A thesis of the Degree of Doctor of Philosophy in Science

Research on Multiple Reaction Monitoring – Mass Spectrometry (MRM-MS) for Large-scale Clinical Proteomics: Development of Automated Workflow for Serum Sample Preparation and Inclusive Quantification Assay for DCP Proteoforms

대규모 임상 단백체학을 위한 다중반응검지 질량분석법 연구: 혈청 시료의 자동화 전처리 과정 및 DCP 단백질형 포괄적 정량 어세이 개발

August 2022

Graduate School of Biomedical Sciences

Seoul National University

Major in Biomedical Sciences

Jihyeon Lee

이학박사 학위논문

대규모 임상 단백체학을 위한 다중반응검지 질량분석법 연구: 혈청 시료의 자동화 전처리 과정 및 DCP 단백질형 포괄적 정량 어세이 개발

Research on Multiple Reaction Monitoring – Mass Spectrometry (MRM-MS) for Large-scale Clinical Proteomics: Development of Automated Workflow for Serum Sample Preparation and Inclusive Quantification Assay for DCP Proteoforms

2022년 8월

서울대학교 대학원

의과학과 의과학전공

이 지 현

대규모 임상 단백체학을 위한 다중 반응검지 질량분석법 연구: 혈청 시료의 자동화 전처리 과정 및 DCP 단백질형 포괄적 정량 어세이 개발

지도교수 김 영 수

이 논문을 이학박사 학위논문으로 제출함
2022년 4월

서울대학교 대학원
의과학과 의과학전공
이 지 현

이지현의 이학박사 학위논문을 인준함
2022년 7월

위 원 장	이 용 석	(인)
부위원장	김 영 수	(인)
위 원	한 범	(인)
위 원	김 종 서	(인)
위 원	김 경 곤	(인)

Research on Multiple Reaction Monitoring – Mass Spectrometry (MRM-MS) for Large-scale Clinical Proteomics: Development of Automated Workflow for Serum Sample Preparation and Inclusive Quantification Assay for DCP Proteoforms

by
Jihyeon Lee

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

July 2022

Approved by Thesis Committee:

Professor _____ Chairman

Professor _____ Vice Chairman

Professor _____

Professor _____

Professor _____

ABSTRACT

Research on Multiple Reaction Monitoring – Mass Spectrometry (MRM-MS) for Large-scale Clinical Proteomics: Development of Automated Workflow for Serum Sample Preparation and Inclusive Quantification Assay for DCP Proteoforms

Jihyeon Lee

Major in Biomedical Sciences

Department of Biomedical Sciences

Seoul National University

Graduate School

Introduction: Mass spectrometry (MS)-based proteomics can make high-throughput analysis, based on its ability to detect and quantify thousands of proteins simultaneously. Reproducible sample preparation remains a significant challenge in large-scale clinical research using multiple reaction monitoring−mass spectrometry (MRM−MS), which enables a highly sensitive multiplexed assay. Although automated liquid-handling platforms are specially designed to address this issue, the high cost of their consumables is a drawback that renders routine operation expensive and impractical. Meanwhile, the MS-based approach has great advantages over antibody-based assays in terms of distinguishing and simultaneously

quantifying multiple proteoforms, including isoforms or posttranslational modifications. A typical example of a protein biomarker containing posttranslational modifications is des-γ-carboxyprothrombin (DCP) which is a hepatocellular carcinoma (HCC) serologic surveillance marker. DCP exists in the blood as a mixture of proteoforms that are made from an impaired carboxylation process at glutamic acid (Glu) residues within the N-terminal domain. The heterogeneity of DCP may affect the accuracy of measurements because DCP levels are commonly determined using an immunoassay that relies on antibody reactivity to an epitope in the DCP molecule.

Methods: In chapter 1, I evaluated the reproducibility of quantification results of the MRM-MS assay of 52 peptides in serum samples prepared by the automated workflow. Further, I performed a collateral systematic evaluation of the possibility of a cost-reduced workflow in a liquid-handling platform. In chapter 2, I aimed to improve the DCP measurement assay by applying a mass spectrometry (MS)-based approach for a more inclusive quantification of various DCP proteoforms. I developed an MRM-MS assay to quantify multiple non-carboxylated peptides included in the various des-carboxylation states of DCP. I performed the MRM-MS assay on 300 patients and constructed a robust diagnostic model that simultaneously monitored three non-carboxylated peptides.

Results: In chapter 1, I evaluate the feasibility of the automated workflow of serum sample preparation. I demonstrated that the automated workflow ensures stable serum sample preparation as evidenced by the average value of total CVs (15.3%). In collateral comparison, I found it possible to save 37% of the total experimental cost with the automated workflow with a cost-optimized method when compared to

the standard procedure, while maintaining nearly equivalent reproducibility. In chapter 2, the MS-based quantitative assay for DCP had reliable surveillance power, which was evident from the area under the receiver operating characteristic curve (AUROC) values of 0.874 and 0.844 for the training and test sets, respectively. It was equivalent to conventional antibody-based quantification, which had AUROC values at the optimal cutoff (40 mAU/mL) of 0.743 and 0.704 for the training and test sets, respectively. The surveillance performance of the MS-based DCP assay was validated using an independent validation set consisting of 318 patients from an external cohort, resulting in an AUROC value of 0.793.

Conclusions: In chapter 1, the routine operation of liquid-handling platforms can enable researchers to process large-scale samples with high throughput, adding credibility to their findings by minimizing human error. In chapter 2, due to higher diagnostic performance and high reproducibility, the quantitative DCP assay using the MRM-MS method is superior to the antibody-based quantification assay.

Keywords: Clinical Proteomics; Mass spectrometry; Multiple Reaction Monitoring; Biomarker; Automation; Assay development

Student number: 2017-24628

* This work is published in Journal of proteome Research Journal. Cost-Effective Automated Preparation of Serum Samples for Reproducible Quantitative Clinical Proteomics (J. Lee, H. Kim, A. Sohn, I. Yeo, and Y. Kim). Published 15 Apr 2019/

Journal of Proteome Research 10.1021/acs.jproteome.9b00023. (1)

* This work is published in Hepatology Communications Journal. Inclusive Quantification Assay of Serum Des-γ-Carboxyprothrombin Proteoforms for HCC Surveillance via Targeted Mass Spectrometry (J. Lee, Y. Lim, J-H. Lee, G. Gwak, M. Do, I. Yeo, D. Shin, D. Han, T. Park, and Y. Kim). Published 12 June 2021/ Hepatology Communications 10.1002/hep4.1752. (2)

TABLE OF CONTENTS

Abstract i

Table of Contents v

List of Tables vii

List of Figures x

List of Abbreviations xiv

General Introduction 1

Chapter 1 6

Feasibility Study on the Automated Workflow of Serum Sample Preparation for Reproducible Quantitative Clinical Proteomics

1. **Introduction** 7
2. **Materials and Methods** 10
3. **Results** 32
4. **Discussion** 49
5. **Conclusion** 53

Chapter 2 56

Development of the Inclusive Quantification MRM-MS Assay of Des-γ-Carboxyprothrombin (DCP) Proteoforms for Hepatocellular Carcinoma (HCC) Surveillance

1. Introduction **57**

2. Materials and Methods **60**

3. Results **84**

4. Discussion **118**

5. Conclusion **125**

General Conclusion **126**

References **130**

Abstract in Korean **140**

LIST OF TABLES

Chapter 1

Table 1-1. Total workflow of in-solution digestion protocol 15

Table 1-2. Properties of target peptides 20

Table 1-3. Volumes needed when dispensing 10 μL of reagents for each set..... 34

Table 1-4. Variation in each set for 52 target peptides 38

Table 1-5. Quantification of each set for 52 target peptides..... 42

Table 1-6. Approximate experimental cost per sample for in-solution digestion 47

Chapter 2

Table 2-1. Clinical characteristics of the subjects in each cohort..... 64

Table 2-2. Total workflow of the multi-enzyme, in-solution digestion protocol 71

Table 2-3. MRM-MS parameters for the quantification of four Glu-peptides from the DCP Gla domain 75

Table 2-4. Results of the reversed response curve analysis for four Glu-peptides 77

Table 2-5. Summary of reversed response curves for each Glu-peptide. 86

Table 2-6. Analytical sensitivity from analysis results of zero sample and LLOQ sample by MRM-MS assay..... 87

Table 2-7. Analytical specificity (interference) of six different matrices by MRM-MS assay 89

Table 2-8. Carryover measurements using zero sample run after the highest calibrator..... 92

Table 2-9. Reproducibility of the MRM-MS assay over 5 days..... 95

Table 2-10. Summary of the DCP 3-Glu-peptide logistic regression model..... 98

Table 2-11. Comparison of diagnostic performance of 3-Glu-peptide panel, AFP level, and combined model in each data set 110

Table 2-12. Comparison of diagnostic performance of 3-Glu-peptide

panel, AFP level, and combined model for discriminating early-stage HCC from stratified controls 116

Table 2-13. Prospective combinations of 3 Glu-peptides that represent 10 DCP variants 122

LIST OF FIGURES

Chapter 1

Figure 1-1. Arrangement of the 9 decks in the Bravo liquid handling platform 13

Figure 1-2. A schematic overview of automated in-solution digestion of human blood samples 13

Figure 1-3. Sequential tasks required in the reagents dispensing procedure 14

Figure 1-4. Three factors for optimizing the aspiration height to minimize wasted volumes 18

Figure 1-5. Distribution of molecular weights of target proteins and retention times for target peptides 19

Figure 1-6. Experimental scheme for evaluation of feasibility of automated in-solution digestion process 28

Figure 1-7. A systematic evaluation of feasibility for automated digestion process and further optimization through exploring the possibility of re-using pipette tips to reduce cost 29

Figure 1-8. Visualization of the formation to calculate intra- and inter-CV 31

Figure 1-9. Optimization of aspiration height for water and serum ... 33

Figure 1-10. Variations of the automated sample preparation process 35

Figure 1-11. Box and whisker plots of intraday and interday variation testing for each set 37

Figure 1-12. Time spent for dispensing 10 µL of reagent to every 24 wells 41

Figure 1-13. The relationship between variation and cost 47

Figure 1-14. The overall scheme for further evaluation of the robustness of the automated sample preparation method in biological replication experiments 51

Figure 1-15. The potential of the robust automated sample preparation method for the overall analytical method validation process to reduce human labor 54

Chapter 2

Figure 2-1. Design of the study 62

Figure 2-2. Candidate non-carboxylated peptides (Glu-peptides) within the Gla domain identified by in silico digestion with chymotrypsin and trypsin 68

Figure 2-3. Scheme of the adapted sample preparation workflow using semi-separate, multi-enzyme digestions for the simultaneous procurement and analysis of tryptic and chymotryptic peptides in a single batch 70

Figure 2-4. The overall scheme of the optimal DCP model and combined model construction and evaluation 82

Figure 2-5. Reversed response curves for the quantifier ions of ANTFLEEVR (A), ERECVEETCSY (B), ESSTATDVF (C), and EEVRKGNL (D) 85

Figure 2-6. Assessment of the diagnostic performance of the DCP 3-Glu-peptide panel in the training and test sets 98

Figure 2-7. Comparison of the relative peptide levels of the three Glu-peptides for the training set (A) and the test set (B) 100

Figure 2-8. Interactive dot plots for the HCC diagnosis performance of the serum DCP level (measured by the immunoassay) and the DCP panel (measured by the MRM-MS) in the training set (A) and the test set (B) 102

Figure 2-9. Correlation between DCP values by immunoassay and

logit(P) values, calculated based on the DCP panel 103

Figure 2-10. Comparison of the diagnostic performances of the DCP 3-Glu-peptide panel, serum AFP levels, and a combined model (DCP 3-Glu-peptide panel and serum AFP level) in the training (A) and test sets (B) 105

Figure 2-11. Diagnostic performances of the DCP 3-Glu-peptide panel, serum AFP levels, and a combined model 106

Figure 2-12. Independent validation of the diagnostic performance of the DCP 3 Glu-peptides panel, serum AFP levels, and the combined model (DCP 3 Glu-peptides panel and serum AFP level) 108

Figure 2-13. The HCC diagnostic abilities of the DCP 3 Glu-peptides panel and the combined model (DCP 3 Glu-peptides panel and serum AFP level) in AFP- and DCP-negative subgroup of the independent validation set 112

Figure 2-14. The HCC surveillance abilities of the DCP 3 Glu-peptides panel and the combined model (DCP 3 Glu-peptides panel and serum AFP level) in the independent validation set 114

Figure 2-15. The specific order of carboxylation modifications for ten Glu residues resulting in the ten possible intermediate DCP proteoforms 121

Figure 2-16. Possible experiment scheme for fidelity analysis of protease digestion to increase the credibility of quantification results from the MRM-MS assay 123

LIST OF ABBREVIATIONS

AFP: alpha-fetoprotein

AUROC: area under the receiver operating curve

BCA: bicinchoninic acid

CI: confidence interval

CV: coefficient of variation

DCP: des-γ-carboxyprothrombin

DTT: dithiothreitol

FDA: Food and Drug Administration

FFPE: formalin-fixed paraffin-embedded

FIA: fluorescence immunoassay

Gla: carboxyglutamic acid

Glu: glutamic acid

HBV: hepatitis B virus

HCC: hepatocellular carcinoma

HCV: hepatitis C virus

LC: liquid chromatography

MRI: magnetic resonance imaging

MRM-MS: multiple reaction monitoring-mass spectrometry

MS: mass spectrometry

PAR: peak area ratio

PIVKA-II: protein induced by vitamin K absence or antagonist-II

PTM: post-translational modification

SDS: sodium dodecyl sulfate

SIS: stable isotope-labeled internal standard

US: ultrasonography

GENERAL INTRODUCTION

Biomarkers are evaluated as indicators of biological processes or pathological conditions, or drug responses. In recent years, the importance of biomarkers as tools for disease diagnosis, personalized medicine, and drug discovery has increased (3, 4). As proteins are the main functional components of biological pathways which are altered during disease conditions, protein-based biomarkers are key players in identifying and understanding the biological mechanism of diseases. Thus, clinical proteomics is a promising approach to the discovery of biomarkers for disease diagnosis or predictions through the systematic analysis of protein structure, expression, interactions, and posttranslational modifications (PTMs).

Clinical proteomics requires large-scale analysis for the discovery or validation of disease biomarkers to increase the statistical power. One of the powerful proteomic technologies is mass spectrometry (MS)-based proteomics, which has tremendous potential in the clinical field. Multiple reaction monitoring-mass spectrometry (MRM-MS) assay, which is a targeted MS-based method for accurately quantifying thousands of proteins simultaneously, can enable high-throughput analysis. Thus, thousands of proteins in biological materials can be multiplexed into a single MRM-MS assay and precisely quantified at femtomole levels with low coefficients of variation (CVs). A single MRM-MS assay for multiple proteins enables large cohorts to be analyzed at high throughput. Analysis through bioinformatics methods such as machine learning for large-scale MS-based proteomics data can make a great opportunity for elucidating complex disease-related mechanisms.

Reproducible quantification results of MRM-MS assay are mandatory for large-scale analysis of clinical proteomics. In this regard, the reproducible sample preparation process for MRM-MS analysis is the fundamental step in obtaining reliable results. Proteins are digested by a protease to generate peptides before MS analysis. A series of steps for proteomic sample preparation is very labor-intensive and time-consuming. Also, the proficiency of researchers affects the quality of proteomic results. Therefore, securing the reproducibility of sample preparation for high-throughput proteomic analysis in large-scale cohorts remains a bottleneck. To address this, the utilization of robotic liquid handling platforms is an emerging innovation to improve the throughput and reproducibility of protein sample preparation while also reducing human labor. The automated workflows in proteomics will enhance the consistency and precision, which is crucial to making accurate and reliable quantification results for validated clinical assays.

Currently, many efforts have been made to utilize liquid handling platforms for various proteomic sample preparation methods, such as in-gel, in-solution, or in-tip digestion (5-15). However, few studies have been focusing on the automated protein sample preparation for targeted quantitative assay using MRM-MS analysis. In this aspect, I evaluated the feasibility of introducing a highly automated sample preparation workflow to MRM-MS assay for large-scale clinical proteomic analysis. In chapter 1, I developed a high-throughput in-solution protein digestion method for serum samples using an Agilent Bravo liquid handling platform with a 96-well format. To assess the robustness of the automated system, identical pooled samples were analyzed as technical replications over three consecutive days. The automated in-solution digested samples displayed high reproducibility, whose coefficient of

variation (CV) was under 20%. Thus, the introduction of automation could open a new era for clinical proteomics studies.

In further optimization, I evaluated the potential for cost reduction to relieve the expenses of an automated workflow for their routine operation in large-scale studies, for which automation is sorely needed. Therefore, a systematic evaluation of the digestion in a liquid-handling platform was conducted while reducing the cost of consumables. I compared the targeted proteins quantification results obtained from the cost-optimized method with results obtained by the standard method operated at the maximum cost of consumables. As a result, I found that the cost-optimized method with reduced total cost by 37% and maintained comparable levels of reproducibility. Overall, the automated in-solution protein sample preparation workflow is a step closer to realizing high-throughput applications that enable robust and reliable large-scale proteomics analysis.

Meanwhile, immunoassays have been regarded as standard techniques for measuring proteomic biomarkers. For example, alpha-fetoprotein (AFP) and des-γ-carboxyprothrombin (DCP), which are protein biomarkers for hepatocellular carcinoma (HCC) surveillance, have been quantified by fluorescence immunoassay (FIA). However, immunoassays are prone to producing false-negative results due to their antibody-dependency and limit of quantification. Furthermore, the proteoforms containing PTMs in the epitope region of antibodies show decreased binding affinity which could result in inaccuracy and imprecise protein quantification. Overcoming these limitations is imperative in ensuring reproducible protein quantification results in large-scale clinical cohorts. In this regard, targeted MS-based quantitative proteomics can supplement the limitations of the immunoassays. MS-based assays

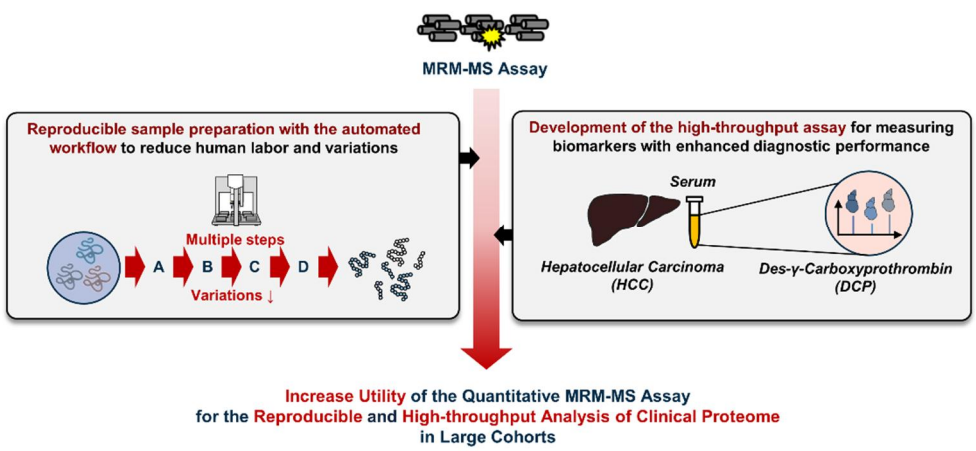
have great advantages over immunoassays in terms of distinguishing and simultaneously quantifying multiple proteoforms, including isoforms or PTMs (16).

In chapter 2, I developed the powerful diagnostic MS-based assay of des-γ-carboxyprothrombin (DCP) by inclusively quantifying a wider range of proteoforms with various des-carboxylation states. DCP is a biomarker for hepatocellular carcinoma (HCC) surveillance, also known as a protein induced by vitamin K absence or antagonist-II (PIVKA-II) or abnormal prothrombin. Normal prothrombin is synthesized as a precursor containing 10 glutamic acid (Glu) residues in the N-terminal domain (Gla domain), and then undergoes posttranslational carboxylation of the Glu residues resulting in the conversion of Glu to γ-carboxylated glutamic acid (Gla) by vitamin K-dependent glutamyl gamma-carboxylase. Meanwhile, when carboxylation is impaired under conditions of vitamin K deficiency, warfarin treatment, or liver dysfunction, DCP is released into the bloodstream as a mixture of 10 proteoforms with up to 10 des-carboxylated Glu residues.

The conventional assay for measuring DCP concentrations is an antibody-based assay using a monoclonal antibody produced by the MU3 cell line, which binds predominantly with DCP molecules containing 9-10 Glu residues. It had been identified that their binding affinity is weak against proteoforms possessing 6-8 Glu residues, and rarely binds with those that have less than 5 Glu residues. Recently, several studies aimed to develop discriminative quantification immunoassays for the detection of DCP proteoforms with lower Glu content using additional immunoassays that feature other antibodies, which recognize different epitopes than those that are detected by the MU3 antibody. Although these studies reported the value of assays for the detection of lower Glu residues containing DCP proteoforms,

these immunoassays are costly and susceptible to batch effects because of the additional immunoassays required. Therefore, I examined potential non-carboxylated peptides (referred to as Glu-peptides) within the Gla domain. A robust MRM-MS assay that can quantify four Glu-peptides was developed and analyzed 618 serum samples, which were obtained from patients with HCC and at-risk control patients. As a result of machine learning analysis in these large cohorts, I could achieve reliable HCC surveillance power of the MRM-MS assay for inclusively quantifying DCP proteoforms with lower Glu residues.

From the results of chapters 1 and 2, I generated a significant step toward reproducible large-scale clinical proteomics through the utilization of high-throughput MRM-MS assays, as illustrated below.



CHAPTER 1

Feasibility Study on the Automated Workflow of Serum Sample Preparation for Reproducible Quantitative Clinical Proteomics

1.1. INTRODUCTION

Multiple reaction monitoring (MRM) is a targeted mass spectrometry (MS) method for accurately quantifying proteins (17-19). MRM−MS has advantages over antibody-based assays that rely on the recognition of epitopes because it can distinguish variants and isoforms of proteins that cannot be recognized by antibodies (20-22). In addition, many proteins can be multiplexed into a single MRM−MS assay, with most proteins detected at femtomole levels and, in most cases, with low coefficients of variation (CVs) (17, 23-30). Thus, quantitative multiplexed assays that are based on MRM−MS are an established technology in the research of protein biomarkers in complex biofluids and have recently been translated into a clinical application (31-36).

MRM−MS-based multiplexed assays that allow the precise quantification of peptides of widely varying masses. The MRM-MS assays are applicable to a large number of samples, which is necessary for clinical verification and validation, and are only achievable with highly reproducible digestion steps (37). However, parallel digestion of large-scale samples by manual preparation takes longer to process and requires highly skilled manpower. Manual handling of large numbers of samples can lead to handling errors during pipetting that can result in a significant loss of biological information. Moreover, the time that is required for processing the first and last samples cannot be controlled when handling a large number of samples, requiring them to be split into batches, which, in turn, can increase the time that is required for sample processing. This implies that not all samples are provided with

the ideal conditions for reproducibility.

Thus, MRM−MS methods must secure reproducibility of the sample preparation process with high throughput for their widespread adoption in biomarker validation and clinical application, while maintaining high precision during routine operation. To address this issue, many efforts have been made to introduce automated liquid handlers into various digestion protocols (38-43). The automation of sample preparation can increase sample processing throughput, reduce inter-sample variability, and eliminate the need for skilled labor in performing repetitive tasks (44). Parallel sample processing in 96-well format using a robotic liquid handler has been widely embraced for routine analysis in biomarker discovery and development (45-47). In particular, the introduction of an automation system is a key task for handling nanoliter scale samples with high throughput, such as single-cell proteomics studies (48, 49). Employment of automation via liquid handling platforms provides handling of accurate volume in a reproducible manner.

