM, et al. Application of selected reaction monitoring for multiplex quantification of clinically validated biomarkers in formalin-fixed, paraffin-embedded tumor tissue. *J Mol Diagn* 2013;15:454-65.
158. Steiner C, Tille JC, Lamerz J, Kux van Geijtenbeek S, McKee TA, Venturi M, et al. Quantification of HER2 by targeted mass spectrometry in formalin-fixed paraffin-embedded (FFPE) breast cancer tissues. *Mol Cell Proteomics* 2015;14:2786-99.

159. Hembrough T, Thyparambil S, Liao WL, Darfler MM, Abdo J, Bengali KM, et al. Selected reaction monitoring (SRM) analysis of epidermal growth factor receptor (EGFR) in formalin fixed tumor tissue. *Clin Proteomics* 2012;9:5.
160. Ingthorsson S, Andersen K, Hilmarsson B, Maeldansmo GM, Magnusson MK, Gudjonsson T. HER2 induced emt and tumorigenicity in breast epithelial progenitor cells is inhibited by coexpression of EGFR. *Oncogene* 2016;35:4244-55.
161. Yoo SH, Park IA, Chung YR, Kim H, Lee K, Noh DY, et al. A histomorphologic predictive model for axillary lymph node metastasis in preoperative breast cancer core needle biopsy according to intrinsic subtypes. *Hum Pathol* 2015;46:246-54.
162. Lae M, Freneaux P, Sastre-Garau X, Chouchane O, Sigal-Zafrani B, Vincent-Salomon A. Secretory breast carcinomas with ETV6-NTRK3 fusion gene belong to the basal-like carcinoma spectrum. *Mod Pathol* 2009;22:291-8.
163. Cantaloni C, Tonini RE, Eccher C, Morelli L, Leonardi E, Bragantini E, et al. Diagnostic value of automated HER2 evaluation in breast cancer: A study on 272 equivocal (score 2+) HER2 immunoreactive cases using an FDA approved system. *Appl Immunohistochem Mol Morphol* 2011;19:306-12.
164. Plancoulaine B, Laurinaviciene A, Meskauskas R, Baltrusaityte I, Besusparis J, Herlin P, Laurinavicius A. Digital immunohistochemistry wizard: Image analysis-assisted stereology tool to produce reference data set for calibration and quality control. *Diagn Pathol* 2014;9 Suppl 1:S8.
165. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010;26:966-8.
166. Vivo-Truyols G, Schoenmakers PJ. Automatic selection of optimal savitzky-golay smoothing. *Anal Chem* 2006;78:4598-608.
167. Sharma V, Eckels J, Taylor GK, Shulman NJ, Stergachis AB, Joyner SA, et al. Panorama: A targeted proteomics knowledge base. *J Proteome Res*

- 2014;13:4205-10.
168. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. Proteomexchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 2014;32:223-6.
 169. Vaught DB, Stanford JC, Young C, Hicks DJ, Wheeler F, Rinehart C, et al. HER3 is required for HER2-induced preneoplastic changes to the breast epithelium and tumor formation. *Cancer Research* 2012;72:2672-82.
 170. Zhang LQ, Li WH. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular Biology and Evolution* 2004;21:236-9.
 171. Stein SE, Scott DR. Optimization and testing of mass spectral library search algorithms for compound identification. *J Am Soc Mass Spectrom* 1994;5:859-66.
 172. Abbatiello SE, Mani DR, Keshishian H, Carr SA. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin Chem* 2010;56:291-305.
 173. Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, et al. Targeted peptide measurements in biology and medicine: Best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol Cell Proteomics* 2014;13:907-17.
 174. Prentice R. Use of the logistic model in retrospective studies. *Biometrics* 1976;32:599-606.
 175. Fluss R, Faraggi D, Reiser B. Estimation of the youden index and its associated cutoff point. *Biom J* 2005;47:458-72.
 176. Kennedy JJ, Abbatiello SE, Kim K, Yan P, Whiteaker JR, Lin C, et al. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nat Methods* 2014;11:149-55.
 177. David G. Hicks, Kulkarni S. HER2+ breast cancer: Review of biologic relevance and optimal use of diagnostic tools. *Am J Clin Pathol* 2008;129:263-73.
 178. Kao KJ, Tai CH, Chang WH, Yeh TS, Chen TC, Lee GB. A fluorescence in