Currently, the evaluation of previous liquid-handling-based methods for sample preparation has focused on determining whether they generate results equivalent to those of manual processing and improve the reproducibility and throughput of protein sample preparation (10, 14, 46). One of the previous studies that quantified 600 peptides from 367 *E.coli* proteins presented comparable reproducibility for the automated sample preparation workflow to the manual workflow (8). The median CV for peptide measurement was 15.8 and 14.3% for the automated and manual workflow, respectively. Also, the median CV for protein measurement was 13.8 and 13.6% for the automated and manual workflow, respectively. However, few studies have focused on the automated protein sample preparation for targeted quantitative

assay using MRM-MS analysis for clinical assay. In this regard, the goal of this study was to develop an automated in-solution protein digestion method for serum samples using a liquid-handling platform for the clinical application of MRM−MS-based peptide quantification. As a result, the automated in-solution digestion process met the general CV value of 20%, which is often considered to be desirable or sufficient for clinical requirements. Despite their comparability, the high cost of liquid-handling platforms, directed toward exclusive tips or plates, is a major obstacle to their routine use in large-scale studies, for which automation is sorely needed. Therefore, I additionally performed a systematic evaluation of the possibility of a cost-reduced digestion workflow in a liquid-handling platform. The cost-reduced workflows were compared with standard experiments that were performed in parallel and operated at the maximum cost of consumables. As a result, I found that the total cost can be reduced by 37%, while nearly equivalent levels of reproducibility are maintained.

In summary, the automated workflow of in-solution digestion for serum sample preparation can increase throughput in a less labor-intensive manner, and reduce inter-sample variability in performing MRM-MS analysis in large-scale clinical cohorts. I expect that this automated workflow can be adopted for the automation of certain aspects of clinical studies in other human samples, such as tissue and formalin-fixed paraffin-embedded (FFPE) samples, with little modification.

1.2. MATERIALS AND METHODS

1.2.1. Materials

Pooled normal serum was purchased from a U.S. Food and Drug Administration (FDA)-approved facility (Innovative Research, Novi, MI). All high-performance liquid chromatography (HPLC)-grade solutions were purchased from Fisher Scientific (Loughborough, U.K.), including water, acetonitrile, formic acid, 0.1% formic acid in water, and 0.1% formic acid in acetonitrile. RapiGest surfactant was purchased from Waters (Milford, MA). Ammonium bicarbonate (ABC) solution was acquired from iNtRON Biotechnology (Sungnam, Korea). Dithiothreitol (DTT) was obtained from Amresco (Solon, OH). Iodoacetamide (IAA) was purchased from Sigma Aldrich (St. Louis, MO). Trypsin was obtained from Promega (sequencing-grade modified, Fitchburg, WI). Unpurified stable isotope-labeled standard (SIS) peptides that contained isotopically labeled (^{13}C and ^{15}N) arginine or lysine were obtained from JPT (Berlin, Germany) (30−70% purity, according to the manufacturer).

The Bravo Automated Liquid-Handling Platform (G5409A), with a 96-large-transfer tip head (Series III) and controlled using VWorks, was purchased from Agilent Technologies (Santa Clara, CA). The platform includes an orbital shaking station and a Peltier thermal station that was operated using an Inheco Single TEC Controller. Nunc DeepWell plates (1 mL) were purchased from Thermo Fisher Scientific (Waltham, MA) as reagent stock plates, and 96-well twin.tec PCR plates (150 μL, skirted) were purchased from Eppendorf (Hamburg, Germany) as sample

plates.

1.2.2. Establishing an automated in-solution digestion workflow utilizing liquid handling platform

The Bravo liquid-handling platform (Agilent), which is widely used in many laboratories, is comprised of nine decks that were accessible by the liquid-handling head, as shown in **Figure 1-1(A)**. The liquid-handling head can transfer an accurate volume of liquid or move plates from one deck to another. The nine deck locations were numbered and fixed from 1 to 9. I arranged certain stationary hardware, plates, and lids in consideration of their accessibility (**Figure 1-1(B)**). Deck location 5 was used for the sample plate and orbital shaking station because it was the only position that was accessible under every configuration of pipette channels. The Peltier thermal station had to be installed at deck location 4 or 6 due to its height, per the user guide. I installed a Peltier thermal station at deck location 4 to prevent collision with the liquid-handling head when it dispensed the reagent into the sample plate. The new tip box had to be positioned at deck location 2 to use the tips in the box effectively (**Figure 1-1(C)**). Other deck locations were arranged efficiently, taking into accounts the flow of the liquid-handling head movements.

The automated in-solution digestion workflow was operated by using three protocol files consecutively, as shown in **Figure 1-2**. The automated workflow begins with transferring serum samples into a 96-well sample plate (protocol file 1). Then, the main reaction for in-solution digestion was performed by protocol file 2: denaturation, alkylation, proteolysis, and quench steps. After the quench step, protocol file 3 was used for spiking SIS peptides. Based on the configuration and

operating principle of the Bravo liquid handling platform, dispense of each reagent was conducted through sequential 7 tasks: delid two plates, tips on the head, aspirate, dispense, tips off from the head, relid two plates, and shake sample plate (**Figure 1-3**). In case of reaction steps where specific temperature conditions existed, task 8 (place plate) should be conducted additionally.

The conditions and parameters of reaction for entire automated workflow are summarized in **Table 1-1**. Samples were prepared by sequentially loading 3 μL of HPLC-grade water and 2 μL of pooled serum sample onto a sample plate using individual pipet tips for each well. The plate was covered with a homemade lid and shaken at 1000 rpm for 10 s. Samples were denatured with 10 μL of denaturant (0.15% RapiGest, 10 mM DTT, 75 mM ABC, pH 8.0) at 60 °C for 60 min. Denatured samples were alkylated with 10 μL of reducing agent (50 mM IAA) at room temperature in the dark for 30 min. Tryptic digestion was performed with 10 μL of trypsin solution (1:50 (w/w)) at 37 °C for 4 h and stopped with 10 μL of 4.50% formic acid (final concentration of 1%) at 37 °C for 30 min. The plate was covered and shaken at 1000 rpm for 10 s after each reagent was added. Following the quenching step, 5 μL of SIS peptides (100 fmol/μL) was added to each well of the sample plate with individual tips. The plate was covered and shaken at 1000 rpm for 15 s. Each sample was moved manually to an individual microcentrifuge tube and centrifuged at 15000 rpm for 30 min at 4 °C to remove insoluble chemicals, such as byproducts of RapiGest surfactant. The supernatant (50 μL) was transferred to a new tube and subjected to online desalting and analysis by the MRM assay.

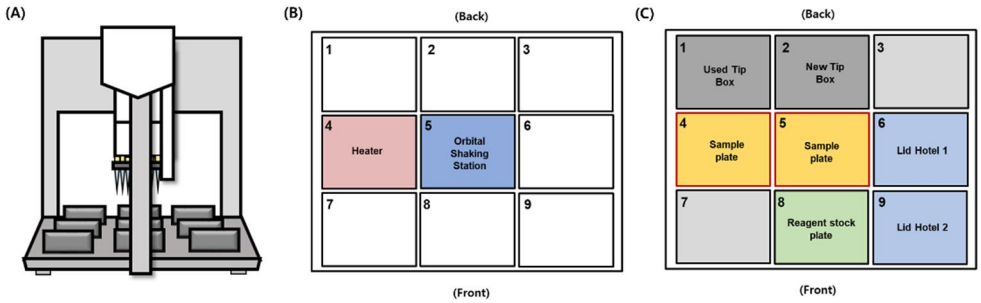


Figure 1-1. Arrangement of the 9 decks in the Bravo liquid handling platform.

(A) The Bravo platform has 9 decks into which a user can install an accessory, such as a heater and shaker. (B) The 2 fixed positions in the platform are shown. The heater station was installed at deck location 4 (pink), whereas the orbital shaking station was fixed at deck location 5 (blue). (C) Other positions are used according to the efficiency of management, as described. The pink decks were the positions at which the sample plate could be moved according to the reaction step. The tip boxes are located at deck locations 1 and 2 due to their height (dark gray decks). The lid hotel is the position for placing plate lids temporarily during aspiration and dispensing of reagents (blue decks). The reagent stock plate was prepared at deck location 8 (green deck). Deck locations 3 and 7 (gray decks) are empty positions.

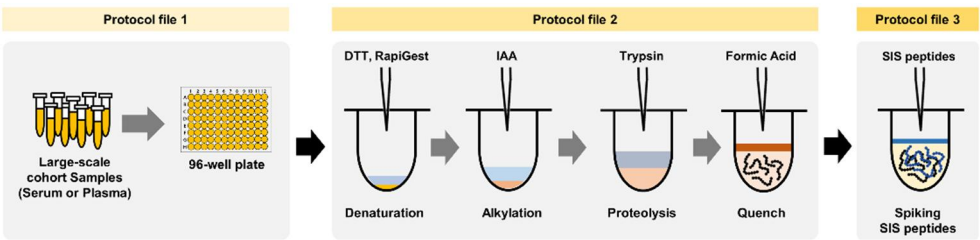


Figure 1-2. A schematic overview of automated in-solution digestion of human blood samples

The automated in-solution sample preparation is conducted on the Bravo liquid

handling platform by implementing three protocol files sequentially. Protocol file 1: transferring serum samples into a 96-well sample plate. Protocol file 2: in-solution digestion process (denaturation, alkylation, proteolysis, and quench steps). Protocol file 3: spiking SIS peptides into digested samples.

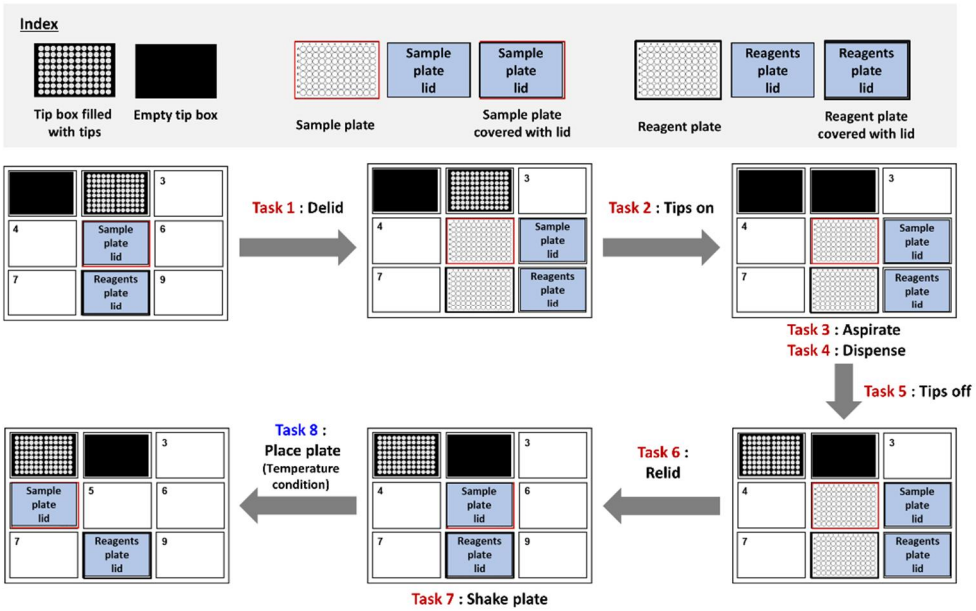


Figure 1-3. Sequential tasks required in the reagents dispensing procedure

Dispense of each reagent was conducted through seven sequential tasks (red fonts). Task 1: delid two plates. Task 2: tips on the head. Task 3: aspirate reagents. Task 4: dispense reagents into wells. Task 5: tips off from the head. Task 6: relid two plates. Task 7: shake sample plate. Task 8 (blue fonts, optional): place a plate at the thermal station.

Table 1-1. Total workflow of in-solution digestion protocol.

No.	Step	Reagents Component	Stock Concentration	Volume (µL)	Cumulative Volume (µL)	Temperature	Time
1	Adjust volume	HPLC water		3	3	25 °C	
2	Sample	Serum		2	5	25 °C	
3	Shake plate					25 °C	10 sec
4	Denaturation & Reduction	DTT, RapiGest	0.15% RapiGest, 15 mM DTT, 75 mM ABC, pH 8.0	10	15	25 °C	
5	Shake plate					25 °C	10 sec
6	Incubation					60 °C	60 min
7	Alkylation	IAA	50 mM IAA, 50 mM ABC	10	25	25 °C	
8	Shake plate					25 °C	10 sec
9	Incubation					25 °C	30 min (in dark reaction)
10	Tryptic digestion	Trypsin	0.3 µg/µL [1:50 (w/w)], 50 mM ABC	10	35	25 °C	
11	Shake plate					25 °C	10 sec
12	Incubation					37 °C	4 hours
13	Quenching	Formic Acid	4.50%	10	45	25 °C	
14	Shake plate					25 °C	10 sec
15	Incubation					37 °C	30 min

16	Spike SIS peptide mixture	SIS peptides	100 fmol/μL	5	50	25 °C	
17	Shake plate					25 °C	15 sec
18	Centrifugation					4 °C	30 min (at 15,000 rpm)

DTT, dithiothreitol; ABC, ammonium bicarbonate; IAA, iodoacetamide; SIS, stable isotope-labeled standard.

1.2.3. Optimization of aspiration height to minimize wasted reagents

The amount of reagent in the plate wells needed for reproducible aspiration of accurate volume depends on the distance between the end of the tips and the bottom of the wells. Thus, I determined the smallest such distance that minimizes the reagent volume to reduce the cost incurred by wasted reagents. The experiment was designed to test the accuracy and variation of three repeated aspiration steps by measuring the weight of the aspirated volume. This test was conducted with six aspiration parameter settings, encompassing all combinations of three distances from the bottom of the well (0.1, 0.3, and 0.5 mm) and two aspiration volumes (5 and 10 μL). The volume that remained in the plate well after aspiration was 0−25 μL at 5 μL intervals (**Figure 1-4**). I performed this experiment for two types of fluids: water with high surface tension and serum with high viscosity.

The accuracy and precision of the aspiration task were calculated as follows. First, I dispensed the fluids (water or serum) into wells from 5 to 30 μL at 5 μL intervals and repeated this step twice to establish three sets of six wells. Then, the 5 μL aspiration task was implemented for the three sets, adjusting the height of the tip that was mounted on the liquid handler to 0.1, 0.3, and 0.5 mm. The aspirated volume was transferred to an empty tube. After the fluids had been dispensed, the weights of the tubes that contained fluid were measured. The weight of the aspirated volume was calculated as the difference between the weights of the empty tube and fluid-containing tube. This procedure was repeated three times. The average CV value was calculated and represented the precision of the aspiration. The weight of the aspiration volume was divided by the theoretically expected weight and converted into a percentage to represent the accuracy of aspiration. This procedure was

repeated for the 10 μL aspiration, except the fluids were dispensed into the six wells in the range of 10−35 μL at 5 μL intervals.

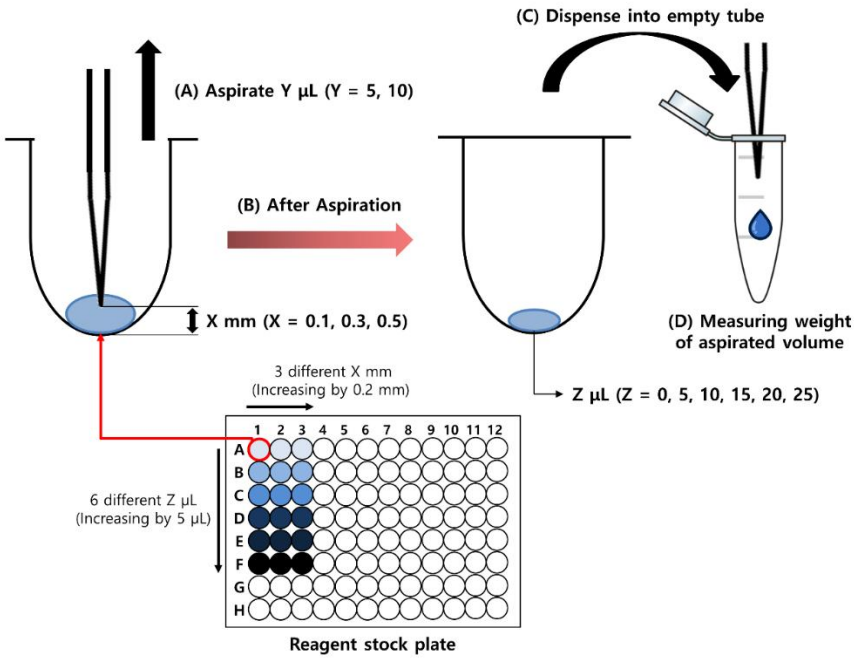


Figure 1-4. Three factors for optimizing the aspiration height to minimize wasted volumes.

Workflow of aspiration. Y μL (A) of fluid is aspirated from the well containing (Y + Z) μL of fluid at X mm (height) from the bottom of the well. After aspiration, Z μL of fluid remains in the well, and the aspirated fluid is transferred to an empty tube, previously measured for weight. The weight of the tube containing dispensed fluid is measured to calculate the weight of the fluid alone (D). The results from 3 replicates were used to optimize the combination of height (X mm) and the extra volume in the well (Z μL) after aspiration, with high reproducibility. I repeated this process for 2 types of fluid (water and serum).

1.2.4. Target peptides for MRM-MS assay

An MRM-MS assay of 52 peptides in identical pooled samples prepared by the automated sample preparation workflow was conducted to assess the feasibility of the process in the aspect of reproducibility of targeted quantification results. 52 peptides from 26 human serum proteins with varying hydrophobicities and molecular weights (50) were selected from the SRM Atlas (51) and CPTAC portal (52) as MRM−MS assay targets (**Figure 1-5**). All 52 peptides were detectable in a nondepleted normal pooled serum sample; Six transitions for each peptide were observed to confirm their detectability and to choose the best transition for each peptide. Transitions with high intensity and reproducibility (CV < 0.2) and low interference based on the AuDIT data were selected (53). Detailed information on the target peptides and transitions are presented in **Table 1-2**.

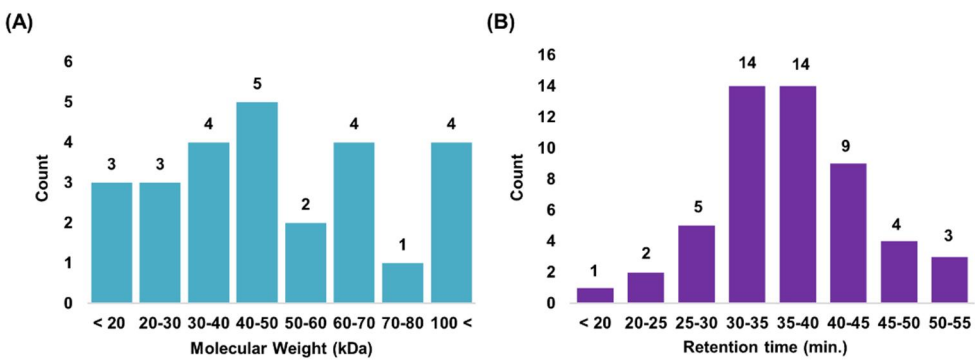


Figure 1-5. Distribution of molecular weights of target proteins and retention times for target peptides.

The molecular weights of target proteins were distributed evenly over the entire range (A). The diversity of the hydrophobicity of target peptides is shown as the diversity in retention time (B).

Table 1-2. Properties of target peptides

No.	Protein Name	Uniprot ID	MW (Da)	Peptide Sequence	RT	Precursor Ion		Product Ion			Result of AuDIT	
						m/z	Charge State	m/z	Charge State	Type	*p*	CV
1	**A1AT**	P01009	46,737	AVLTIDEK	28.32	444.8	2	718.4	1	y6	0.990	0.038
				AVLTIDEK*		448.8	2	726.4	1	y6		
				LSITGTYDLK	36.07	555.8	2	797.4	1	y7	0.247	0.057
				LSITGTYDLK*		559.8	2	805.4	1	y7		
2	**A2MG**	P01023	163,291	AIGYLNTGYQR	32.49	628.3	2	1071.5	1	y9	0.001	0.047
				AIGYLNTGYQR*		633.3	2	1081.5	1	y9		
3	**AMBP**	P02760	38,999	EYCGVPGDGDEELLR	36.99	854.9	2	1100.5	1	y10	0.676	0.090
				EYCGVPGDGDEELLR*		859.9	2	1110.5	1	y10		
				HHGPTITAK	15.50	321.2	3	319.2	1	y3	0.593	0.054
				HHGPTITAK*		323.9	3	327.2	1	y3		
				TVAACNLPIVR	34.33	607.3	2	1013.6	1	y9	0.638	0.100
				TVAACNLPIVR*		612.3	2	1023.6	1	y9		
4	**ANT3**	P01008	52,602	DDLYVSDAFHK	33.38	437.2	3	704.3	1	y6	0.150	0.089
				DDLYVSDAFHK*		439.9	3	712.4	1	y6		
				FATTFYQHLADSK	35.95	510.3	3	219.1	1	b2	0.509	0.003
				FATTFYQHLADSK*		512.9	3	219.1	1	b2		
				FDTISEK	24.08	420.2	2	692.3	1	y6	1.000	0.033

				FDTISEK*		424.2	2	700.4	1	y6			
				LPGIVAEGR	30.45	456.3	2	531.3	1	y5	0.913	0.150
				LPGIVAEGR*		461.3	2	541.3	1	y5		
				VAEGTQVLELPFK	43.94	715.9	2	391.2	1	y3	0.544	0.048
				VAEGTQVLELPFK*		719.9	2	399.2	1	y3		
5	**APOA1**	P02647	30,778	ATEHLSTLSEK	23.77	405.9	3	363.2	1	y3	0.873	0.137
				ATEHLSTLSEK*		408.6	3	371.2	1	y3		
				DLATVYVDVLK	45.39	618.3	2	936.5	1	y8	0.986	0.088
				DLATVYVDVLK*		622.4	2	944.6	1	y8		
				DYVSQFEGSALGK	39.66	700.8	2	279.1	1	b2	0.996	0.032
				DYVSQFEGSALGK*		704.8	2	279.1	1	b2		
				EQLGPVTQEFWDNLEK	48.93	645.0	3	804.4	1	y6	0.457	0.020
				EQLGPVTQEFWDNLEK*		647.7	3	812.4	1	y6		
6	**APOB**	P04114	515,605	FPEVDVLTK	38.60	524.3	2	803.5	1	y7	0.996	0.103
				FPEVDVLTK*		528.3	2	811.5	1	y7		
				GFEPTLEALFGK	50.24	654.8	2	975.6	1	y9	0.992	0.066
				GFEPTLEALFGK*		658.9	2	983.6	1	y9		
				SVGFHLPSR	31.25	333.9	3	472.3	1	y4	0.130	0.091
				SVGFHLPSR*		337.2	3	482.3	1	y4		
				TEVIPPLIENR	38.68	640.9	2	330.2	1	b3	0.712	0.013
				TEVIPPLIENR*		645.9	2	330.2	1	b3		

7	**APOH**	P02749	38,298	FICPLTGLWPINTLK	52.88	591.7	2	685.4	1	y6	0.181	0.030
				FICPLTGLWPINTLK*		594.3	2	693.4	1	y6		
				TFYEPGEEITYSCKPGYVSR	37.21	795.0	3	921.9	2	y16	1.000	0.148
				TFYEPGEEITYSCKPGYVSR*		798.4	3	926.9	2	y16		
8	**BTD**	P43251	61,133	VDLITFDTPFAGR	47.59	726.4	2	1011.5	1	y9	0.098	0.052
				VDLITFDTPFAGR*		731.4	2	1021.5	1	y9		
9	**CD14**	P08571	40,076	FPAIQNLALR	42.34	571.8	2	898.5	1	y8	0.081	0.061
				FPAIQNLALR*		576.8	2	908.6	1	y8		
10	**CHLE**	P06276	68,418	IFFPGVSEFGK	44.94	614.3	2	820.4	1	y8	0.277	0.071
				IFFPGVSEFGK*		618.3	2	828.4	1	y8		
11	**CO3**	P01024	187,148	IHWESASLLR	36.12	404.6	3	488.3	1	y4	0.875	0.097
				IHWESASLLR*		407.9	3	498.3	1	y4		
				TGLQEVEVK	28.16	501.8	2	731.4	1	y6	0.768	0.060
				TGLQEVEVK*		505.8	2	739.4	1	y6		
				VHQYFNVELIQPGAVK	41.86	614.7	3	471.3	1	y5	0.973	0.046
				VHQYFNVELIQPGAVK*		617.3	3	479.3	1	y5		
12	**CYTC**	P01034	15,799	ALDFAVGEYNK	36.84	613.8	2	1042.5	1	y9	0.064	0.172
				ALDFAVGEYNK*		617.8	2	1050.5	1	y9		
13	**FA10**	P00742	54,732	NTEQEEGGEAVHEVEVVIK	34.50	699.3	3	216.1	1	b2	0.923	0.117
				NTEQEEGGEAVHEVEVVIK*		702.0	3	216.1	1	b2		
14	**FA5**	P12259	251,703	SQHLDNFSNQIGK	28.23	496.6	3	646.4	1	y6	0.103	0.076

				SQHLDNFSNQIGK*		499.3	3	654.4	1	y6		
15	**FA9**	P00740	51,778	NCELDVTCNIK	31.65	683.3	2	275.1	1	b2	0.937	0.052
				NCELDVTCNIK*		687.3	2	275.1	1	b2		
				SALVLQYLR	41.70	531.8	2	692.4	1	y5	0.321	0.120
				SALVLQYLR*		536.8	2	702.4	1	y5		
16	**HPT**	P00738	45,205	TEGDGVYTLNNEK	28.03	720.3	2	881.4	1	y7	0.333	0.033
				TEGDGVYTLNNEK*		724.3	2	889.5	1	y7		
				VGYVSGWGR	31.64	490.8	2	562.3	1	y5	0.973	0.066
				VGYVSGWGR*		495.8	2	572.3	1	y5		
				VVLHPNYSQVDIGLIK	41.12	599.0	3	658.4	1	y6	0.641	0.019
				VVLHPNYSQVDIGLIK*		601.7	3	666.4	1	y6		
17	**IC1**	P05155	55,154	GVTSVSQIFHSPDLAIR	43.82	609.7	3	157.1	1	b2	0.951	0.104
				GVTSVSQIFHSPDLAIR*		613.0	3	157.1	1	b2		
				TNLESILSYPK	42.55	632.8	2	216.1	1	b2	0.912	0.010
				TNLESILSYPK*		636.8	2	216.1	1	b2		
18	**PLF4**	P02776	10,845	ICLDLQAPLYK	42.91	667.4	2	1060.6	1	y9	0.754	0.084
				ICLDLQAPLYK*		671.4	2	1068.6	1	y9		
19	**PROC**	P04070	52,071	LGDDLLQCHPAVK	33.24	489.3	3	839.4	1	y7	0.553	0.089
				LGDDLLQCHPAVK*		491.9	3	847.4	1	y7		
20	**PROS**	P07225	75,123	FSAEFDFR	38.56	509.7	2	784.4	1	y6	0.894	0.007
				FSAEFDFR*		514.7	2	794.4	1	y6		

				IETISHEDLQR	26.59	447.6	3	884.4	1	y7	0.613	0.094
				IETISHEDLQR*		450.9	3	894.4	1	y7		
21	**SHBG**	P04278	43,779	IALGGLLFPASNLR	50.64	721.4	2	804.4	1	y7	0.745	0.106
				IALGGLLFPASNLR*		726.4	2	814.4	1	y7		
				QAEISASAPTSLR	30.43	665.9	2	889.5	1	y9	0.039	0.166
				QAEISASAPTSLR*		670.9	2	899.5	1	y9		
22	**TFR1**	P02786	84,871	ILNIFGVIK	49.00	508.8	2	790.5	1	y7	0.943	0.069
				ILNIFGVIK*		512.8	2	798.5	1	y7		
23	**THRB**	P00734	70,037	ELLESYIDGR	38.44	597.8	2	839.4	1	y7	0.074	0.107
				ELLESYIDGR*		602.8	2	849.4	1	y7		
				SEGSSVNLSPPLEQCVPDR	37.58	691.0	3	605.8	2	y10	0.052	0.057
				SEGSSVNLSPPLEQCVPDR*		694.3	3	610.8	2	y10		
				TATSEYQTFFNPR	38.50	781.4	2	272.2	1	y2	0.710	0.086
				TATSEYQTFFNPR*		786.4	2	282.2	1	y2		
24	**TRFE**	P02787	77,064	DGAGDVAFVK	30.16	489.7	2	735.4	1	y7	0.932	0.022
				DGAGDVAFVK*		493.8	2	743.4	1	y7		
				EFQLFSSPHGK	33.83	426.2	3	612.3	1	y6	0.812	0.023
				EFQLFSSPHGK*		428.9	3	620.3	1	y6		
				YLGEEYVK	31.57	500.8	2	724.4	1	y6	0.599	0.037
				YLGEEYVK*		504.8	2	732.4	1	y6		
25	**TRFL**	P02788	78,182	YYGYTGAFR	32.91	549.3	2	771.4	1	y7	0.701	0.160

				YYGYTGAFR*		554.3	2	781.4	1	y7		
26	**TTHY**	P02766	15,887	AADDTWEPFASGK	36.98	697.8	2	921.4	1	y8	0.900	0.026
				AADDTWEPFASGK*		701.8	2	929.5	1	y8		

The molecular weight of a protein is shown as the value of isoform 1 when it has multiple isoforms.

The heavy-labeled lysine (Lys, K) or arginine (Arg, R) is asterisked (*).

The automated detection of inaccurate and imprecise transitions (AuDIT) algorithm was used to determine whether the highest transition that I used to identify each peptide target was affected by matrix interference. The results of AuDIT are shown in last 2 columns, and all transitions were marked as having “good” status.

Abbraviations: m/z, mass-to-charge ratio. AuDIT, automated detection of inaccurate and imprecise transitions. CV, coefficient of variation.

1.2.5. Quantitative MRM−MS analysis

Target peptides were quantified on an Agilent 6490 triple quadrupole (QQQ) mass spectrometer (Agilent) with a Jetstream electrospray source that was equipped with a 1260 Infinity HPLC system (Agilent). Ten microliters of tryptic peptides were injected into a guard column (2.1 mm id ×30 mm, 1.8 μm particle size, Agilent Zorbax SB-C18) for online desalting and then passed through to a reverse-phase analytical column (150 mm × 0.5 mm id, 3.5 μm particle size, Agilent Zorbax SB-C18) that was maintained at 40 °C. The total run time for the liquid chromatography (LC) step was 70 min. Mobile phases A (water 0.1% (v/v) formic acid) and B (acetonitrile 0.1% (v/v) formic acid) were used to create a binary gradient of 3 to 35% acetonitrile/0.1% formic acid through the column for 45 min at 10 μL/min to separate the peptides. The column was equilibrated for the next run by ramping the gradient to 70% B for 5 min and 3% B for 10 min.

MassHunter (vB06.01, Agilent Technologies) was used to establish an MRM−MS scheduling method and control the LC−MS system for data acquisition. The ion spray capillary voltage and nozzle voltage were 2500 and 2000 V, respectively. The temperature and flow rate of the gas were set to 250 °C at 15 L/min for the drying gas and 350 °C at 12 L/min for the sheath gas. The voltage of the cell accelerator was adjusted to 5 V, the fragment voltage was 380 V, and the delta electron multiplier voltage (EMV) was 200 V. A total of three batches were sequentially analyzed in order of preparation, whereas samples in each batch were analyzed in random sequence. All raw data files from the MRM−MS were imported into and aligned by Skyline (McCoss Lab, University of Washington) for quantitative analysis (54).

1.2.6. Evaluation of feasibility for automated in-solution digestion workflow

As shown in **Figure 1-6**, the identical pooled serum samples were prepared by the automated sample preparation process established in the study. To evaluate the possibility of obtaining significant reproducibility when introducing automation, the same assay was performed over three consecutive days. Also, I performed a systematic evaluation of digestion for further optimization by exploring the possibility of re-using pipette tips for reagent dispense tasks to reduce cost (**Figure 1-7**). A total of 96 identical samples was divided into four sets of 24 samples, according to the number of wells that were to be dispensed with a tip. Thus Set 1 used 24 tips for transferring each reagent to 24 wells (1 well/tip), Set 2 used 8 tips (3 wells/tip), Set 3 used 4 tips (6 wells/tip), and Set 4 used 1 tip (24 wells/tip). In Sets 1 and 3, the 24 samples were arranged in 4 × 6 format, whereas the 24 samples were arranged in 3 × 8 format for Sets 2 and 4. In each set, the volume that was aspirated in each tip included an additional 5 μL of reagent that was successively dispensed in each well.

In Set 1, which is the standard method, the reagents were aspirated with 24 tips and dispensed into 24 wells simultaneously. For Set 2, eight tips were mounted simultaneously and aspirated the reagents that were to be dispensed to three wells by each tip. The reagents were then dispensed into eight wells at the same time. This procedure was repeated two additional times to dispense the reagents into 24 wells. Similarly, for Set 3, four tips were equipped concurrently and aspirated the reagents that were to be dispensed to six wells by each tip. Then, the reagent was dispensed into four wells at the same time. This process was repeated five additional times for

dispensing into 24 wells. In the experiment for Set 4, one tip was equipped and aspirated the reagents that were to be dispensed to 24 wells. The reagent was dispensed into one well. This procedure was repeated 23 additional times by advancing it sequentially toward the adjacent well, resulting in 24 wells that received reagent. To avoid cross-contamination, the liquid dispensing heights in each task were adjusted such that the tips never touch the sample plate or the surface.

The Set 1 experiment used the most tips, and thus its sample preparation was the fastest and most accurate, but its cost for consumables was also the highest. For this reason, I examined cost-effective methods of using fewer tips while maintaining comparable precision levels with Set 1.

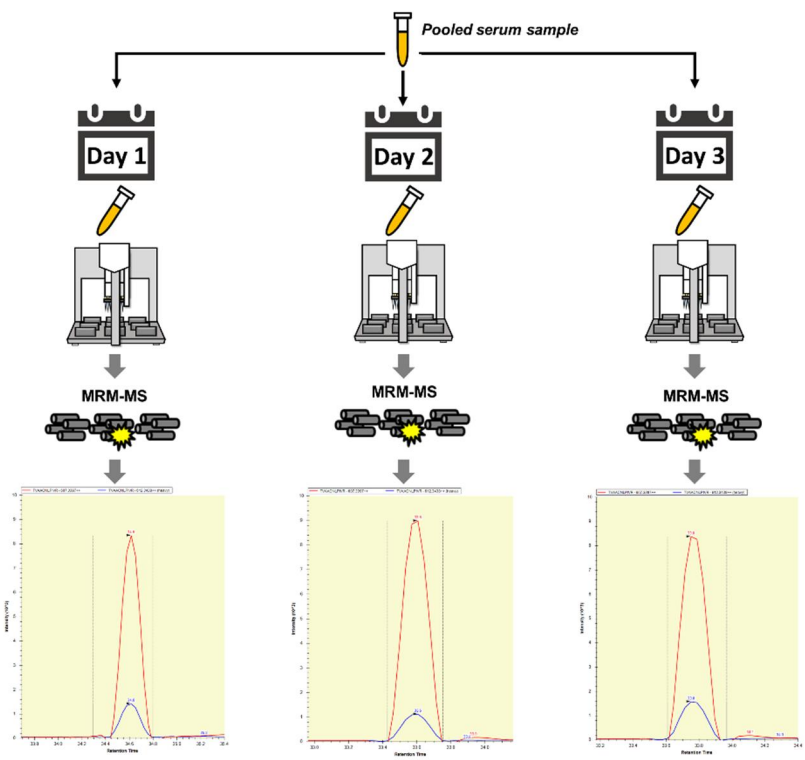


Figure 1-6. Experimental scheme for evaluation of feasibility of automated in-solution digestion process.

The automated sample preparation process established in the study was evaluated in aspects of reproducibility obtained over three successive days. The identical pooling serum samples were prepared by the automated process and analyzed by MRM-MS assay. The quantification results were assessed whether the automated workflow could achieve reasonable levels of variations.

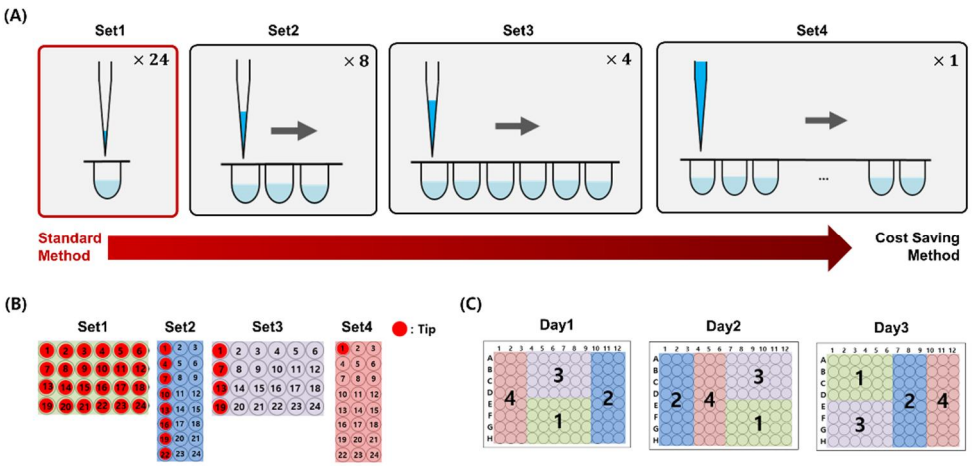


Figure 1-7. A systematic evaluation of feasibility for automated digestion process and further optimization through exploring the possibility of re-using pipette tips to reduce cost

(A) Set 1 is the standard method dispensing reagents to 24 samples with individual tips. The other three sets are cost-saving methods of dispensing reagents with re-used tips. Dispensing reagents began from the first column and moved to the right side of each dataset in Sets 2 and 3. Set 4 dispensed reagents from the first to last well sequentially (the well number in Set 4 corresponds to the dispensed order) (B) The 96-well plate was divided into 4 sets of 24 identical pooled serum samples for dispensing reagents with varying numbers of tips. Set 1 (green) was dispensed using 24 tips, versus 8, 4, and 1 tip for sets 2 (blue), 3 (purple), and 4 (pink). The red circle

shows the position of the tips at the initial dispensing in each dataset. (C) The entire workflow was replicated over three days, wherein the location of the four sets in a plate varied daily. The plate compartment varied on each day to exclude compound effects from positional differences in one plate. The green box represents Set 1, and the blue box represent Set 2. Sets 3 and 4 are represented by a purple and pink box, respectively. Finally, the variation of each of the 4 sets was calculated and compared with each other.

1.2.7. Assess reproducibility of the automated sample preparation process

The peak area ratio (PAR) was calculated by dividing the peak areas of endogenous peptides by those of the SIS peptides for each peptide in all individual samples. The results over the 3 days were used to analyze CVs to compare the variability according to the number of tips that were used to prepare 24 samples. The PAR values for the 24 samples that were prepared daily were averaged, and the results for the 3 days were calculated as the average CV value to represent the intraday variation of each data set (**Figure 1-8 (A)**). The interday variation was expressed as the CV of the 24 PAR values that were obtained using the average PAR value over the 3 days from the same well location in the plate (**Figure 1-8 (B)**).

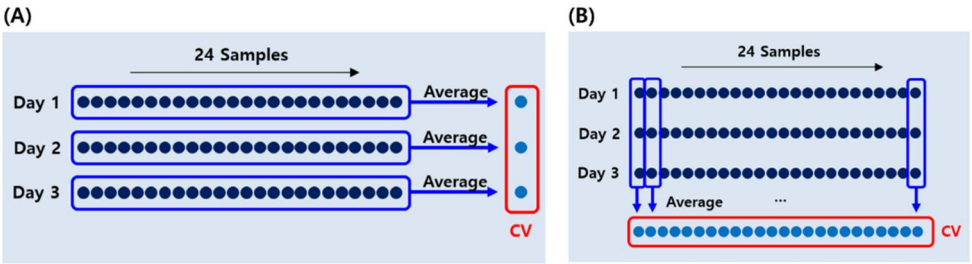


Figure 1-8. Visualization of the formation to calculate intra- and inter-CV.

The 24 navy circles represent samples prepared by one dispensing method on each day. Intraday variation was calculated as the coefficient of variation (CV) of three average values of 24 samples on each day for each dataset (A). The samples prepared on different days and with the same well position in 1 dataset were averaged. The variation between the 24 averaged values over 3 days was calculated as the interday variation (B).

1.3. RESULTS

1.3.1. Optimization of aspiration height

I optimized the distance between the ends of the tips and the bottom of the well to minimize the cost of the volume that remained after aspiration. A test was conducted to determine the minimum distance and extra volume for which an accurate volume could be reproducibly aspirated into the tip. As shown in **Figure 1-9**, an accurate volume could be reproducibly aspirated 0.1 mm from the bottom of the well when it contained at least 5 μL of fluid as extra reagent, regardless of the fluid type or the aspiration amount. As the height of aspiration rose by 0.2 mm, an extra 5 μL of fluid was needed to reproducibly aspirate an accurate volume. On the basis of these results, the protocol was set to aspirate each reagent at 0.1 mm from the bottom of the well, containing an extra 5 μL over the total aspiration volume to minimize cost.

The total volume that was aspirated in the tip was set to contain an additional 5 μL of reagent volume, considering the loss of reagent that remained in the tip after the dispensing (See **Table 1-3**). Thus, greater tip consumption accompanies the increased use of reagents to supply this extra volume in each well of the reagent stock plates. The aspiration of water was generally less accurate and reproducible than that of serum samples, due to the higher surface tension and static repulsion of the former (55, 56). Thus, when aspirating a fluid with high surface tension and relatively low cost, such as water, over 5 μL of fluid should be placed in the plate well.

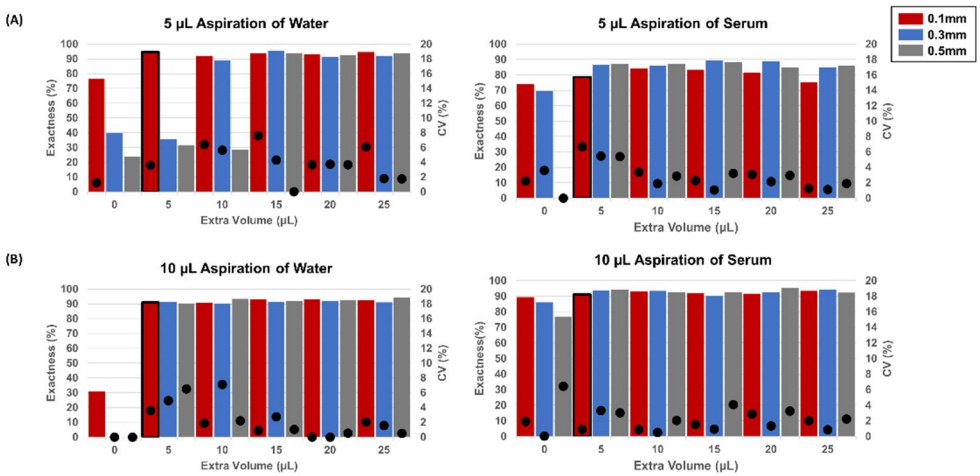


Figure 1-9. Optimization of aspiration height for water and serum.

The accuracy (bar graphs) and reproducibility (point graphs) were calculated for each dataset according to aspirated volume: 5 µL (A) and 10 µL (B). The exactness of the aspiration is calculated by comparing the measured weight of the aspirated volume with its expected weight. The reproducibility is presented as CV values (points with CV values over 20% are not shown). (A) The graphs on the left represent the 5 µL of water aspirated from the well containing 5 µL plus extra reagent volume (x-axis). The graphs on the right represent the 5 µL of serum aspirated from the well containing 5 µL plus extra reagent volume. (B) The graphs on the left represent the 10 µL of water aspirated from the well containing 10 µL plus extra reagent volume (x-axis). The graphs on the right represent the 10 µL of serum aspirated from the well containing 10 µL plus extra reagent volume. The results of aspiration at 0.1 mm from the bottom of the well (red color) showed reasonable exactness and reproducibility when the plate well contained at least 5 µL of fluid as extra reagent. The minimum extra volume increased to 10 µL as the height of aspiration rose to 0.3 mm (blue). The most cost-consuming height was 0.5 mm (gray), requiring at least 15 µL of additional reagent for reproducible aspiration of an accurate volume. After

considering all combination of parameters synthetically, aspiration at 0.1 mm from the well containing an additional 5 µL with aspiration volume (red bar with edge) is the ideal means of reducing the wasted volume, maintaining high exactness and low variation of aspiration. The optimization process is detailed in **Figure 1-4**.

Table 1-3. Volumes needed when dispensing 10 µL of reagents for each set.

	# of Tips	Well/Tip	Dispensed volume (µL)[a]	Aspiration volume (µL)[b]	Well volume (µL)[c]	Total extra volume (µL)[d]
Set 1	24	1	10	15	20	240
Set 2	8	3	10	35	40	80
Set 3	4	6	10	65	70	40
Set 4	1	24	10	245	250	10

[a] Volume dispensed to each sample well.

[b] Volume aspirated in each tip.

[c] Minimum volume of reagent in each well considering the aspiration height.

[d] Volume disposed after dispensing.

1.3.2. Reproducibility assessment of the automated in-solution digestion process

The robustness of the automated workflow was evaluated by the reproducibility of results for all samples prepared daily. The CV values of standard method (Set 1, 1 well/tip) for the 52 peptides are shown in **Figure 1-10**. The intra-, inter-, and total-CV values were under 20% for all peptides, except for seven peptides that had high variation (total-CV >20%). Thus, the standard method (Set 1, 1 well/tip) had reasonable reproducibility, with an average intra-CV of 7.9%, an average inter-CV of 12.0%, and an average total-CV of 15.3% (**Figure 1-10 (D)**).

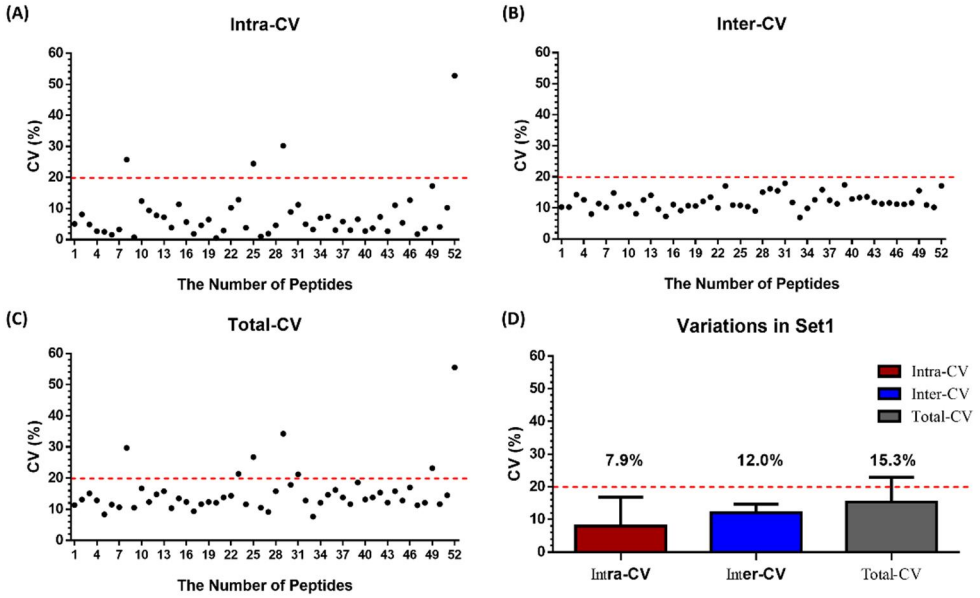


Figure 1-10. Variations of the automated sample preparation process

The variations in the 52 peptides in Set 1 (standard method) are shown. (A) The 52 peptides showed intra-CV values under 20%, except for four outliers. (B) All the 52 peptides showed inter-CV values under 20%. (C) The 52 peptides showed total-CV values under 20%, except for seven outliers with a CV of >20%. (D) The graph was plotted by mean with SD of CV values of 52 peptides. The average CV values of 52 peptides was 7.9% (red bar graph) for intra-CV, 12.0% (blue bar graph) for inter-CV, and 15.3% (gray bar graph) for total-CV. CV, coefficient of variation. SD, standard deviation.