- situ hybridization (FISH) microfluidic platform for detection of HER2 amplification in cancer cells. *Biosens Bioelectron* 2015;69:272-9.
179. Zhou W, Xu F, Li D, Chen Y. Improved detection of HER2 by a quasi-targeted proteomics approach using aptamer-peptide probe and liquid chromatography-tandem mass spectrometry. *Clin Chem* 2018;64:526-35.
 180. Sapino A, Maletta F, Verdun di Cantogno L, Macri L, Botta C, Gugliotta P, et al. Gene status in HER2 equivocal breast carcinomas: Impact of distinct recommendations and contribution of a polymerase chain reaction-based method. *Oncologist* 2014;19:1118-26.
 181. Marchione DM, Ilieva I, Devins K, Sharpe D, Pappin DJ, Garcia BA, et al. Hypersol: High-quality data from archival FFPE tissue for clinical proteomics. *J Proteome Res* 2020;19:973-83.
 182. Zhu Y, Weiss T, Zhang Q, Sun R, Wang B, Yi X, et al. High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. *Mol Oncol* 2019;13:2305-28.
 183. Cheng AL, Huang WG, Chen ZC, Zhang PF, Li MY, Li F, et al. Identifying cathepsin D as a biomarker for differentiation and prognosis of nasopharyngeal carcinoma by laser capture microdissection and proteomic analysis. *Journal of Proteome Research* 2008;7:2415-26.
 184. Broeckx V, Boonen K, Pringels L, Sagaert X, Prenen H, Landuyt B, et al. Comparison of multiple protein extraction buffers for GELC-MS/MS proteomic analysis of liver and colon formalin-fixed, paraffin-embedded tissues. *Mol Biosyst* 2016;12:553-65.
 185. Coscia F, Doll S, Bech JM, Schweizer L, Mund A, Lengyel E, et al. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol* 2020;251:100-12.
 186. Zhang Y, Muller M, Xu B, Yoshida Y, Horlacher O, Nikitin F, et al. Unrestricted modification search reveals lysine methylation as major modification induced by tissue formalin fixation and paraffin embedding. *Proteomics* 2015;15:2568-79.

ABSTRACT IN KOREAN

국문 초록

서론: 질량분석학 기반 단백질체학적 접근법은 미량의 시료에서 수백 개의 차등적으로 발현되는 단백질을 발굴하기 위해 수천 개의 단백질을 동시에 스크리닝할 수 있는 능력을 기반으로, 특정 질병과 연관된 바이오마커를 식별하는 데 점점 더 많이 적용되고 있다. 일반적으로 임상 코호트로부터 수집된 체액과 포르말린 고정 파라핀 포매조직절편 (FFPE)과 같은 병리학적 검체를 분석한다. 단백질 분석에 있어 높은 처리량과 감도를 가진 질량분석학 기반 접근법은 바이오마커 발굴 및 임상 진단 분야에서 강력한 도구로 활용될 수 있다. 또한 단백질 연구는 질병의 생물학적 메커니즘을 이해하는데 도움을 준다.