1.3.3. Evaluate the possibility of cost-saving automated processes

The possibility of reducing experimental costs was evaluated by comparing variations obtained from other three cost-saving methods that reduced consumables (**Figure 1-11**). Total variation results are detailed in **Table 1-4**. The CVs of Sets 2

and 3 had equivalent intra- and inter-CV as Set 1. The average intra-CV and inter-CV were 8.5 and 13.5%, respectively, for Set 2 and 9.0 and 13.5% for Set 3. Thus, reducing the number of tips is a means of minimizing the cost incurred by expensive consumables of automation while maintaining reproducibility.

Set 4, which was prepared with only one tip for 24 wells, had an average intra-CV of 24.2% and an average inter-CV of 26.4%, higher than in the other three sets, likely due to the inaccurate dispensing of each reagent, resulting from the long time that it took to dispense into all 24 wells with one tip (**Figure 1-12**). It took nearly twice as much time to complete the dispensing to the 24 wells, whereas this time differed slightly between the other three sets.

The PAR of each peptide is shown as the average value of 24 samples for each set in **Table 1-5**. I compared the quantification results for each set with those of Set 1, the standard experiment with regard to bias values. The bias indicates that the quantification results of Sets 2 (0.8% on average) and 3 (−3.4% on average) are similar to those of Set 1. Set 4 showed higher bias than the other sets (−15.5% on average). These differences appear to be attributed to the difference in digestion efficiency for each set because SIS peptide was spiked equally in all sets. The digestion efficiency was generally lower in Sets 3 and 4. The quantification results for Set 2 were nearly equivalent to those of Set 1 based on their low bias.

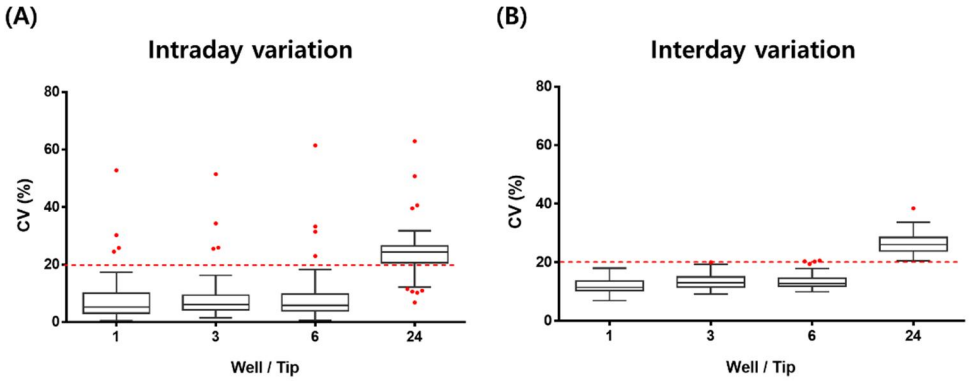


Figure 1-11. Box and whisker plots of intraday and interday variation testing for each set.

Each CV value was calculated for 52 peptides in each dataset. Experiments for Sets 1, 2, 3, and 4 were conducted with 24, 8, 4, and 1 tip for dispensing to the 24 wells, respectively. The well/tip on the x-axis was calculated for each dataset, dividing the number of wells by the number of tips used for dispensing the reagent. Sets 1 (1 well/tip), 2 (3 well/tip), and 3 (6 well/tip) had similar levels of intraday variation, with average intra-CVs lower than 15% (7.9%, 8.5%, 9.0%, respectively), whereas Set 4 (24 well/tip) had a higher average intra-CV (24.2%) (A). Also, the inter-CVs were similar in Sets 1–3 (B). The average inter-CVs were lower than 15% in the first 3 sets (12.0% for Set 1, 13.5% for Set 2, 13.5% for Set3), whereas Set 4 had an inter-CV of 26.4%.

Table 1-4. Variation in each set for 52 target peptides.

No.	Protein Name	Peptide Sequence	Set1			Set2			Set3			Set4		
			Intra CV	Inter CV	Total CV	Intra CV	Inter CV	Total CV	Intra CV	Inter CV	Total CV	Intra CV	Inter CV	Total CV
1	**A1AT**	AVLTIDEK	5.1	10.3	**11.4**	3.9	10.9	**11.6**	0.6	12.1	**12.2**	29.3	22.2	**36.7**
		LSITGTYDLK	8.1	10.3	**13.1**	5.9	12.2	**13.6**	3.9	11.9	**12.5**	26.6	23.7	**35.6**
2	**A2MG**	AIGYLNTGYQR	4.9	14.3	**15.1**	6.2	12.8	**14.3**	5.1	13.3	**14.2**	24.5	27.3	**36.7**
3	**AMBP**	EYCGVPGDGDEELLR	2.7	12.6	**12.9**	4.3	14.2	**14.9**	4.3	13.0	**13.7**	25.3	25.7	**36.1**
		HHGPTITAK	2.5	8.0	**8.4**	6.6	11.3	**13.1**	8.4	12.0	**14.6**	22.5	25.8	**34.2**
		TVAACNLPIVR	1.6	11.4	**11.5**	2.6	11.8	**12.1**	2.4	12.1	**12.3**	28.0	25.4	**37.8**
4	**ANT3**	DDLYVSDAFHK	3.3	10.2	**10.7**	4.1	11.5	**12.2**	6.6	14.2	**15.6**	22.8	22.4	**32.0**
		FATTFYQHLADSK	25.8	14.8	**29.7**	25.9	16.7	**30.8**	22.9	12.6	**26.2**	40.6	24.4	**47.4**
		FDTISEK	0.8	10.4	**10.5**	3.7	10.8	**11.4**	5.4	12.5	**13.6**	25.9	23.5	**35.0**
		LPGIVAEGR	12.4	11.1	**16.7**	10.9	11.0	**15.5**	8.2	12.1	**14.6**	10.6	23.6	**25.9**
		VAEGTQVLELPFK	9.4	8.1	**12.4**	9.4	9.5	**13.4**	3.8	10.9	**11.5**	16.1	24.4	**29.2**
5	**APOA1**	ATEHLSTLSEK	7.8	12.6	**14.8**	5.1	13.7	**14.6**	9.3	10.7	**14.2**	28.2	33.6	**43.9**
		DLATVYVDVLK	7.2	14.1	**15.8**	9.2	16.8	**19.1**	11.2	15.4	**19.0**	19.9	26.5	**33.1**
		DYVSQFEGSALGK	3.9	9.6	**10.4**	4.7	13.8	**14.6**	6.3	11.3	**12.9**	19.2	26.3	**32.6**
		EQLGPVTQEFWDNLEK	11.4	7.3	**13.5**	10.2	11.7	**15.5**	9.9	13.3	**16.6**	16.7	24.5	**29.6**
6	**APOB**	FPEVDVLTK	5.7	11.1	**12.4**	4.2	13.0	**13.6**	6.4	13.6	**15.0**	21.8	29.0	**36.3**
		GFEPTLEALFGK	1.9	9.2	**9.4**	3.7	12.8	**13.4**	3.2	11.8	**12.3**	25.2	24.1	**34.9**

		SVGFHLPSR	4.6	10.7	**11.7**	5.4	12.9	**14.0**	4.2	12.0	**12.7**	25.4	24.5	**35.3**
		TEVIPPLIENR	6.5	10.6	**12.4**	5.9	12.6	**13.9**	4.1	12.4	**13.1**	26.0	23.7	**35.2**
7	**APOH**	FICPLTGLWPINTLK	0.5	12.1	**12.1**	5.6	13.3	**14.5**	5.1	12.8	**13.8**	23.6	24.3	**33.8**
		TFYEPGEEITYSCKPGYVSR	2.9	13.5	**13.8**	7.2	17.5	**19.0**	7.1	13.6	**15.3**	25.0	32.1	**40.7**
8	**BTD**	VDLITFDTPFAGR	10.3	10.0	**14.4**	10.7	9.1	**14.1**	6.9	11.8	**13.6**	16.4	22.5	**27.9**
9	**CD14**	FPAIQNLALR	12.9	17.1	**21.4**	5.8	20.0	**20.8**	17.6	20.3	**26.8**	10.9	38.5	**40.0**
10	**CHLE**	IFFPGVSEFGK	3.8	10.9	**11.6**	3.5	11.1	**11.6**	3.9	12.7	**13.3**	21.5	25.4	**33.2**
11	**CO3**	IHWESASLLR	24.5	10.8	**26.8**	25.5	11.6	**28.0**	33.2	14.5	**36.3**	39.5	20.4	**44.5**
		TGLQEVEVK	1.0	10.4	**10.5**	2.5	10.8	**11.1**	3.0	10.6	**11.0**	28.3	22.7	**36.2**
		VHQYFNVELIQPGAVK	1.9	9.0	**9.2**	2.7	12.6	**12.8**	4.3	11.4	**12.2**	25.4	25.7	**36.2**
12	**CYTC**	ALDFAVGEYNK	4.6	15.1	**15.8**	9.5	16.9	**19.4**	5.5	20.6	**21.3**	21.4	29.1	**36.1**
13	**FA10**	NTEQEEGGEAVHEVEVVIK	30.2	16.2	**34.3**	34.3	17.8	**38.6**	31.4	19.4	**36.9**	50.7	21.1	**54.9**
14	**FA5**	SQHLDNFSNQIGK	8.9	15.5	**17.9**	9.1	11.6	**14.7**	7.1	16.0	**17.5**	17.0	26.5	**31.5**
15	**FA9**	NCELDVTCNIK	11.2	17.9	**21.2**	7.8	19.3	**20.8**	6.0	20.3	**21.2**	22.8	28.3	**36.3**
		SALVLQYLR	5.0	11.8	**12.8**	7.7	13.2	**15.2**	15.3	13.8	**20.6**	24.4	26.4	**36.0**
16	**HPT**	TEGDGVYTLNNEK	3.3	6.9	**7.7**	2.1	9.9	**10.1**	1.5	11.3	**11.4**	20.3	27.3	**34.0**
		VGYVSGWGR	6.9	9.9	**12.1**	7.5	14.0	**15.9**	2.2	13.8	**14.0**	24.2	28.5	**37.4**
		VVLHPNYSQVDIGLIK	7.5	12.6	**14.7**	5.7	13.3	**14.5**	5.2	13.4	**14.4**	27.3	25.9	**37.6**
17	**IC1**	GVTSVSQIFHSPDLAIR	3.1	15.9	**16.2**	9.6	13.3	**16.4**	11.0	14.7	**18.4**	10.2	27.2	**29.0**
		TNLESILSYPK	5.8	12.5	**13.8**	3.3	13.5	**13.9**	4.0	10.5	**11.2**	12.1	23.9	**26.8**
18	**PLF4**	ICLDLQAPLYK	3.1	11.3	**11.7**	4.4	11.6	**12.4**	4.0	11.1	**11.8**	23.9	23.7	**33.7**

19	**PROC**	LGDDLLQCHPAVK	6.6	17.4	**18.6**	5.3	17.8	**18.6**	8.4	17.5	**19.5**	24.5	31.3	**39.8**
20	**PROS**	FSAEFDFR	2.8	12.9	**13.2**	8.1	11.9	**14.4**	3.3	12.6	**13.0**	23.7	29.6	**37.9**
		IETISHEDLQR	3.7	13.3	**13.8**	7.2	13.0	**14.9**	5.6	11.0	**12.4**	22.5	28.7	**36.5**
21	**SHBG**	IALGGLLFPASNLR	7.3	13.6	**15.4**	8.5	16.4	**18.5**	3.7	16.1	**16.5**	6.8	23.0	**24.0**
		QAEISASAPTSLR	2.7	11.8	**12.2**	6.8	17.3	**18.6**	7.7	16.9	**18.6**	24.5	29.8	**38.6**
22	**TFR1**	ILNIFGVIK	11.1	11.3	**15.8**	1.4	14.0	**14.1**	11.7	13.0	**17.5**	11.5	29.6	**31.8**
23	**THRB**	ELLESYIDGR	5.4	11.6	**12.8**	6.4	12.7	**14.3**	11.0	14.3	**18.0**	22.5	28.7	**36.5**
		SEGSSVNLSPPLEQCVPDR	12.7	11.2	**17.0**	11.7	10.6	**15.7**	16.9	12.5	**21.0**	29.1	28.0	**40.4**
		TATSEYQTFFNPR	1.8	11.2	**11.3**	1.4	15.1	**15.2**	3.9	13.6	**14.1**	24.6	27.4	**36.8**
24	**TRFE**	DGAGDVAFVK	3.6	11.6	**12.1**	1.5	13.4	**13.5**	3.8	9.9	**10.6**	25.9	25.6	**36.4**
		EFQLFSSPHGK	17.3	15.6	**23.2**	16.2	16.3	**23.0**	18.3	11.5	**21.6**	31.8	26.1	**41.1**
		YLGEEYVK	4.1	11.0	**11.7**	4.7	11.7	**12.7**	2.8	10.9	**11.3**	21.8	26.9	**34.7**
25	**TRFL**	YYGYTGAFR	10.3	10.2	**14.5**	12.3	18.4	**22.2**	6.6	15.4	**16.7**	28.9	28.0	**40.3**
26	**TTHY**	AADDTWEPFASGK	52.8	17.1	**55.5**	51.4	13.3	**53.1**	61.4	17.8	**64.0**	62.9	29.1	**69.3**

Total CV = $\sqrt{\text{IntraCV}^2 + \text{InterCV}^2}$.

Total CV values over 20% are underlined in the column for Set 1. The 7 peptide targets have lower reproducibility in all sets.

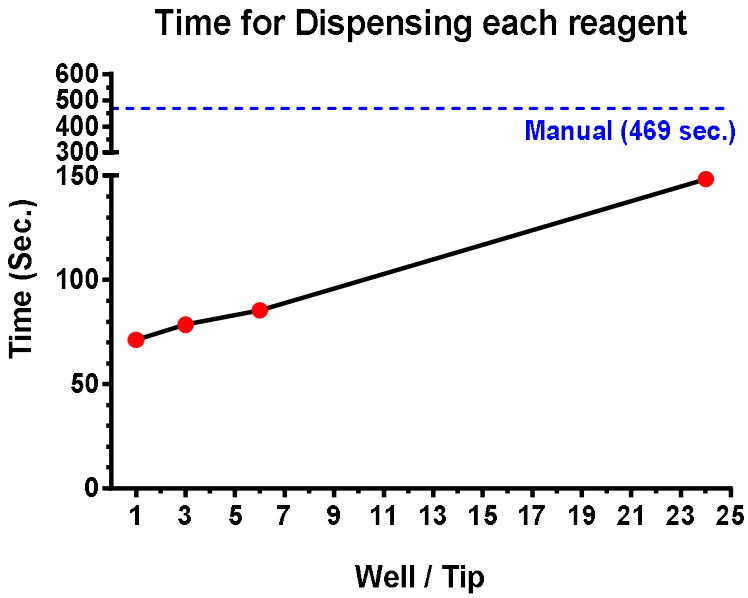


Figure 1-12. Time spent for dispensing 10 µL of reagent to every 24 wells.

The well/tip on the x-axis was calculated, dividing the number of wells by the number of tips used for dispensing the reagent. The total time required for dispensing 10 µL of reagent to all 24 wells is presented for each set. Dispensing a reagent takes more time when fewer tips are used for the same amounts of wells. The blue line presents the time required for dispensing 10 µL of reagent to all 24 wells manually.

Table 1-5. Quantification of each set for 52 target peptides.

No.	Protein Name	Peptide Sequence	Set1		Set2		Set3		Set4	
			PAR[a] (L/H)	Bias[b] (%)	PAR[a] (L/H)	Bias[b] (%)	PAR[a] (L/H)	Bias[b] (%)	PAR[a] (L/H)	Bias[b] (%)
1	**A1AT**	AVLTIDEK	**68.42**	0.0	**68.58**	0.2	**66.91**	-2.2	**56.65**	-17.2
		LSITGTYDLK	**119.99**	0.0	**124.73**	4.0	**118.14**	-1.5	**100.71**	-16.1
2	**A2MG**	AIGYLNTGYQR	**20.63**	0.0	**20.54**	-0.5	**20.11**	-2.6	**16.78**	-18.7
3	**AMBP**	EYCGVPGDGDEELLR	**6.01**	0.0	**6.14**	2.1	**5.83**	-3.1	**4.93**	-18.0
		HHGPTITAK	**2.62**	0.0	**2.61**	-0.3	**2.53**	-3.3	**2.14**	-18.4
		TVAACNLPIVR	**6.57**	0.0	**6.57**	0.1	**6.28**	-4.3	**5.29**	-19.5
4	**ANT3**	DDLYVSDAFHK	**2.70**	0.0	**2.75**	1.8	**2.71**	0.4	**2.45**	-9.1
		FATTFYQHLADSK	**7.09**	0.0	**7.56**	6.5	**6.83**	-3.7	**5.84**	-17.6
		FDTISEK	**1.63**	0.0	**1.62**	-0.6	**1.56**	-4.7	**1.39**	-15.2
		LPGIVAEGR	**1.08**	0.0	**1.08**	-0.2	**1.11**	2.7	**1.07**	-1.2
		VAEGTQVLELPFK	**1.34**	0.0	**1.39**	3.6	**1.34**	-0.2	**1.31**	-2.1
5	**APOA1**	ATEHLSTLSEK	**73.46**	0.0	**75.47**	2.7	**68.80**	-6.4	**61.82**	-15.9
		DLATVYVDVLK	**84.41**	0.0	**87.12**	3.2	**78.87**	-6.6	**73.78**	-12.6
		DYVSQFEGSALGK	**58.82**	0.0	**59.31**	0.8	**56.88**	-3.3	**49.16**	-16.4
		EQLGPVTQEFWDNLEK	**20.98**	0.0	**21.05**	0.3	**20.27**	-3.4	**19.31**	-7.9
6	**APOB**	FPEVDVLTK	**2.09**	0.0	**2.10**	0.8	**1.98**	-5.1	**1.64**	-21.6
		GFEPTLEALFGK	**2.61**	0.0	**2.63**	0.7	**2.52**	-3.5	**2.14**	-18.2

		SVGFHLPSR	**1.91**	0.0	**1.93**	1.2	**1.83**	-4.3	**1.55**	-18.6
		TEVIPPLIENR	**5.41**	0.0	**5.41**	0.1	**5.12**	-5.3	**4.31**	-20.3
7	**APOH**	FICPLTGLWPINTLK	**7.97**	0.0	**8.14**	2.2	**7.60**	-4.6	**6.78**	-15.0
		TFYEPGEEITYSCKPGYVSR	**1.51**	0.0	**1.51**	-0.3	**1.46**	-3.7	**1.15**	-24.1
8	**BTD**	VDLITFDTPFAGR	**0.22**	0.0	**0.21**	-2.3	**0.20**	-7.9	**0.20**	-9.3
9	**CD14**	FPAIQNLALR	**0.04**	0.0	**0.04**	6.5	**0.04**	-5.0	**0.04**	-2.3
10	**CHLE**	IFFPGVSEFGK	**0.06**	0.0	**0.06**	0.0	**0.06**	-4.5	**0.05**	-17.1
11	**CO3**	IHWESASLLR	**0.65**	0.0	**0.66**	2.1	**0.62**	-5.0	**0.54**	-16.2
		TGLQEVEVK	**7.22**	0.0	**7.31**	1.3	**7.09**	-1.8	**5.85**	-19.0
		VHQYFNVELIQPGAVK	**7.63**	0.0	**7.64**	0.1	**7.46**	-2.2	**6.33**	-17.1
12	**CYTC**	ALDFAVGEYNK	**0.19**	0.0	**0.20**	3.4	**0.19**	-4.1	**0.16**	-17.6
13	**FA10**	NTEQEEGGEAVHEVEVVIK	**1.52**	0.0	**1.55**	2.1	**1.59**	4.7	**1.17**	-22.9
14	**FA5**	SQHLDNFSNQIGK	**0.12**	0.0	**0.11**	-7.0	**0.12**	-5.2	**0.11**	-14.3
15	**FA9**	NCELDVTCNIK	**0.53**	0.0	**0.54**	1.9	**0.53**	-0.3	**0.41**	-22.9
		SALVLQYLR	**0.22**	0.0	**0.23**	2.2	**0.22**	-2.1	**0.18**	-21.2
16	**HPT**	TEGDGVYTLNNEK	**21.32**	0.0	**21.70**	1.8	**20.80**	-2.4	**18.52**	-13.1
		VGYVSGWGR	**45.71**	0.0	**46.11**	0.9	**43.58**	-4.7	**37.36**	-18.3
		VVLHPNYSQVDIGLIK	**1.63**	0.0	**1.61**	-1.4	**1.59**	-2.2	**1.30**	-20.4
17	**IC1**	GVTSVSQIFHSPDLAIR	**1.89**	0.0	**1.80**	-4.7	**1.85**	-1.9	**1.75**	-7.5
		TNLESILSYPK	**1.30**	0.0	**1.31**	0.7	**1.31**	1.0	**1.20**	-7.4
18	**PLF4**	ICLDLQAPLYK	**1.34**	0.0	**1.40**	4.8	**1.33**	-0.3	**1.17**	-12.6

19	**PROC**	LGDDLLQCHPAVK	**0.15**	0.0	**0.16**	4.4	**0.15**	-0.1	**0.12**	-17.6
20	**PROS**	FSAEFDFR	**1.63**	0.0	**1.56**	-4.4	**1.44**	-12.1	**1.37**	-16.2
		IETISHEDLQR	**2.54**	0.0	**2.50**	-1.6	**2.36**	-7.2	**2.14**	-15.7
21	**SHBG**	IALGGLLFPASNLR	**0.09**	0.0	**0.09**	-2.7	**0.08**	-5.8	**0.09**	-3.9
		QAEISASAPTSLR	**0.22**	0.0	**0.22**	-0.8	**0.20**	-11.5	**0.18**	-21.0
22	**TFR1**	ILNIFGVIK	**0.02**	0.0	**0.02**	0.7	**0.02**	-0.5	**0.02**	-12.5
23	**THRB**	ELLESYIDGR	**1.80**	0.0	**1.82**	1.1	**1.73**	-3.7	**1.48**	-17.5
		SEGSSVNLSPPLEQCVPDR	**2.96**	0.0	**2.90**	-2.1	**2.91**	-1.7	**2.42**	-18.4
		TATSEYQTFFNPR	**1.50**	0.0	**1.51**	0.6	**1.47**	-1.8	**1.22**	-18.6
24	**TRFE**	DGAGDVAFVK	**25.81**	0.0	**26.60**	3.1	**25.35**	-1.8	**21.58**	-16.4
		EFQLFSSPHGK	**12.69**	0.0	**11.84**	-6.7	**12.03**	-5.2	**9.74**	-23.2
		YLGEEYVK	**32.31**	0.0	**32.86**	1.7	**31.19**	-3.5	**26.26**	-18.7
25	**TRFL**	YYGYTGAFR	**0.13**	0.0	**0.13**	4.0	**0.12**	-4.6	**0.11**	-15.3
26	**TTHY**	AADDTWEPFASGK	**0.49**	0.0	**0.51**	4.3	**0.47**	-4.5	**0.44**	-10.2

The quantification results are presented as peak area ratio (PAR) values. The bias is calculated against the results from Set 1, which was the standard method using equal numbers of tips and wells.

[a]PAR = Peak area of endogenous peptide / Peak area of SIS peptide

[b]$Bias_{Set\,X}$ (%)= $(PAR_{Set\,X} - PAR_{Set\,1}) / PAR_{Set1} \times 100$

1.3.4. Evaluation of feasibility for the introduction of the automated in-solution digestion workflow in clinical fields

In a series of analyses so far, I have confirmed that the introduction of automation yields high reproducible results. For general application in the clinical field, the validity of the automated process was evaluated in terms of time and cost. As shown in **Figure 1-12**, more time is needed for dispensing reagents with fewer tips for the same number of wells. However, it remains shorter compared with the time that is required for manual dispensing of reagents. Although it takes longest with one tip, it takes less than one-third of the time to dispense each reagent to 24 samples by manual preparation (～500 s.). Those methods have a tremendous advantage over manual preparation because automation can minimize the variation in reaction times for each sample and prevent protein degradation.

Meanwhile, for daily operations of the assay in clinical practice, the cost aspect should be considered to reduce the patient's financial burden. The total experimental cost per sample preparation for each set and the manual preparation are shown in **Table 1-6**. The cost per sample preparation was calculated for measurable consumables, such as tips, plates, sample vials, expensive reagents (RapiGest and trypsin), and so on. The total cost for the automated workflow was twice as high as that of the manual workflow. Thus, despite the comparable reproducibility, the total cost of the automation system should be optimized to increase utility. Therefore, I evaluated the potential for minimizing cost, by analyzing the relationship between reproducibility and experimental cost based on the number of tips **(Figure 1-13)**. As a result, the methods that used eight or four tips in preparing the 24 wells had high reproducibility (CVs <15%), similar to the method in which reagents were dispensed

using an individual tip for each well (Set 1 experiment). The variation increased as reagents were dispensed to more wells with one tip. Thus, greater use of consumable items raises the cost, but higher reproducibility is ensured. I found that a considerable amount of cost for the automated platform could be saved by methods with eight or four tips (29 and 37% reduction in total experimental cost, respectively). The bulk of the cost reduction comprised trypsin and RapiGest reagent. Therefore, reducing wasted volumes by using fewer tips could be a key factor in lowering costs (see **Table 1-6**). It is most effective to use one tip for five wells– the point at which the cost and variation (CV value) graphs intersect (**Figure 1-13**); this method minimizes the variability in sample preparation while keeping expenditures low.

Table 1-6. Approximate experimental cost per sample for in-solution digestion.

Consumables	Set1	Set2	Set3	Set4	Manual
	24 Tips	8 Tips	3 Tips	1 Tip	24 Tips
Tip	0.551	0.184	0.092	0.023	0.040
Reagent stock plate (or tube)	0.089	0.030	0.015	0.004	0.010
Sample plate (or tube)	0.075	0.075	0.075	0.075	0.079
RapiGest	1.867	1.244	1.089	0.972	0.933
Trypsin	6.277	4.185	3.662	3.269	3.139
Glass vial for LC	1.884	1.884	1.884	1.884	1.884
Total	**10.74**	**7.60**	**6.82**	**6.23**	**6.08**
Savings against Manual (%)	**-76.56**	**-24.93**	**-12.02**	**-2.34**	
Savings against Set 1 (%)	-	**29.24**	**36.55**	**42.03**	

The cost unit is US$. The price is based on 2019 index. Prices for other consumables and reagents are excluded when negligible against the total experimental cost. Labor and facility costs for the instrument are also excluded, because these costs vary depending on the region. LC, liquid chromatograph.

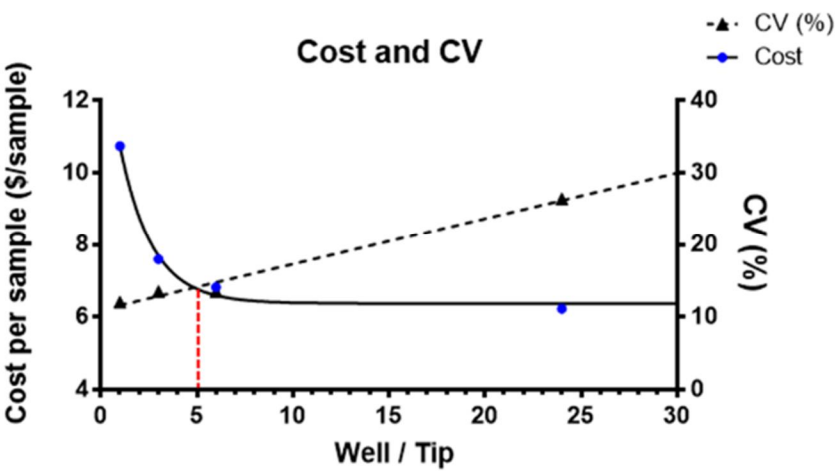


Figure 1-13. The relationship between variation and cost.

The well/tip on the x-axis was calculated for each dataset, dividing the number of wells by the number of tips used for dispensing the reagent. The total experimental

cost per sample in each set is represented as a dotted line (details in Table 1-6). The CVs were plotted based on the average interday CV value of 52 peptides. Since optimal balance between reasonable reproducibility and the cost is obtained at the intersection of the graph, using 1 tip for 5 wells.

1.4. DISCUSSION

Automated sample preparation is essential for the development of routine clinical MRM−MS assays. Furthermore, automated processing can reduce the systematic biases that are associated with transferring a multistep assay between laboratories and even between technicians in the same laboratory, thereby increasing reproducibility and facilitating the wider adoption of MRM−MS protein assays. This work focused on developing a robust sample processing workflow that facilitates the development of high-confidence clinical MRM−MS assays while reducing human labor. When combined with online diversion for the final cleanup step, the CV for the entire procedure was under 20%.

Although automated platforms are attractive concerning their high reproducibility, the tremendous cost of consumables is one of the obstacles to the routine application of the MRM−MS assay in clinical practice. There have been few endeavors to reduce the costs that arise during the digestion process, although many automated digestion processes have been optimized. Therefore, this research is a meaningful initial attempt to automate the digestion process with a lower cost burden. Based on the comparison of reproducibility, I have found a potential for the automated workflow with the cost-optimized method by using fewer tips, reagents, and plates for preparation, ensuring stable automated serum sample preparation. It is most effective to use one tip for five wells− this method minimizes the variability in sample preparation while keeping expenditures low. Although this automated workflow was tested and constructed with an Agilent Bravo apparatus, it can be

adopted by other similar liquid-handling systems, such as those for which tips can be placed on specific pipet channels, 96-well plates can be used, and the aspiration height can be adjusted. With slight adaptations, the platform is expected to perform well for various sample types, other than serum, to cope with a massive number of samples.

Despite their comparability evaluated here, there are several limitations to this study. First, the robustness of the automated preparation workflow was evaluated only in the analysis of the technical replication experiment using identical pooled samples. Therefore, the feasibility of this automated process should be verified in the analysis of the biological replication experiment using individual samples or samples obtained from various origins. Second, this study was focused on the feasibility of the automated serum sample preparation for MRM-MS assays, the reproducibility was evaluated for the targeted quantification results for 52 peptides. Also, the absence of cross-contamination was not evaluated. To overcome these limitations, further study is required to evaluate the robustness of the automated sample preparation method in biological replication experiments for additional target peptides as well as to determine whether cross-contamination has occurred between samples placed within the 96-well plate (**Figure 1-14**).

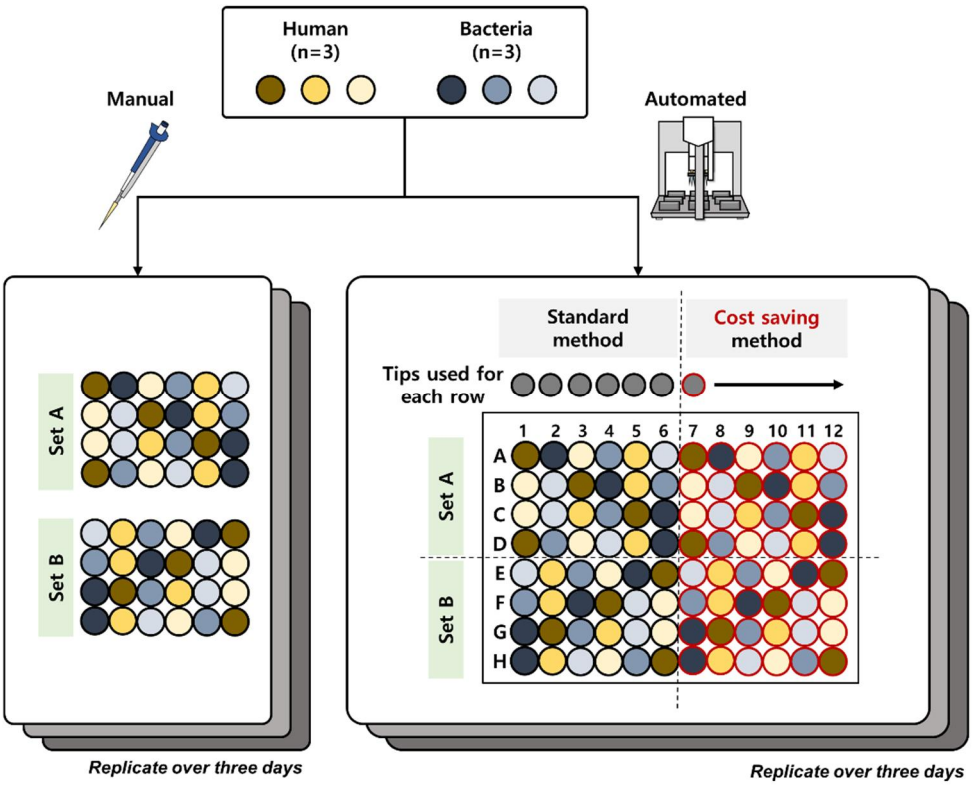


Figure 1-14. The overall scheme for further evaluation of the robustness of the automated sample preparation method in biological replication experiments. The robustness of the automated workflows (standard and cost-saving methods) can be evaluated in biological replication experiments using individual samples of humans and bacteria. Three individual samples per each sample origin (humans and bacteria) are prepared by the automated workflow and the manual process over three successive days. Three human-derived and three bacterial samples are placed alternately in each row of Set A in random order, while the sequence is inverted in Set B. The robustness of the automated workflow for biological replications can be assessed by comparing the reproducibility and accuracy obtained from each preparation method. For the cost-saving automated method, the cross-contamination can be determined by quantification results for bacterial proteome in human samples

prepared immediately next to the bacteria samples, and vice versa. Thus, the robustness of the automated workflow and the cross-contamination can be evaluated within a single experiment.

1.5. CONCLUSION

The reproducibility of the technical replication experiment identified in the identical pooled samples is meaningful in that it can be applied to the analytical method validation process for biomarkers. The analytical method validation processes are requiring labor-intensive steps to prepare an identical matrix digested from the pooled samples, as well as individual sample preparation for each validation criteria (**Figure 1-15)**. Thus, it can significantly reduce human labor and increase the throughput and reliability of the analytical validation study after further evaluation of the robustness of the automated process in biological replication experiments. I hope that this automated platform will help implement MRM−MS protein assays in clinical practice.

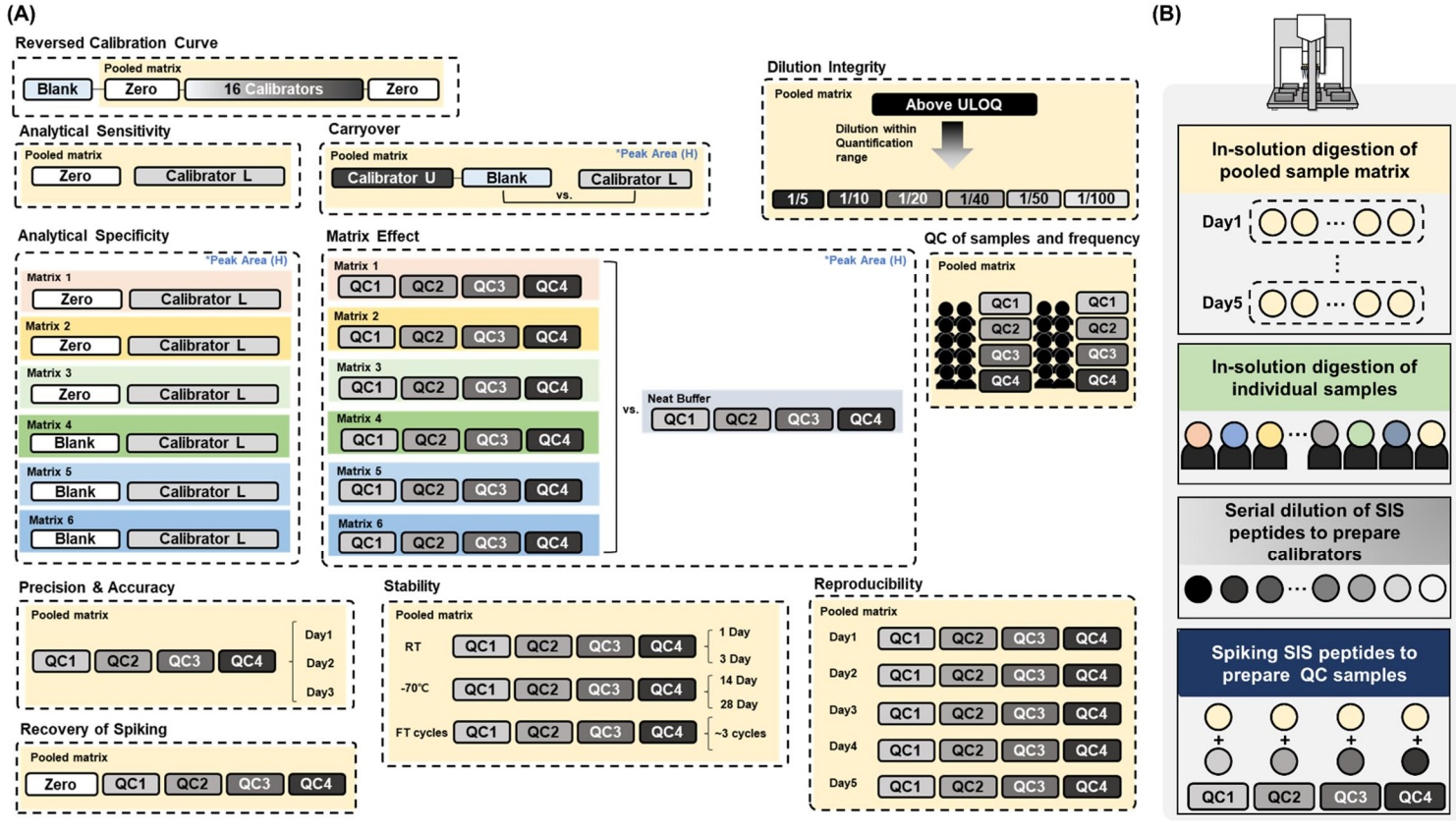


Figure 1-15. The potential of the robust automated sample preparation method for the overall analytical method validation process to reduce human labor.

(A) The analytical method validation processes are required labor-intensive steps to prepare an identical matrix digested from the pooled samples, individual sample preparation, or SIS peptides with different concentrations for each validation item. The detailed criteria of analytical method validation are summarized in the previous study (35). (B) The current automated workflow can perform in-solution digestion of pooled or individual samples over five successive days with high reproducibility, serial dilution of SIS peptides, or spiking SIS peptides to the matrix for making QC samples. It will be useful for reducing human labor and increasing the throughput and reliability of the analytical validation studies. QC1: LLOQ concentration, QC2: low concentration, QC3: medium concentration, QC4: high concentration, Calibrator U and L are representing ULOQ and LLOQ respectively.

QC, Quality Control, LLOQ: Lower Limit of Quantification, ULOQ: Upper Limit of Quantification.

CHAPTER 2

Development of the Inclusive Quantification MRM-MS Assay of Des-γ-Carboxyprothrombin (DCP) Proteoforms for Hepatocellular Carcinoma (HCC) Surveillance

2.1. INTRODUCTION

Liver cancer is the seventh most prevalent cancer worldwide and is the second leading cause of cancer-related deaths (57, 58). The most common type of primary liver cancer is hepatocellular carcinoma (HCC), which accounts for approximately 75% of all liver cancer cases (58-61). A primary risk factor for HCC is chronic liver cirrhosis due to chronic hepatitis B virus (HBV) or hepatitis C virus (HCV) infection (62, 63).

The prognosis for HCC remains poor, with a 5-year survival rate of less than 20% in most countries (58, 64-66). Consequently, diagnosing HCC at earlier stages is the treatment strategy due to it is related with better survival rates in early-stage HCC (70%) (67-69). Currently, ultrasonography (US) and serum alpha-fetoprotein (AFP) detection are widely used to surveil at-risk individuals for the development of HCC (70-72). However, these methods can often result in misdiagnosis due to the imprecise identification of small tumors in liver cirrhosis backgrounds using US or fluctuations in AFP levels that are caused by benign liver diseases (72-75). Further, certain HCCs with normal AFP levels can contribute to the low sensitivity of serum AFP. Therefore, ongoing research has attempted to develop more effective surveillance methods with enhanced sensitivity that can be used independently from or in conjunction with US or serum AFP (35, 76-78).

Another available marker for HCC surveillance is des-γ-carboxyprothrombin (DCP), also known as protein induced by vitamin K absence or antagonist-II, or abnormal prothrombin, which is found at elevated levels in patients with HCC (79,

80). Several studies have reported that DCP can be used to complement AFP for the early diagnosis of HCC (77, 81). Normal prothrombin is synthesized as a precursor containing 10 glutamic acid (Glu) residues in the N-terminal domain (Gla domain), at positions 6, 7, 14, 16, 19, 20, 25, 26, 29, and 32 (82). Under normal conditions, the precursors undergo posttranslational carboxylation of the Glu residues resulting in the conversion of Glu to γ-carboxylated glutamic acid (Gla) by vitamin K-dependent glutamyl gamma-carboxylase in the specific order of 26, 25, 16, 29, 20, 19, 14, 32, 7, and 6 (83, 84). Carboxylation is impaired under conditions of vitamin K deficiency, warfarin treatment, or liver dysfunction, (85, 86) resulting in DCP being released into the bloodstream as a mixture of proteoforms with up to 10 des-carboxylated Glu residues (87).

Concentrations of DCP have been determined using a conventional antibody-based assay featuring a monoclonal antibody produced by the MU3 cell line (88). The DCP epitope that is recognized by the MU3 antibody is located within the Gla domain at amino acids 17-27, which includes four Glu residues (19, 20, 25, and 26). Thus, the DCP proteoforms containing some Gla residues at the antibody epitope could have reduced affinity for the MU3 antibody compared with that of the totally non-carboxylated DCP. According to previous studies, the MU3 antibody binds predominantly with DCP molecules containing 9-10 Glu residues, weakly with those that possess 6-8 Glu residues, and rarely with those that have less than 5 Glu residues (88, 89). Recently, several studies aimed to develop discriminative quantification immunoassays for the detection of DCP proteoforms with lower Glu content to overcome this limitation and improve the diagnostic performance of DCP measurement (87, 90, 91). These studies used additional immunoassays that feature

other antibodies, such as 19B7, P-11, and P-16, which recognize different epitopes than those that are detected by the MU3 antibody. Although these studies reported the value of these assays for the detection of DCP proteoforms that contain fewer Glu resides, the performance of extra and separate immunoassays is costly and subject to batch effects.

Multiple-reaction monitoring–mass spectrometry (MRM-MS) is a powerful analytical method that can be used to accurately quantify peptides and proteins with high throughput. Recently, the MRM-MS assay has been shown to be advantageous compared with conventional antibody-based assays in terms of throughput and the ability to distinguish protein isoforms with common epitopes (20, 22, 32). In the previous study, they developed an MRM-MS assay to quantify DCP using a surrogate peptide; this found that the MRM-MS assay had comparable diagnostic power compared with the conventional immunoassay (36, 92). However, this MRM-MS assay remains limited because it only quantifies a surrogate peptide that represents just a small portion of the existing DCP proteoforms.

The objective of the present study was to improve the diagnostic power of the MRM-MS assay for DCP by inclusively quantifying a wider range of proteoforms with various des-carboxylation states. In brief, I examined potentially non-carboxylated peptides (referred to as Glu-peptides) within the Gla domain and developed a robust MRM-MS assay to quantify multiple Glu-peptides for the inclusive quantification of DCP proteoforms.

2.2. MATERIALS AND METHODS

2.2.1. Chemicals and reagents

High-performance liquid chromatography (HPLC)-grade solutions, including water, acetonitrile, formic acid, 0.1% formic acid in water, and 0.1% formic acid in acetonitrile, were purchased from Fisher Scientific (Loughborough, United Kingdom). Ammonium bicarbonate (200 mM) solution was purchased from iNtRON Biotechnology (Sungnam, Korea). Dithiothreitol and iodoacetamide were obtained from Merck Co. (Darmstadt, Germany) and Sigma-Aldrich (St. Louis, MO). RapiGest surfactant was obtained from Waters Corp. (Milford, MA,). Sequencing-grade chymotrypsin and trypsin were obtained from Promega (Madison, WI). Formic acid was purchased from Fisher Scientific. Stable isotope-labeled standard (SIS) peptides (heavy peptides) were synthesized by SynPeptide Co., Ltd. (Shanghai, China) (with >99% isotope purity and about >60% purity for individual peptides). Heavy peptides for two tryptic peptides were labeled (^{13}C and ^{15}N) at a C-terminal arginine (Arg) and lysine (Lys). The SIS peptides for three chymotryptic peptides were labeled (^{13}C and ^{15}N) at C-terminal tyrosine (Tyr), phenylalanine (Phe), and leucine (Leu).

2.2.2. Clinical specimens and study design

A total of 618 serum samples were obtained from patients with HCC and at-risk control patients with chronic hepatitis B (CHB), chronic hepatitis C (CHC), or liver cirrhosis (LC). All patients were recruited from two medical centers in Korea (Asan

Medical Center and Samsung Medical Center) and provided informed consent before enrollment. This study was approved by the institutional review boards of the Asan (IRB No. 2017-1049) and Samsung (IRB No. 2017-08-164) Medical Centers. The cohort of 300 patients that was used for model construction and assessment was recruited from Asan Medical Center (cohort A) and included 100 cases of HCC and 200 at-risk controls (**Figure 2-1**). Seventy percent of the patients were randomly defined as the training set (n = 210), and the remaining patients were defined as the test set (n = 90). The training set was used to construct a diagnostic model, while the test set was used to assess model performance. To validate the model performance in an external cohort, an independent validation set was recruited comprising 318 samples collected from Samsung Medical Center (cohort B), including 184 cases of HCC and 134 at-risk controls.

The HCC diagnosis was confirmed by the results of a histological examination or typical imaging features obtained by US, computed tomography, or magnetic resonance imaging, according to regular clinical practice guidelines (93). The stages of HCC were defined according to the Barcelona Clinic Liver Cancer (BCLC) classification as follows: very-early-stage (BCLC stage 0, single nodule <2 cm) and early-stage (BCLC stage A, a single 2-5 cm lesion or two to three lesions that are each <3 cm). Cirrhosis was defined clinically or radiologically using the following criteria: coarse liver echotexture and nodular liver surface on US, clinical features of portal hypertension (e.g., ascites, splenomegaly, or varices), or thrombocytopenia (platelet count <150 × 1,000/mm^3). CHB was defined as the presence of serum hepatitis B surface antigen for more than 6 months. Patients with persistent anti-HCV and HCV RNA for more than 6 months were defined as CHC. Ages and sex

distributions were matched between the control and case groups to the greatest extent possible; the clinical information for each data set and reference values (94) are detailed in **Table 2-1**.