방법: 1 장과 2 장에서, 고분해능 질량분석기 기반 단백질체학 분석을 수행하여 췌장낭종액 시료에서 췌관내 유두상 점액 종양 (IPMN)의 악성도를 예측할 수 있는 바이오마커를 발굴하였다. 2 장에서는 실제 임상 상황을 더 잘 반영하기 위해 췌관내 유두상 점액 종양 뿐만 아니라 점액성 낭성 종양 (MCN)과 장액성 낭성 종양 (SCN)을 추가한 확장된 코호트에서 시료를 수집하였다. 3 장에서, 표적 단백질 기술인 다중반응검지법을 FFPE 조직에 적용하여 유방암 환자의 인간 상피 증식 인자 수용체 2 (HER2) 상태를 결정할 수 있는 새로운 분석법을 확립하였다.

결과: 1 장에서, IPMN 환자의 췌장낭종액 시료에서 2,992 개의 단백질을 동정하였다. IPMN 의 조직학적 등급에 따라 차등적으로 발현되는 18 개의 바이오마커 후보군이 발견되었으며, 그 중 일부는 독립적인 코호트를 이용하여 웨스턴 블롯으로 검증하였다. 그 결과는 단백질 데이터와 일치하였다. 2 장에서, IPMN, MCN, SCN 환자의 췌장낭종액에서 5,834 개의 단백질을 동정하였다. IPMN 이형성증간에 차별적으로 발현되는 364 개의 단백질 중, 19 개의 최종 바이오마커 후보군은 IPMN 의 악성도에 따라 연속적으로 증가하거나 감소하였다. 독립코호트에서 CD55 단백질을 효소면역측정법 (ELISA), 웨스턴블롯, 면역화학염색 (IHC)을 통해 검증하였으며, 단백질 데이터와 일치하는 결과를 얻을 수 있었다. 3 장에서, 우리는 HER2 상태를 구별하기 위한 기존 방법을 개선하는 다중반응검지법을 확립하였다. 충분한 양의 단백질을 확보하는 데 필요한 FFPE 슬라이드 수를 산출하고, 상피 세포 특이적 단백질의 발현량을 HER2 발현량 측정을 위한 정규화 인자로 사용함으로써 시료 전처리를 단순화하였다. 이에 따라 HER2 단백질 정량의 정확성과 정밀도가 향상되었다.

결론: 1 장과 2 장에서, 우리는 현존하는 최대의 췌장낭종액 단백질 데이터를 생성했으며, IPMN 이형성증의 잠재적 바이오마커를 발굴하였다. 췌장낭종액 바이오마커의 발굴은 IPMN 의 악성도를 정확하게 평가하는데 도움이 되며, 외과적 의사 결정을 효과적으로 도울 수 있을 것이다. 궁극적으로, 발굴한 마커가 임상에서 유용하게 사용될 수 있다면, IPMN 이형성증에 대한 정확한 평가를 통해 저위험군 IPMN

환자의 불필요한 수술적 절제를 방지하는데 기여할 수 있을 것이다. 3 장에서, 모호한 HER2 그룹을 구별할 수 있는 우리의 프로토콜은 형광동소혼성화 (FISH) 테스트가 필요한 사례의 수를 줄임으로써 유방암 환자의 진단에 필요한 시간과 비용을 잠재적으로 줄일 수 있다. 또한 우리가 개발한 단순화된 분석 절차는 MRM 분석법을 임상에 적용하기 위한 진입 장벽을 낮춰준다. 우리가 확립한 MRM 분석법은 IHC 에 비해 보다 정확한 HER2 발현 수준을 산출함으로서, 임상 의가 유방암 환자를 위한 적절한 치료 방침을 결정하는데 도움을 줄 수 있을 것이다.

주요어: 단백질체학; 질량분석학; 다중반응검지법; 체장낭종액; 체관내 유두상 점액 종양; 바이오마커; 인간 상피 증식 인자 수용체 2; 포르말린 고정 파라핀 포매조직절편

학 번: 2014-25061

* 본 내용은 학술지에 게재된 세 논문, Clinical Proteomics (2018, 15:17), Cancers (2020, 12:9), 그리고 Clinical Chemistry (2020, 66:10)을 바탕으로 작성하였음.