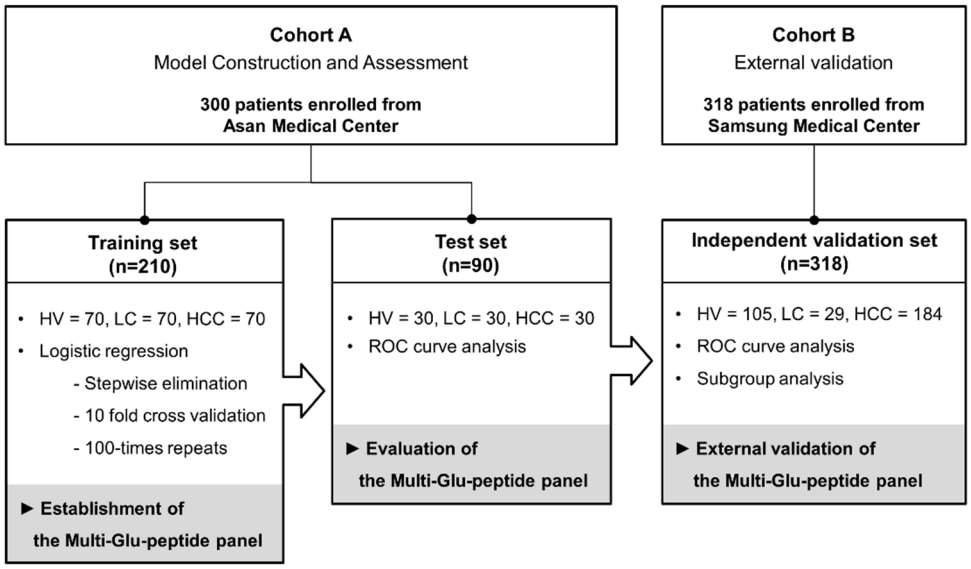


Figure 2-1. Design of the study.

The 618 patients were enrolled in this study from two cohorts (Asan and Samsung Medical Center). The 300 patients from cohort A were randomly divided into the training and test sets to construct the diagnostic model and assess its performance. The diagnostic model was established by stepwise logistic regression in the training set, consisting of serum samples from 70 HCC patients and 140 high-risk controls (70 with HV and 70 with LC). The diagnostic performance of the established model was assessed using the test set, comprised of 90 serum samples from 30 HCC patients and 60 high-risk controls (30 with HV and 30 with LC), by ROC curve analysis. Then, the model performance was validated in an independent validation set from cohort B, including 318 patients, consisting of 184 HCC patients and 134 high-risk

patients (105 with HV and 29 with LC). Additional evaluations for HCC surveillance performance of the model were conducted using subgroups of independent validation sets, such as patients with AFP- and DCP-negative, or early-stage and very-early-stage HCC. Abbreviations: HV, patients with hepatitis virus infections; LC, patients with liver cirrhosis; HCC, hepatocellular carcinoma.

Table 2-1. Clinical characteristics of the subjects in each cohort.

	Cohort A						Cohort B		
	Train set (n = 210)			Test set (n = 90)			Independent validation set (n = 318)		
Characteristics	**Controls**	**Cases**	***P* value**	**Controls**	**Cases**	***P* value**	**Controls**	**Cases**	***P* value**
n	140	70		60	30		134	184	
Age (years) [range]	53.7 ± 10.0 [28.0-75.0]	56.9 ± 9.3 [24.0-78.0]	*0.0262*	51.2 ± 8.1 [28.0-65.0]	59.7 ± 8.8 [46.0-74.0]	*<0.001*	50.8 ± 1.7 [24.0-80.0]	57.9 ± 9.4 [35.0-79.0]	*<0.001*
Gender, n (%)			*0.5926*			*1.0000*			*0.2115*
Male	108 (77.1%)	57 (81.4%)		47 (78.3%)	23 (76.7%)		99 (73.9%)	148 (80.4%)	
Female	32 (22.9%)	13 (19.6%)		13 (21.7%)	7 (23.3%)		35 (26.1%)	36 (19.6%)	
AFP (ng/mL), median [IQR]	2.6 [1.9-4.1]	6.7 [2.8-32.3]	*<0.001*	2.35 [1.9-3.3]	15.85 [4.0-909.9]	*<0.001*	2.8 [1.8-3.9]	8.3 [3.4-6.1]	*<0.001*
DCP (mAU/mL), median [IQR]	15.0 [13.0-18.5]	40.0 [24.3-121.3]	*<0.001*	16.0 [14.0-18.5]	36.0 [23.8-131.9]	*<0.001*	19.5 [15.0-23.8]	71.5 [27.8-268.3]	*<0.001*
AST (U/L)	34.5	30.5	*0.2189*	32.9	30.7	*0.4686*	44.0	156.6	*<0.001*
ALT (U/L)	33.0	30.1	*0.7168*	36.3	27.6	*0.5403*	46.8	143.8	*<0.001*
Prothrombin time (INR)	1.1 ± 0.1	1.0 ± 0.2	*0.5661*	1.1 ± 0.2	1.0 ± 0.1	*0.7840*	1.3 ± 1.9	1.2 ± 0.1	*<0.001*
Albumin (g/dL)	4.2 ± 0.4	3.9 ± 0.3	*<0.001*	4.3 ± 0.4	3.9 ± 0.4	*<0.001*	4.5 ± 0.4	3.7 ± 0.5	*<0.001*
Bilirubin (mg/dL)	1.2 ± 0.6	0.7 ± 0.4	*<0.001*	1.2 ± 0.9	0.6 ± 0.4	*<0.001*	0.9 ± 0.7	1.3 ± 0.6	*<0.001*
Creatinine (mg/dL)	0.8 ± 0.2	0.8 ± 0.2	*0.3185*	0.9 ± 0.2	0.9 ± 0.2	*0.1606*	1.0 ± 1.1	0.9 ± 0.2	*0.3119*
Etiology, n (%)			*0.8571*			*<0.001*			*<0.001*
HBV	96 (68.6%)	52 (74.3%)		51 (85.0%)	20 (66.7%)		117 (87.3%)	140 (76.1%)	

HCV	10 (7.1%)	4 (5.7%)		1 (1.7%)	2 (6.7%)		15 (11.2%)	8 (4.3%)	
HBV and HCV	3 (2.1%)	1 (1.4%)		2 (3.3%)	0		0	1 (0.5%)	
Others	31 (22.1%)	13 (18.6%)		6 (10.0%)	8 (26.7%)		2 (1.5%)	35 (19.0%)	
Child-Pugh score, n (%)			*0.0290*			*<0.001*			*<0.001*
5	121 (86.4%)	58 (82.9%)		54 (90.0%)	22 (73.3%)		127 (94.8%)	100 (54.3%)	
6	10 (7.1%)	12 (17.1%)		2 (3.3%)	8 (26.7%)		6 (4.5%)	60 (32.6%)	
7	5 (3.6%)	0		3 (5.0%)	0		0	22 (12.0%)	
8	4 (2.9%)	0		1 (1.7%)	0		0	1 (0.5%)	
9	0	0		0	0		1 (0.7%)	1 (0.5%)	
Child-Pugh class, n (%)			*0.0476*			*0.3659*			*<0.001*
0	1 (0.7%)	2 (2.9%)		0	0		0	0	
A	130 (92.9%)	68 (97.1%)		56 (93.3%)	30 (100%)		133 (99.3%)	160 (87.0%)	
B	9 (6.4%)	0		4 (6.7%)	0		1 (0.7%)	24 (13.0%)	
Tumor size (cm)	-	2.6 ± 1.3		-	2.8 ± 1.7		-	3.4 ± 2.4	
Tumor number, n (%)									
1	-	67 (95.7%)		-	28 (93.3%)		-	172 (93.5%)	
2	-	2 (2.9%)		-	2 (6.7%)		-	12 (6.5%)	
≥ 3	-	1 (1.4%)		-	0		-	0	
BCLC stage, n (%)									
0	-	27 (38.6%)		-	11 (36.7%)		-	36 (19.6%)	
A	-	40 (57.1%)		-	17 (56.7%)		-	123 (66.8%)	
B	-	3 (4.3%)		-	2 (6.6%)		-	25 (13.6%)	

Data are presented as mean ±SD (AFP and DCP: median [IQR]). P-values for gender, etiology, Child-Pugh score, Child-Pugh class were calculated by chi-squared test. P-values for others were calculated by t-test. P-values less than 0.001 are shown as '<0.001'. The reference ranges for each liver function test are based on laboratory reference values from Massachusetts General Hospital (MGH) (43): AST (0-35 U/L), ALT (0-35 U/L), albumin (3.5-5.5 g/dL), bilirubin (0.3-1.0 mg/dL), creatinine (<1.5 mg/dL), and prothrombin time (11.1-13.1 sec). The reference range for prothrombin time as INR is <1.1, per the Mayo Clinic (www.mayoclinic.org).

Abbreviations: ALT, alanine aminotransferase; AST, aspartate transaminase; HBV, hepatitis B virus; HCV, hepatitis C virus; INR, international normalized ratio; IQR, interquartile ratio

2.2.3. Candidate non-carboxylated peptides within the Gla domain of DCP

The Gla domain of prothrombin consists of 46 amino acids at the N-terminus, including 10 sporadic Gla sequences. I identified potential non-carboxylated peptides (Glu-peptides) that originated from the Gla domain through an *in silico* proteolytic digestion using Skyline (McCoss Laboratory, University of Washington). Due to the decarboxylation (neutral loss of CO_2) that occurred in the fragmentation process, I have considered only the non-carboxylated peptides to obtain reproducible quantification results (95). The *in silico* digestion was performed separately with trypsin and chymotrypsin, and peptides with six to 30 amino acids and without methionine residues were selected to ensure reproducible quantification. Two tryptic peptides (ANTFLEEVR, ECVEETCSYEEAFEALESSTATDVFWAK) and three chymotryptic peptides (EEVRKGNL, ERECVEETCSY, ESSTATDVF) remained as potential targets representing the partially non-carboxylated state of the Gla domain (**Figure 2-2**). MRM-MS analysis was then performed to verify the detectability of these five peptides by MS analysis, using corresponding SIS peptides (also referred to as heavy peptides). Among the five Glu-peptides, the longest tryptic peptide (ECVEETCSYEEAFEALESSTATDVFWAK) did not have any discernable signal due to its long length and hydrophobicity. The other four Glu-peptides were detected by MRM-MS and were chosen for further analysis.

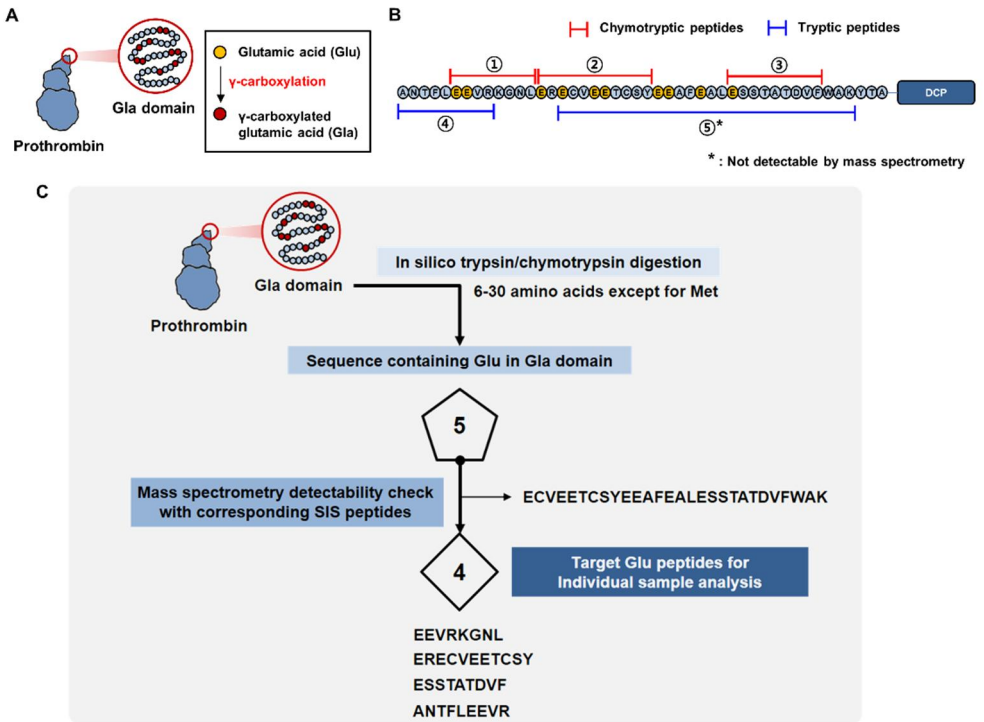


Figure 2-2. Candidate non-carboxylated peptides (Glu-peptides) within the Gla domain identified by in silico digestion with chymotrypsin and trypsin.

(A) The Gla domain is the N-terminal region of prothrombin, consisting of 46 amino acids and containing 10 sporadic Gla residues. The Gla residues at the Gla domain are generated by the γ-carboxylation of the Glu residues in the prothrombin precursor. (B) The *in silico* digestion, based on the Gla domain sequences following chymotrypsin and trypsin digestion, derived three- and two-Glu-peptide sequences, respectively. The asterisked (*) tryptic peptide was non-detectable by mass spectrometry due to the long length. (C) The overall scheme to identify the candidate non-carboxylated peptides in Gla domain.

2.2.4. Multienzyme digestion for sample preparation

All serum samples were randomized within each set before preparation. The complete sample preparation workflow is shown in **Figure 2-3** and detailed in **Table 2-2**. The six most abundant proteins (albumin, immunoglobulin [Ig]G, antitrypsin, IgA, transferrin, and haptoglobin) were depleted using a multiple affinity-removal system column (Hu-6, 4.6 × 100 mm; Agilent Technologies, CA) and their exclusive buffers (buffers A and B). The depleted serum was concentrated using a 3-kDa molecular weight cutoff concentrator (Amicon Ultra-4 3K; Millipore, MA). The proteins in depleted and concentrated serum samples were quantified by the bicinchoninic acid assay, and 200 μg of proteins were denatured, alkylated, and divided into two equal fractions. Each sample pair was separately digested with trypsin and chymotrypsin to obtain peptides without competing for cleavage sites in a single run while minimizing variations due to prior steps. The incubation was performed at 37°C for 4 hours and was stopped by the addition of formic acid. The supernatant was transferred to clean tubes after centrifugation at 16,602g at 4°C for 1 hour to remove the by-products of RapiGest-SF. The individual enzymatic digests were mixed evenly and spiked with corresponding heavy peptides before the MRM-MS analysis.

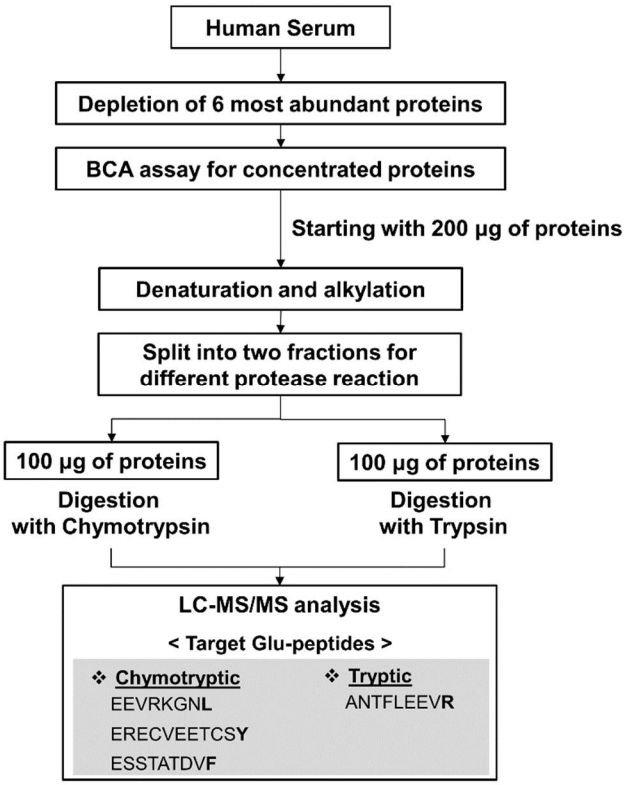


Figure 2-3. Scheme of the adapted sample preparation workflow using semi-separate, multi-enzyme digestions for the simultaneous procurement and analysis of tryptic and chymotryptic peptides in a single batch.

Human serum samples were depleted of the 6 most abundant proteins and concentrated. The concentration of the concentrated sample was measured by BCA assay, and 200 µg of protein was reduced with DTT and alkylated with IAA. The denatured proteins were separated into two fractions and digested separately with trypsin and chymotrypsin. After the digestion reactions were completed, the two fractions were combined into one final vial and spiked with heavy isotope-labeled peptides corresponding to the target peptides and subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS). Abbreviations: BCA, bicinchoninic acid; DTT, dithiothreitol; IAA, iodoacetamide.

Table 2-2. Total workflow of the multi-enzyme, in-solution digestion protocol.

No.	Steps	Stock Concentration	Volume (µL)	Cumulative Volume (µL)		Temp.	Time
1	Adjust sample volume	200 µg of depleted and concentrated serum sample		40		25°C	
2	Denaturation and reduction	0.5% RapiGest, 50 mM DTT, 200 mM ABC, pH 8.0	10	50		60°C	1 hour
3	Alkylation	120 mM IAA, 100 mM ABC	10	60		25°C	30 min (dark reaction)
	Separate into two equal fractions			30 Trypsin	30 Chymotrypsin		
4	Enzymatic digestion	0.1333 µg/µL [1:50 (w/w)], 50 mM ABC	15	45	45	37°C	4 hours

5	Quenching	10% Formic acid	5	50	50	37°C	30 min
6	Centrifuge and transfer the supernatant into a new, empty tube			50	50	4°C	1 hour (at 15,000 rpm)
7	Combine the final digests from both fractions			100			

Abbreviations: DTT, dithiothreitol; ABC, ammonium bicarbonate; IAA, iodoacetamide

2.2.5. Selection of a quantifier ion for each target peptide

I experimentally screened the six intense transitions as an initial list of MRM-MS transitions using the heavy peptides. The best transition was selected as a quantifier ion, considering the results of the reversed response curve analysis and the Automated Detection of Inaccurate and imprecise Transitions (AuDIT) algorithm (53), according to the following criteria: (1) the best linearity of the response curve (based on the correlation coefficient, R^2); (2) the lower limit of quantification (LLOQ) value was lowest among the transitions; and (3) interference-free status from AuDIT results.

The background matrix for the response curves was prepared using 100 μg of proteins from depleted pooled hepatitis serum for each enzyme fraction. The calibration points were generated by mixing the background matrix with variable amounts of heavy peptides from 78.13 fmol to 20 pmol, over a 100-fold range. All calibration points were sequentially analyzed, followed by a blank sample (0.1% formic acid in HPLC water), from zero sample (matrix only) to the highest concentration point in triplicate. The peak area ratio (PAR) was calculated with the peak area of heavy peptides normalized against that of corresponding endogenous (light) peptides existing in a matrix. Linear regression analysis was conducted on the plot in which the PAR of heavy peptides to light peptides was plotted against the nominal concentration of heavy peptides on a $\log_{10}$ scale.

The limit of detection (LOD) and limit of quantification (LOQ) were calculated based on the averaged PAR, plus 3 times and 10 times the SD for a zero sample that was analyzed in triplicate, respectively. The LLOQ was determined as the lowest concentration at which the precision was under 20%, the accuracy was within 20%,

and the signal-to-noise (S/N) ratio over 5, representing the first point of the response curve. Similarly, the upper LOQ (ULOQ) was defined as the highest concentration on the response curve showing the precision under 20% and the accuracy was within 20%, representing the last point of the response curve. The analytical information and AuDIT results for the quantifier ion used for each peptide and their response curves are shown in **Tables 2-3 and 2-4**, respectively.

Table 2-3. MRM-MS parameters for the quantification of four Glu-peptides from the DCP Gla domain.

Protein Name	Peptide Sequence	Fragment Ion Type	RT (min)	Isotope[a]	Precursor ion		Product ion		AuDIT results[b]			
					Q1 (m/z)	Ion charge state	Q3 (m/z)	Ion charge state	Interference[c]		CV (%)[d]	Overall state[e]
									State	*P*-value		
Prothrombin	ANTFLEEVR	y6	32.44	Light	539.780	2	792.425	1	Good	*0.0001*	10.05%	**Good**
				Heavy	544.784	2	802.433	1				
	ERECVEETCSY	y2	24.19	Light	731.285	2	269.113	1	Good	*0.8110*	8.95%	**Good**
				Heavy	736.298	2	279.140	1				
	ESSTATDVF	b7	29.88	Light	478.714	2	692.273	1	Good	*0.0117*	11.76%	**Good**
				Heavy	483.728	2	692.273	1				
	EEVRKGNL	b7	14.77	Light	472.762	2	407.214	2	Good	*0.2163*	9.89%	**Good**
				Heavy	476.270	2	407.214	2				

Abbreviations: RT, retention time; CV, coefficient of variation; m/z, mass-to-charge ratio. AuDIT, automated detection of inaccurate and imprecise transitions.

[a] Light: unlabeled endogenous peptides; Heavy: the stable isotope (^{13}C and ^{15}N)-labeled synthetic peptides at C-terminal arginine (Arg, R), tyrosine (Tyr, Y), phenylalanine (Phe, F), and leucine (Leu, L).

[b] The automated detection of inaccurate and imprecise transitions (AuDIT) algorithm was used to determine whether the ion used to identify each peptide target was affected by matrix interference.

[c] The quantifier ions with the *P*-values larger than the threshold of 10^{-5} are marked as 'Good' in the AuDIT result.

[d] The quantifier ions with CV values less than the threshold of 20% are marked as 'Good' in the AuDIT result.

[e] All quantifier ions were marked as 'good' for overall interference states, satisfying the criteria for both the CV and the *P*-value.

Table 2-4. Results of the reversed response curve analysis for four Glu-peptides.

No.	Peptide Sequence	Measurements	Zero sample	Calibrator 1	Calibrator 2	Calibrator 3	Calibrator 4	Calibrator 5	Calibrator 6	Calibrator 7	Calibrator 8
1	ANTFLEEVR	Expected conc. (fmol)	N/A	78.125	156.250	312.500	625.000	1250.000	2500.000	5000.000	10000.000
		PAR (Mean)[a]	0.0334	0.2844	0.6098	1.0611	2.2631	5.0173	10.0182	21.3615	40.5140
		SD	0.0123	0.0199	0.0997	0.1291	0.2177	0.7421	0.4183	3.2039	6.7917
		CV (%)	36.66	6.98	16.36	12.17	9.62	14.79	4.18	15.00	16.76
		S/N[b]	N/A	8.21	16.50	29.57	59.18	142.29	285.89	582.61	1012.08
2	ERECVEETCSY	Expected conc. (fmol)	N/A		312.500	625.000	1250.000	2500.000	5000.000	10000.000	20000.000
		PAR (Mean)[a]	0.7351		4.1101	9.4825	20.3575	30.5907	66.9806	108.9276	214.7077
		SD	0.2496		0.7603	1.4155	0.2743	6.0207	8.7954	12.6391	18.8996
		CV (%)	33.96		18.50	14.93	1.35	19.68	13.13	11.60	8.80
		S/N[b]	N/A		6.57	15.39	25.50	59.21	116.36	237.26	416.11
3	ESSTATDVF	Expected conc. (fmol)	N/A			625.000	1250.000	2500.000	5000.000	10000.000	20000.000

		PAR (Mean)[a]	0.0846			1.2016	1.9072	3.6504	5.5057	14.7844	24.3178
		SD	0.0053			0.1164	0.1755	0.4095	0.2050	1.9891	2.0658
		CV (%)	6.30			9.69	9.20	11.22	3.72	13.45	8.50
		S/N[b]	N/A			16.69	26.18	49.54	89.47	169.61	286.89
4	EEVRKGNL	Expected conc. (fmol)	N/A	78.125	156.250	312.500	625.000	1250.000	2500.000	5000.000	10000.000
		PAR (Mean)[a]	0.1373	1.3187	2.8958	4.4914	9.2414	19.8764	37.4456	68.8534	189.6540
		SD	0.0175	0.1985	0.5230	0.4541	1.4980	1.4039	1.7719	3.5002	7.1681
		CV (%)	12.77	15.05	18.06	10.11	16.21	7.06	4.73	5.08	3.78
		S/N[b]	N/A	7.23	13.81	26.33	49.92	102.78	216.24	471.10	812.31

Abbreviations: PAR, peak area ratio; SD, standard deviation; CV, coefficient of variation; S/N, Signal-to-noise; N/A, not applicable.

[a] Calculated by dividing the peak area of stable isotope-labeled (SIS, heavy) peptide by that of endogenous (light) peptide.

[b] Calculated by dividing the heavy-peptide peak area of each calibrator by that of the zero sample.

The highest concentration (Calibrator 8) of all reversed response curves represent the upper limit of quantification (ULOQ). The lower limit of quantification (LLOQ) is the lowest concentration of each response curve that meets the criteria (Calibrator 1 for ANTFLEEVR and EEVRKGNL,

Calibrator 2 for ERECVEETCSY, and Calibrator 3 for ESSTATDVF). The summary of the response curve equations is in **Table 2-5**, the LLOQ criteria and results are presented in **Table 2-6**.

2.2.6. Quantitative MRM-MS Analysis

The quantification of target peptides for DCP was performed on an Agilent 6490 triple quadrupole MS (Agilent Technologies) with a Jetstream electrospray source coupled with a 1260 Infinity HPLC system (Agilent Technologies). The liquid chromatography–MS system was controlled by MassHunter (vB06.01; Agilent Technologies) software for the establishment of a scheduled MRM-MS method and data acquisition.

The total liquid chromatography assay was performed over 70 minutes, with a binary gradient consisting of mobile phase A (water 0.1% volume [vol]/vol formic acid) and mobile phase B (acetonitrile 0.1% vol/vol formic acid). Twenty microliters of chymotryptic and tryptic peptides was injected into the guard column (2.1 mm × 30 mm internal diameter [id], 1.8 μm particle size; Agilent Zorbax SB-C18), which was maintained at 40°C. After online desalting for 10 minutes at 5 μL/minute with 3% B, the peptides were subjected to a reversed-phase analytical column (150 mm × 0.5 mm id, 3.5 μm particle size; Agilent Zorbax SB-C18) maintained at 40°C. The separation of the peptides was conducted with a binary gradient of 3% to 35% B through the column for 45 minutes at 40 μL/minute. Equilibration of the column for the next run was performed by raising the gradient to 70% B for 5 minutes and then lowering it to 3% B for 10 minutes.

The ion spray capillary voltage was 2,500 V, and the nozzle voltage was 2,000 V. The drying gas and sheath gas were set to flow at 15 L/minute at 250°C and 12 L/minute at 250°C, respectively. The voltage of the cell accelerator was adjusted to 5 V. The fragment voltage and the delta electron multiplier voltage were set to 380 V and 200 V, respectively. The resolution mode of the first and third quadrupoles

was set to unit mode.

2.2.7. Data analysis

Quantitative analysis after MS analysis was performed using Skyline (McCoss Laboratory), which handled the MRM-MS raw data files from import to alignment and was used to conduct peak area calculations for transitions. The raw data were processed in Skyline, and each data point was smoothed by the Savitzky-Golay method. The PAR of the endogenous peptide to the heavy peptide for each peptide was used to represent the relative abundance of the peptide in each sample.

In the training set, a DCP multi Glu-peptides panel was constructed to discriminate cases from controls by stepwise backward logistic regression with 10-fold cross-validation (repeated 100 times). The stepwise backward elimination strategy was used to maximize the opportunity to identify the best combination of Glu-peptides for the discriminative quantification of DCP proteoforms between cases and control groups. The 10-fold cross-validation approach was used to avoid the overfitting of the model. The logistic regression model was trained and tested using the Classification and Regression Training (CARET) package in R statistical software program (version 3.6.3; R Foundation, Vienna, Austria) (96). The method used for stepwise backward elimination was the glmStepAIC method, which performed automatic stepwise variable selection based on the Akaike information criterion to train the model. Then, the optimal DCP multi Glu-peptides model was selected based on its accuracy. The 'Combined model' was developed by combining the multi-Glu-peptide panel with the serum AFP levels by the logistic regression method in the 'caretStack' function of the 'caretEnsemble' package in the R as a

stacked ensemble model. The overall scheme of model construction and evaluation is shown in **Figure 2-4.**

A receiver operating characteristic (ROC) curve was used to generate area under the ROC curve (AUROC) values to evaluate the predictive ability of the DCP multi-Glu-peptide panel in each data set. The cut-off point was identified by calculating the Youden Index (J = max [sensitivity + specificity –1]) for the training set. The relative differences in abundance for each peptide in the panel were compared between the control and case groups using the Mann-Whitney U test. DeLong`s tests were conducted to compare the AUROC values. All reported *P* values are from two-sided tests, and two-tailed $P < 0.05$ was considered significant.

All statistical analyses were performed using R (version 3.6.3; R Foundation, Vienna, Austria), IBM SPSS (version 25.0; IBM, Chicago, IL), and GraphPad Prism (version 6.0; GraphPad, San Diego, CA).

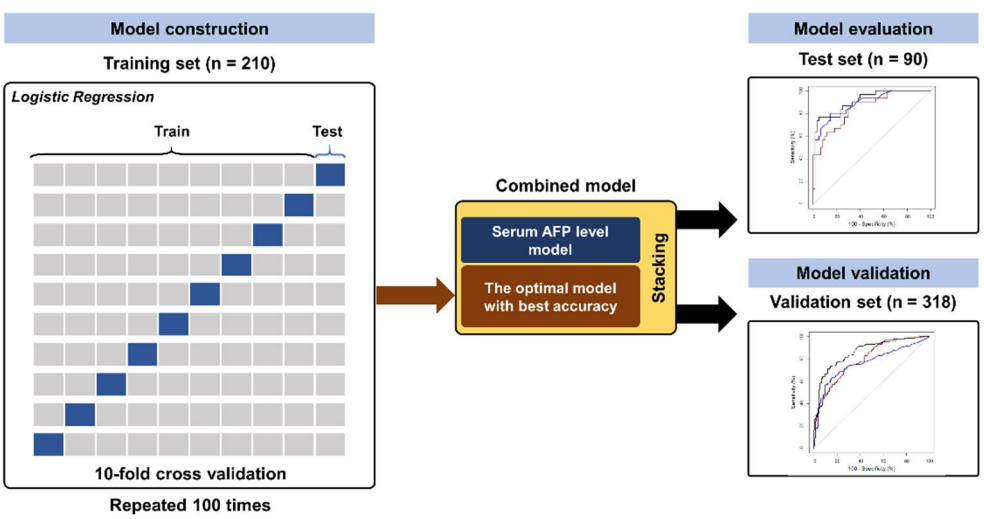


Figure 2-4. The overall scheme of the optimal DCP model and combined model construction and evaluation. In the training set, a DCP multi Glu-peptides panel was constructed by stepwise backward logistic regression with 10-fold cross-

validation repeated 100 times. Then, the optimal model was selected based on its accuracy and the combined model was developed by stacking the DCP multi-Glu-peptide panel with the serum AFP levels by the logistic regression method of the caretStack function in the caretEnsemble package.

2.3. RESULTS

2.3.1. Reversed response curves for four candidate Glu-peptides in depleted human serum

The reversed response curves for four DCP Glu-peptides are shown in **Figure 2-5**. Each curve satisfied the U.S. Food and Drug Administration guidelines for validating response curves (35); more than six calibration points were composed in each curve, and the coefficients of variance (CVs) of the measurements (n = 3) at all points in the curve were below 20% (**Table 2-4**). All correlation coefficients (R^2) of the response curve were above 0.99, except that for the ESSTATDVF peptide, which had an R^2 value that was slightly lower than the others (0.9872). The LOD, LOQ, LLOQ, and ULOQ values for the quantifier ion in each of the four Glu-peptides are summarized in **Figure 2-5**, and the results of the linear regression analyses for the response curve for each peptide are summarized in **Table 2-5**.

The analytical sensitivities of the target peptides at the LLOQ concentration met the requirements for precision, accuracy, and S/N criteria described in the Materials and Methods section (**Table 2-6**). The potential interferences of the analytes in the biological samples were inspected as the analytical specificity of individual serum samples from 6 patients with hepatitis. The interference values of peptides satisfied the standard criteria in all samples (interference <20%), as shown in **Table 2-7**. The average interference values of six matrices for the ANTFLEEVR, ERECVEETCSY, ESSTATDVF, and EEVRKGNL peptides were 6.4%, 3.2%, 7.5%, and 6.8%, respectively. The carryovers were inspected to ensure that the

ULOQ samples would not affect the subsequent sequential analysis of specimens. The average carryover of the four analytes ranged from 3.28% to 12.10%, which met the criteria (carryover <20%; **Table 2-8**).

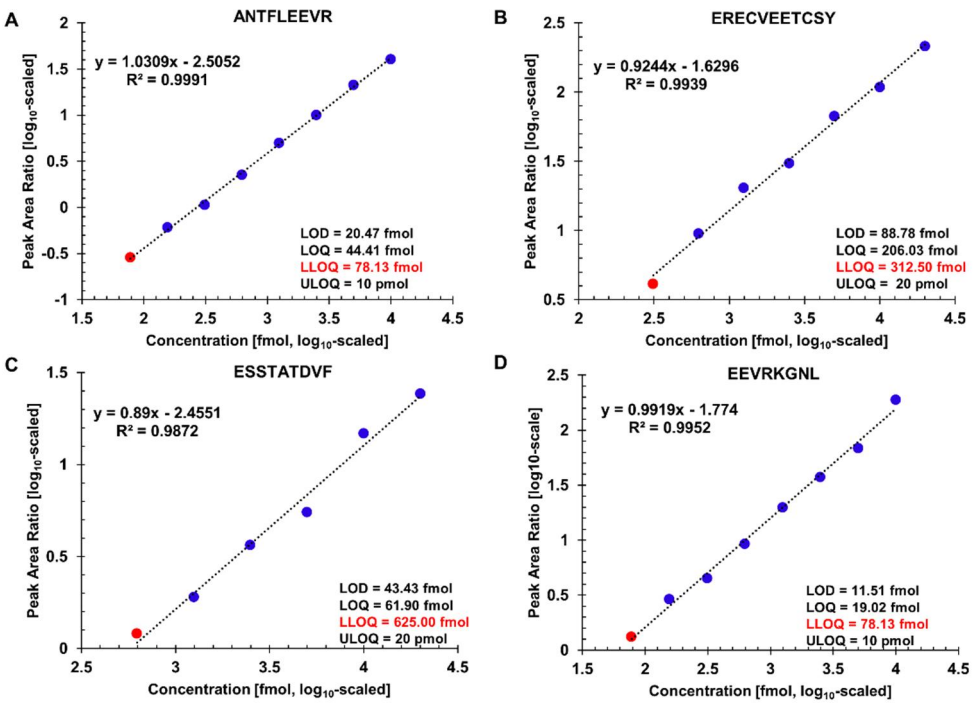


Figure 2-5. Reversed response curves for the quantifier ions of ANTFLEEVR (A), ERECVEETCSY (B), ESSTATDVF (C), and EEVRKGNL (D).

The response curves were generated by linear regression analysis with $\log_{10}$-scaled values, and their equations are summarized in Table 2-5. Each curve consists of over 6 calibrators (blue points and red points), and the lowest calibrator (red point) represents the lower limit of quantification (LLOQ) concentration.

Table 2-5. Summary of reversed response curves for each Glu-peptide.

No.	Peptide Sequence	Slope	Intercept	R^2
1	ANTFLEEVR	1.0309	−2.5052	0.9991
2	ERECVEETCSY	0.9244	−1.6296	0.9939
3	ESSTATDVF	0.8900	−2.4551	0.9872
4	EEVRKGNL	0.9919	−1.774	0.9952

Abbreviations: R^2, coefficient of determination

Each reversed response curve equation was generated with log10-scaled values.

Table 2-6. Analytical sensitivity from analysis results of zero sample and LLOQ sample by MRM-MS assay.

No.	Peptide Sequence	Measurements	Zero sample	LOD[b]	LOQ[c]	LLOQ[d]				
						Expected	Measured[e]	S/N[f]	Precision (%)[g]	Accuracy (%)[h]
1	ANTFLEEVR	PAR (mean)[a]	0.0334	0.0702	0.1560	N/A	0.2844	8.21	6.98	101.79
		Concentration. (fmol)	9.97	20.47	44.4081	78.13	79.52			
2	ERECVEETCSY	PAR (mean)[a]	0.7351	1.4840	3.2314	N/A	4.11010	6.57	18.50	85.52
		Concentration. (fmol)	41.53	88.78	206.03	312.500	267.26			
3	ESSTATDVF	PAR (mean)[a]	0.0846	0.1006	0.1379	N/A	1.20157	16.69	9.69	112.79
		Concentration. (fmol)	35.76	43.43	61.90	625.00	704.93			
4	EEVRKGNL	PAR (mean)[a]	0.1373	0.1898	0.3125	N/A	1.31870	7.23	15.05	103.95
		Concentration. (fmol)	8.30	11.51	19.02	78.13	81.21			

Abbreviations: LOD, limit of detection; LOQ, limit of quantification; LLOQ, lower limit of quantification; PAR, peak area ratio; S/N, Signal-to-noise; N/A, not applicable.

[a] Calculated by dividing the peak area of stable isotope-labeled (SIS, heavy) peptide by that of endogenous (light) peptide.

[b] Calculated based on the averaged peak area ratio plus 3 times the standard deviation (SD) of a triplicate-analyzed zero sample.

[c] Calculated based on the averaged peak area ratio plus 10 times the standard deviation (SD) of a triplicate-analyzed zero sample.

[d] LLOQ is the lowest concentration of the response curve that meets the FDA, EMA criteria: Signal-to-noise (S/N) >5, Precision < 20%, Accuracy within 20%

[e] Concentrations are calculated by inputting the PAR values into the reversed response curve equations.

[f] Calculated by dividing the heavy-peptide peak area at LLOQ by that of the zero sample.

[g] Precision of the triplicate analysis of each LLOQ sample, coefficient of variation (%).

[h] Accuracy (%) = back-calculated concentration at the LLOQ sample / nominal concentration of the LLOQ × 100.

Table 2-7. Analytical specificity (interference) of six different matrices by MRM-MS assay.

Sample Name	Glu-peptides	ANTFLEEVR		ERECVEETCSY		ESSTATDVF		EEVRKGNL	
	Replicate	Peak area (heavy)[a]	Interference (%)[b]	Peak area (heavy)[a]	Interference (%)[b]	Peak area (heavy)[a]	Interference (%)[b]	Peak area (heavy)[a]	Interference (%)[b]
LLOQ[c]	1	7633		5306		2820		4476	
	2	7825	N/A	5735	N/A	2755	N/A	4568	N/A
	3	5887		5690		2921		5069	
	Mean	7115		5577		2832		4704	
Matrix 1	1	7	0.09	263	4.96	59	2.09	253	5.65
	2	344	4.40	163	2.84	155	5.63	107	2.34
	3	60	1.02	266	4.67	316	10.82	356	7.02
	Mean	137	**1.84**	231	**4.16**	177	**6.18**	239	**5.01**
Matrix2	1	146	1.91	400	7.54	42	1.49	223	4.98
	2	541	6.91	82	1.43	80	2.90	386	8.45
	3	988	16.78	267	4.69	103	3.53	206	4.06

	Mean	558	**8.54**	250	**4.55**	75	**2.64**	272	**5.83**
Matrix 3	**1**	751	9.84	80	1.51	190	6.74	188	4.20
	2	303	3.87	31	0.54	108	3.92	317	6.94
	3	207	3.52	79	1.39	100	3.42	177	3.49
	Mean	420	**5.74**	63	**1.15**	133	**4.69**	227	**4.88**
Matrix 4	**1**	145	1.90	221	4.17	167	5.92	541	12.09
	2	327	4.18	76	1.33	106	3.85	384	8.41
	3	438	7.44	289	5.08	501	17.15	396	7.81
	Mean	303	**4.51**	195	**3.52**	258	**8.97**	440	**9.44**
Matrix 5	**1**	52	0.68	24	0.45	325	11.52	328	7.33
	2	355	4.54	14	0.24	150	5.44	142	3.11
	3	878	14.91	463	8.14	414	14.17	269	5.31
	Mean	428	**6.71**	167	**2.94**	296	**10.38**	246	**5.25**
Matrix 6	**1**	1246	16.32	235	4.43	408	14.47	335	7.48
	2	456	5.83	55	0.96	168	6.10	481	10.53

3	650	11.04	163	2.86	454	15.54	691	13.63
Mean	784	**11.06**	151	**2.75**	343	**12.04**	502	**10.55**
Average Interference (%)		**6.40**		**3.18**		**7.48**		**6.82**

Abbreviations: LLOQ, lower limit of quantification

[a] The peak area was from interferents corresponding to the m/z values of the designated heavy peptides for matrix samples.

[b] Calculated by dividing the peak area of each matrix sample by that of the LLOQ and multiplying the result by 100.

[c] The matrix used for the LLOQ analysis is a pooled serum sample from hepatitis patients.

Table 2-8. Carryover measurements using zero sample run after the highest calibrator.

No.	Peptide Sequence	Replicate	Peak area of LLOQ	Peak area of zero sample after ULOQ[a]	Carryover (%)[b]	Average carryover (%)
		1	7633	632	8.28	
1	**ANTFLEEVR**	**2**	7825	700	8.95	**6.46**
		3	5887	127	2.16	
		1	5306	35	0.66	
2	**ERECVEETCSY**	**2**	5735	380	6.63	**3.28**
		3	5690	146	2.57	
		1	2820	185	6.56	
3	**ESSTATDVF**	**2**	2755	194	7.04	**6.37**
		3	2921	161	5.51	
		1	4476	567	12.67	
4	**EEVRKGNL**	**2**	4568	586	12.83	**12.10**
		3	5069	548	10.81	

Abbreviations: LLOQ, lower limit of quantification; ULOQ, Upper limit of quantification.

[a] The peak area was from interferents corresponding to the m/z of the designated heavy peptides.

[b] Calculated by dividing the peak area of the zero sample after ULOQ by that of the LLOQ and multiplying the result by 100.

2.3.2. Reproducibility of the MRM-MS assay using the multi-enzyme digestion workflow

I evaluated the reproducibility of the total MRM-MS assay that used the multienzyme digestion workflow with depleted pooled HCC serum. The serum was prepared over 5 days and analyzed daily in triplicate. The average CV values of each target peptide were under 20% in both the intra-assay and interassay analyses, as shown in **Table 2-9**. The average CV values of the intra-assay analysis ranged from 7.03% to 17.35%. The corresponding values for the interassay analysis ranged from 14.81% to 17.67%. These results demonstrate that the total MRM-MS assay using the multienzyme digestion workflow is stable for the quantitation of four peptides over several days.

Table 2-9. Reproducibility of the MRM-MS assay over 5 days.

	ANTFLEEVR		ERECVEETCSY		ESSTATDVF		EEVRKGNL	
Intra-assay variation (n = 3)	**Average PAR[a]**	**CV (%)**	**Average PAR[a]**	**CV (%)**	**Average PAR[a]**	**CV (%)**	**Average PAR[a]**	**CV (%)**
day1	2.05	12.09	0.40	15.15	1.35	7.60	0.49	3.12
day2	2.37	11.10	0.37	16.59	1.27	18.54	0.55	8.36
day3	1.81	9.12	0.39	20.23	0.96	2.42	0.53	8.90
day4	2.58	13.35	0.50	16.69	1.50	2.94	0.37	8.93
day5	2.36	8.33	0.48	18.10	1.23	10.71	0.39	5.85
Average	**2.23**	**10.80**	**0.43**	**17.35**	**1.26**	**8.44**	**0.47**	**7.03**
Inter-assay variation (n = 5)	**Average PAR[a]**	**CV (%)**	**Average PAR[a]**	**CV (%)**	**Average PAR[a]**	**CV (%)**	**Average PAR[a]**	**CV (%)**
injection1	2.47	13.66	0.38	19.38	1.26	16.48	0.46	9.34
injection2	2.12	16.78	0.49	19.40	1.18	16.91	0.45	19.35
injection3	2.11	13.99	0.42	14.22	1.34	16.04	0.48	20.81
Average	**2.23**	**14.81**	**0.43**	**17.67**	**1.26**	**16.48**	**0.47**	**16.50**

Abbreviations: CV, coefficient of variation; PAR, peak area ratio

[a] Calculated by dividing the peak area of stable isotope-labeled (SIS, heavy) peptide by that of endogenous (light) peptide.

2.3.3. Construction of the DCP multi Glu-peptides panel

As a result of the logistic regression analysis performed on the training set, a multi-Glu-peptide panel containing three Glu-peptides (ANTFLEEVR, ERECVEETCSY, and ESSTATDVF) was established as the best panel for predicting HCC, eliminating a nonsignificant Glu-peptide (EEVRKGNL). The three Glu-peptides contributed significantly to the panel ($P < 0.005$), as indicated by the final logistic model, which is detailed in **Table 2-10**. The three Glu-peptides panel obtained an AUROC of 0.873 (95% confidence interval [CI], 0.818-0.928) (**Figure 2-6A**). The constructed panel achieved a greater AUROC value than each individual Glu-peptide (AUROC values: 0.801 for ERECVEETCSY, 0.734 for ESSTATDVF, and 0.561 for ANTFLEEVR; **Figure 2-6A**). The predictive performance of the three Glu-peptides panel was consistent in the test set, with an AUROC value of 0.844 (95% CI, 0.761-0.928), which was equivalent to the AUROC value of 0.873 obtained for the training set (DeLong`s test, $P = 0.5722$; **Figure 2-6B**).

The levels of three peptides in the 300 individual samples were plotted as scatter dot plots with lines showing the mean and SD, separately for the training and test sets (**Figure 2-7**). By the statistical tests, the levels of the three Glu-peptides were significantly altered in cases compared with controls ($P < 0.05$), except for ANTFLEEVR in the training set ($P = 0.103$). Notably, the level of the ANTFLEEVR peptide was significantly decreased in the HCC case group compared with that in the control group, whereas the levels of the other two peptides were significantly elevated in the HCC case group.

Table 2-10. Summary of the DCP 3-Glu-peptide logistic regression model

	Estimate	Standard error	z value	*P*-value
(Intercepts)	−5.202	1.0066	−5.168	2.37E-07
ANTFLEEVR	−0.9199	0.3111	−2.957	0.00311
ERECVEETCSY	11.1108	1.8697	5.942	2.81E-09
ESSTATDVF	2.1785	0.5023	4.337	1.44E-05

Note: A logistic regression model to predict the probability of having HCC (P) was built with the following equation: logit(P) = log(P/(1 − P)) = −5.202 - 0.9199 × ANTFLEEVR+11.1108× ERECVEETCSY+2.1785× ESSTATDVF. The numeric values of each peptide in the equation were raw values for relative concentrations (peak area ratio of endogenous light peptides to heavy SIS peptides). The optimal cutoff value for the above equation is 0.432.

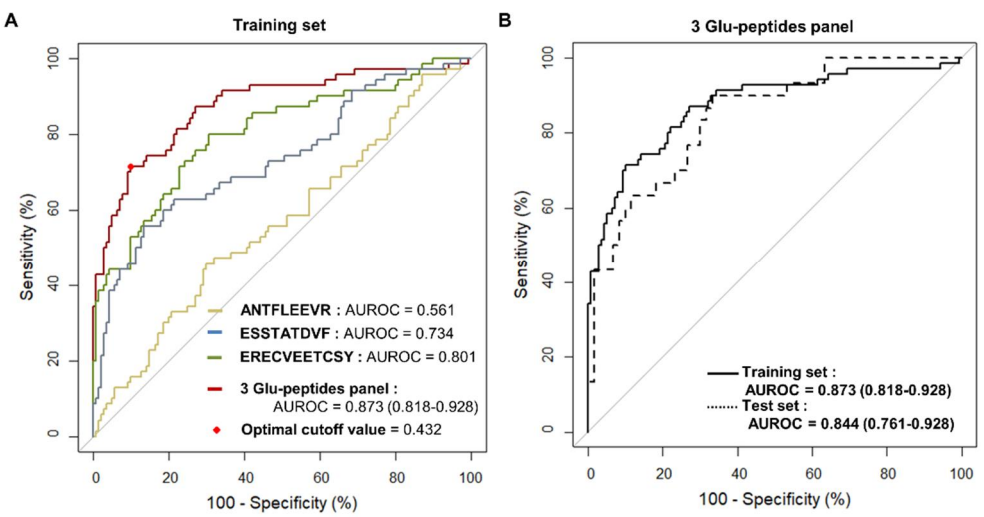


Figure 2-6. Assessment of the diagnostic performance of the DCP 3 Glu-peptide panel in the training and test sets

(A) The ROCs for each Glu-peptide consisting of the 3 Glu-peptides panel were analyzed. The ERECVEETCSY peptide (green line) and ESSTATDVF peptide (blue line) showed acceptable AUROC values ranging from 0.7 to 0.8, whereas the ANTFEEVR peptide (yellow line) had lower discrimination power alone. The combination of all three Glu-peptides (red line) resulted in enhanced diagnostic performance, as shown by the AUROC value of 0.873 (Delong's test, $P < 0.005$ for all comparisons to the AUROC values of each Glu-peptide individually). The optimal cutoff value of the three Glu-peptides panel was 0.432 (red point), presented by the Youden Index for the training set. (B) The ROC curves for the DCP 3 Glu-peptide panel in the training (solid black line) and test sets (dotted black line). The AUROC of the DCP 3-Glu-peptide panel was 0.873 in the training set, which was consistent with the test set, which had an AUROC of 0.844 (DeLong`s test, $P = 0.5722$). All of the AUROC values were summarized with 95% CI for ROC curves. Abbreviations: AUROC, area under the ROC curve; CI, confidence interval; ROC, receiver operating characteristic.

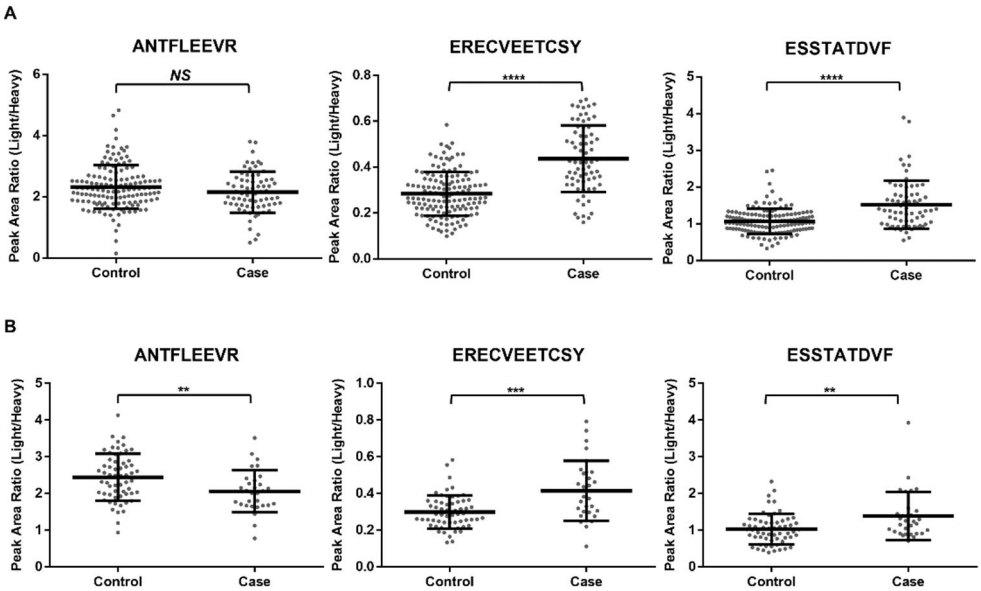


Figure 2-7. Comparison of the relative peptide levels of the three Glu-peptides for the training set (A) and the test set (B).

The relative concentrations of three Glu-peptides were plotted as peak area ratios of light peptides to heavy SIS peptides for individual patients. Middle horizon lines and error bars indicate the mean and the standard deviation, respectively. The relative concentrations of ANTFLEEVR peptide were significantly lower in cases than in controls, whereas those for ERECVEETCSY and ESSTATDVF were significantly higher in cases than in controls. The Mann-Whitney U test was used to calculate P-values for the comparisons of relative concentrations between the ANTFEEVR peptide in the training set, the ERECVEETCSY peptide in the test set, and the ESSTATDVF peptide in both the training and test sets. Welch's t-test was used to calculate *P*-values for comparisons of the relative concentrations of the ERECVEETCSY peptide in the training set. The relative concentrations of the ANTFEEVR peptide in the test set were compared by Student`s t-test. *NS* is no significant difference between groups, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

2.3.4. Comparison of the MRM-MS assay and the immunoassay

To assess whether the diagnostic performance of the DCP three Glu-peptides panel based on the MRM-MS assay was comparable to the diagnostic performance of measuring serum DCP levels using the immunoassay, the confusion matrixes were analyzed at the optimal cutoff (**Figure 2-8**). The optimal cutoff of the serum DCP level was 40 mAU/mL, whereas that of the three Glu-peptides panel was 0.432 (corresponding probability = 0.606), as determined by the Youden Index of the training set. The MRM-MS assay provided higher sensitivity and accuracy than the immunoassay. The sensitivities of the MRM-MS assay were 52.9% and 56.7% for the training and test sets, respectively, whereas those for the immunoassay were 48.6% and 43.3%. This suggests that the MRM-MS assay can work more favorably for surveillance diagnosis than the conventional DCP assay. The accuracies of the MRM-MS assay were 81.4% (95% CI, 75.5%-86.5%) and 80.0% (95% CI, 70.3%-87.7%) for the training and test sets, respectively, whereas those for the immunoassay were 75.8% (95% CI, 68.2%-82.5%) and 73.9% (95% CI, 61.9%-83.8%). Further, the DCP levels by immunoassay and the logit(P) values from the DCP three-Glu-peptide panel correlated weakly (Pearson`s correlation, $R = 0.24$; $P = 2.3e\text{-}05$), as shown in **Figure 2-9**.

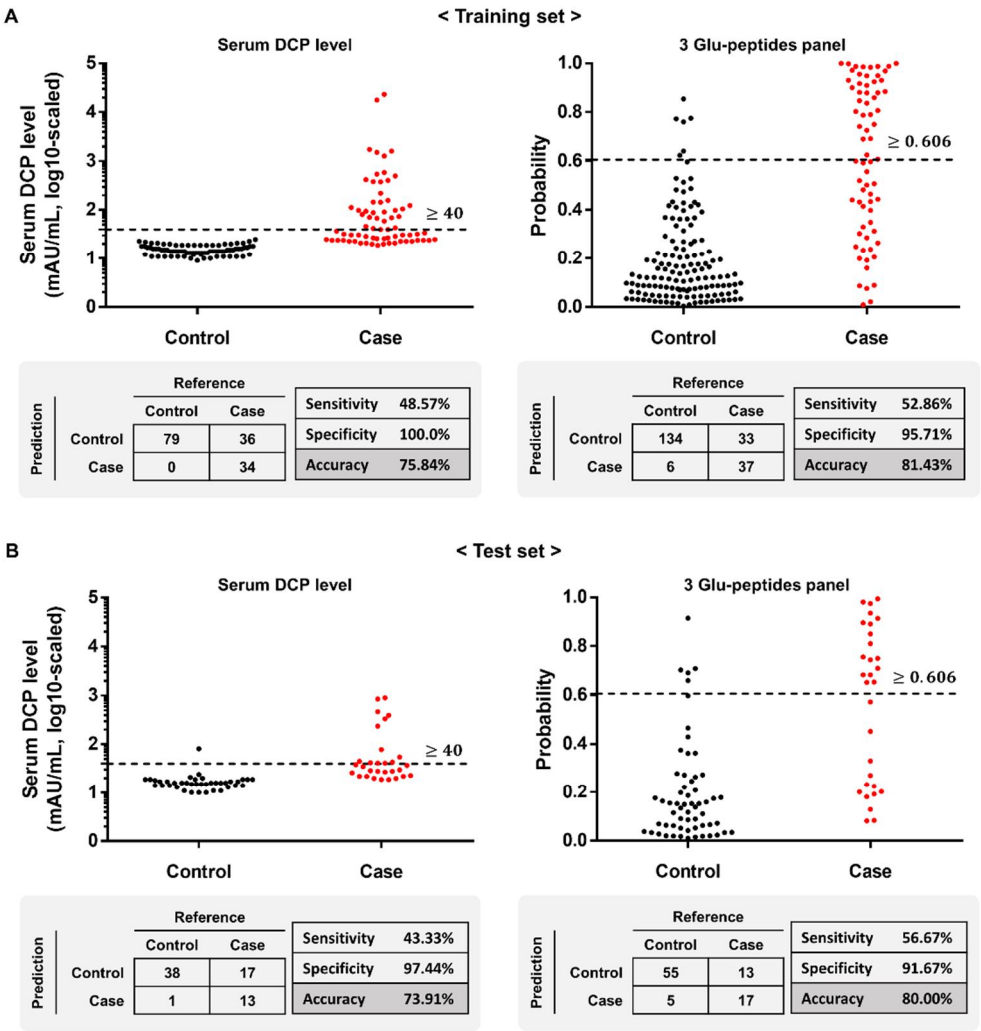


Figure 2-8. Interactive dot plots for the HCC diagnosis performance of the serum DCP level (measured by the immunoassay) and the DCP panel (measured by the MRM-MS) in the training set (A) and the test set (B).

Horizontal dotted lines denote the optimal cutoff values. The serum DCP levels were not available for 65 and 21 at-risk patients in the training set and the test set, respectively. The DCP panel (3 Glu-peptides) could achieve the accuracy higher than 80% at the optimum cutoff which maximizes the sensitivity and specificity.

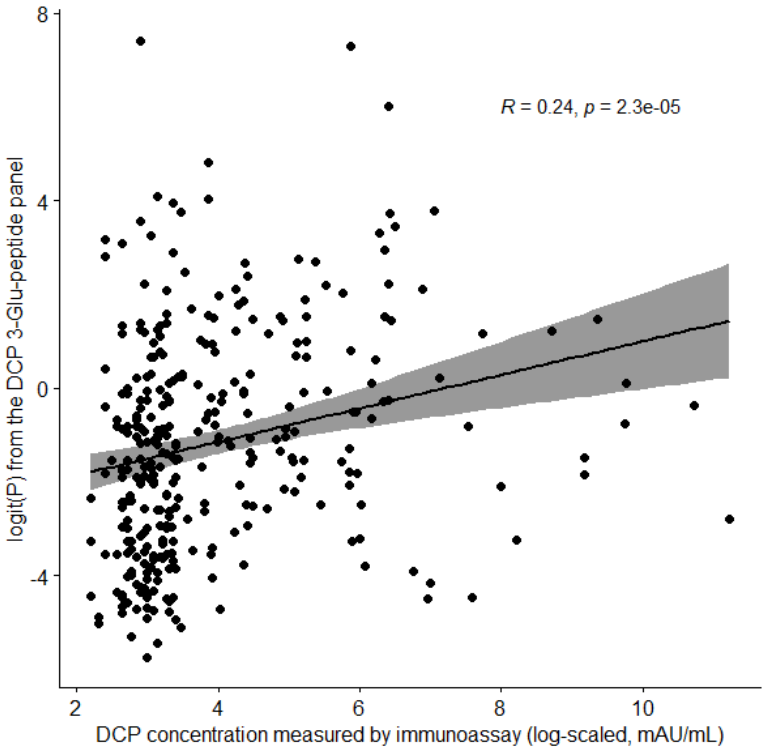


Figure 2-9. Correlation between the DCP levels by immunoassay and logit(P) values, calculated based on the DCP panel.

I analyzed the correlation of standard DCP values, measured by immunoassay, with logit(P) values, calculated based on the DCP 3 Glu-peptides panel in the external validation set. By Pearson`s correlation, there was a weak correlation between the standard DCP and DCP 3 Glu-peptides panels (R = 0.24, P = 2.3e-05). The fitted regression line and observations are presented as line and dots, respectively. The grey area indicates the 95% confidence interval (CI).

2.3.5. Combined model of the DCP 3 Glu-peptides panel and the serum AFP levels

I conducted further logistic regression analyses to determine whether combining the DCP panel with serum AFP levels could enhance the predictive power for HCC detection. The combined model using both the three Glu-peptides panel and serum AFP levels increased the AUROC values to 0.903 (95% CI, 0.855-0.952) for the training set (**Figure 2-10A**). The combined model outperformed serum AFP levels (AUROC, 0.770; 95% CI, 0.698-0.842) for the training set, based on DeLong`s test ($P < 0.05$). Similarly, the AUROC value of the combined model significantly increased from 0.844 to 0.913 (95% CI, 0.851-0.974) for the test set (**Figure 2-10B**). The diagnostic performances including AUROC, sensitivity, specificity, and diagnostic accuracy of each model in each data set were summarized in **Figure 2-11**. Notably, the combined model had greater sensitivity in both the training and test sets (68.9% and 76.7%, respectively) compared with the low sensitivity of serum AFP levels alone (35.7% and 56.7%, respectively), as shown in **Figure 2-11C**.

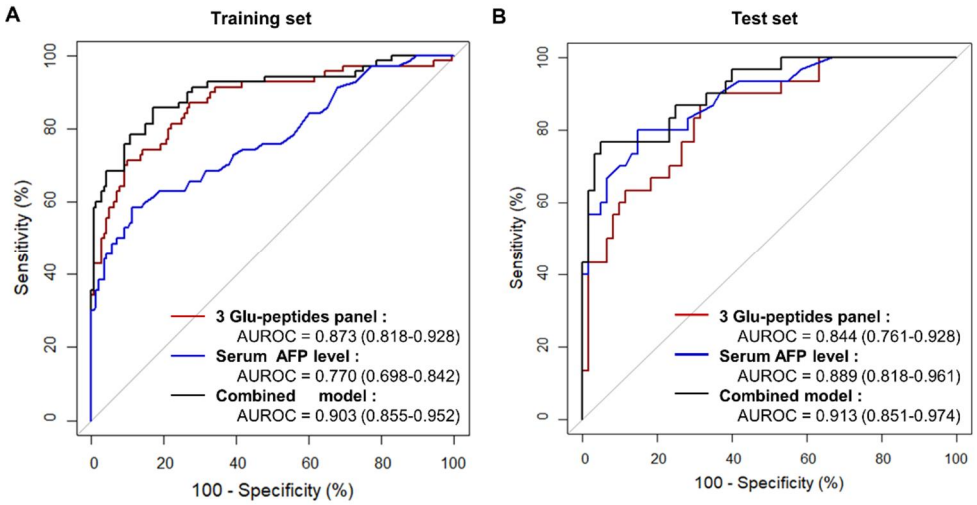


Figure 2-10. Comparison of the diagnostic performances of the DCP 3 Glu-peptides panel, serum AFP levels, and a combined model (DCP 3 Glu-peptides panel and serum AFP level) in the training (A) and test sets (B).

The AUROC values were presented with 95% CI for ROC curves. The combined model (black lines) had a higher AUROC value than the DCP 3 Glu-peptides panel (red lines) or serum AFP levels (blue lines) alone, for all data sets. (A) In the training set, the combined model had a higher AUROC value (0.903) than either AFP levels (0.770) or DCP 3 Glu-peptides panel (0.873) alone. The AUROC value for the combined model was statistically different from that for AFP levels (DeLong`s test, $P < 0.05$) but not from that for the DCP 3 Glu-peptides panel (DeLong`s test, $P = 0.079$). (B) In the test set, the combined model also had a higher AUROC value (0.913) than both the AFP levels (0.889) and the DCP 3 Glu-peptides panel (0.844). The AUROC value for the combined model was statistically different from that for the DCP 3 Glu-peptides panel (DeLong`s test, $P < 0.05$) but not from that for AFP levels (DeLong`s test, $P = 0.484$). Abbreviations: AUROC, area under the ROC curve; CI, confidence interval; ROC, receiver operating characteristic.

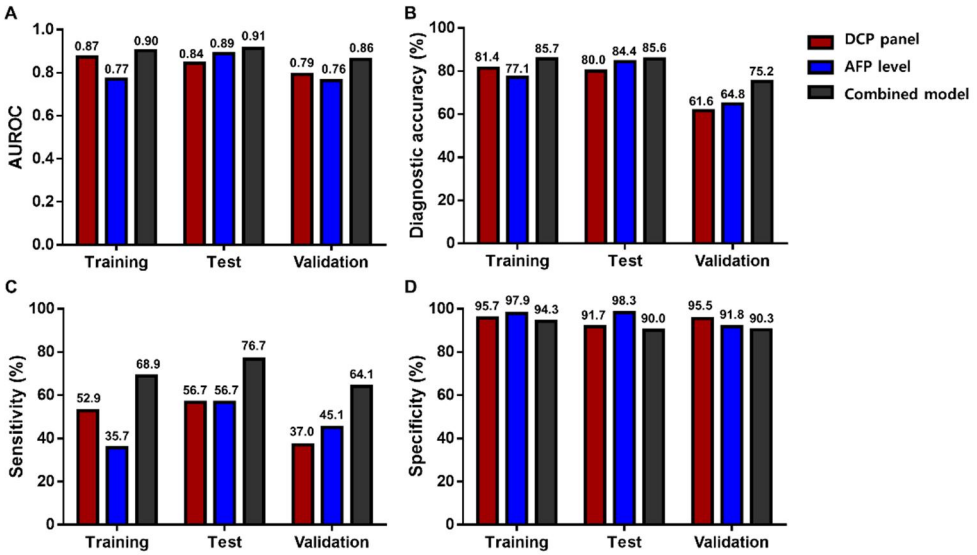


Figure 2-11. Diagnostic performances of the DCP 3 Glu-peptides panel, serum AFP levels, and a combined model.

(A) Area under the receiver operating characteristic curve (AUROC) values, (B) diagnostic accuracy, (C) sensitivity, and (D) specificity for the DCP 3 Glu-peptides panel (red bars), serum AFP levels (blue bars), and the combined model of them (black bars). All diagnostic performances were analyzed using the optimal cutoff value derived from the training set.

2.3.6. Validation of the DCP 3 Glu-peptides panel and the combined model with AFP levels in an external cohort

I analyzed another 318 samples from an external cohort as an independent validation set; this consisted of 134 controls and 184 cases (**Table 2-1**). The AUROC values of the three Glu-peptides panel and the combined model with AFP levels were 0.793 (95% CI, 0.745-0.842) and 0.863 (95% CI, 0.822-0.903), respectively, for the independent validation set (**Figure 2-12A**). The AUROC values for the three Glu-peptides panel and the combined model were statistically equivalent to those identified in the test set (0.844 and 0.913, respectively) based on the results of DeLong`s test ($P > 0.05$). Moreover, the Mann-Whitney U test revealed that the levels of each peptide were also significantly different between the control and case 11 groups, as shown in **Figure 2-12B** ($P < 0.0001$). The combined model had greater sensitivity compared with serum AFP levels alone in the independent validation set (45.1% to 64.1%), whereas the other diagnostic performances of the panel remained equivalent, as detailed in **Table 2-11**.

I examined the diagnostic abilities of the three Glu-peptides panel in the AFP-negative and DCP-negative subgroup, consisting of 127 patients at risk and 39 patients with HCC, with AFP and DCP levels below the reference values. The AUROC values of the three Glu-peptides panel and the combined model with AFP levels were 0.803 (95% CI, 0.726-0.880) and 0.821 (95% CI, 0.739-0.903), respectively, for the AFP-negative and DCP-negative subgroup (**Figure 2-13**). Notably, the three Glu-peptides panel could discriminate 18 patients with HCC, corresponding to approximately half of the 39 patients with HCC with AFP and DCP levels below the reference values, reducing the false-negative rate.

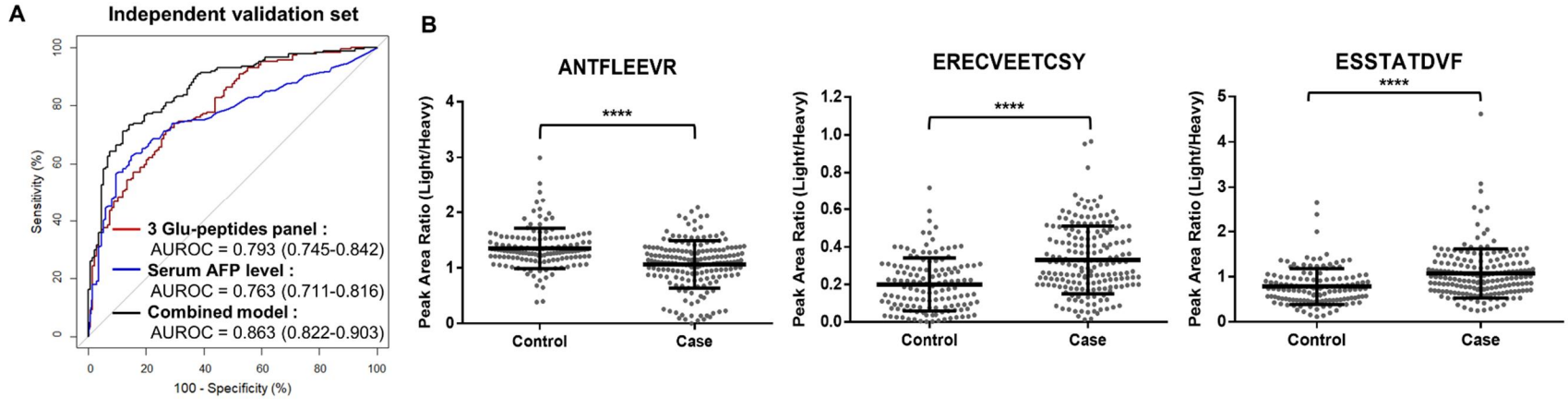


Figure 2-12. Independent validation of the diagnostic performance of the DCP 3 Glu-peptides panel, serum AFP levels, and the combined model (DCP 3 Glu-peptides panel and serum AFP level).

(A) The ROC curves for the DCP 3 Glu-peptides panel (red lines), serum AFP levels (blue lines), and the combined model (black lines) for the independent validation set. The AUROC values were presented with 95% CI for ROC curves. The differences between the AUROC value for the combined model and those for the DCP 3 Glu-peptides panel and serum AFP levels were significant for the independent validation set (DeLong`s test, $P < 0.005$). The diagnostic performance of the DCP panel was consistent in the independent validation set, although the performance was slightly higher in the training set (AUROC = 0.873) than in the independent validation set (DeLong`s test, $P = 0.0336$). (B) The relative concentrations of three Glu-peptides in cohort B. The tendencies for the relative concentrations of each peptide were consistent in the independent

validation set with those observed in the training set. The relative concentrations of three Glu-peptides were plotted as peak area ratios of the light peptides to heavy SIS peptides for individual patients. Middle horizon lines and error bars indicate the mean and the standard deviation, respectively. *P*-values were calculated using the Mann-Whitney U test to compare the relative concentrations of each peptide. ****$P < 0.0001$. Abbreviations: AUROC, area under the ROC curve; CI, confidence interval; ROC, receiver operating characteristic.

Table 2-11. Comparison of diagnostic performance of DCP panel (3 Glu-peptides), AFP level, and combined model in each data set

Diagnostic performance	AUROC (95% CI)		Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (95% CI)
Training set							
		P-value					
DCP panel (3 Glu-peptides)	0.873 (0.818-0.928)		52.9	95.7	86.1	80.2	0.814 (0.755-0.865)
AFP level	0.770 (0.698-0.842)	*<0.05* †	35.7	97.9	89.3	75.3	0.771 (0.709-0.826)
DCP panel (3 Glu-peptides) + AFP level	0.903 (0.855-0.952)	*0.079* ‡	68.9	94.3	85.7	85.7	0.857 (0.802-0.902)
Test set							
		P-value					
DCP panel (3 Glu-peptides)	0.844 (0.761-0.928)		56.7	91.7	77.3	80.9	0.800 (0.703-0.877)
AFP level	0.889 (0.818-0.961)	*0.393* †	56.7	98.3	94.4	81.9	0.844 (0.7528-0.912)
DCP panel (3 Glu-peptides) + AFP level	0.913 (0.851-0.974)	*<0.05* ‡	76.7	90.0	79.3	88.5	0.856 (0.766-0.921)
Independent validation set							
		P-value					
DCP panel (3 Glu-peptides)	0.793 (0.745-0.842)		37.0	95.5	91.9	52.5	0.616 (0.561-0.670)

AFP level	0.764 (0.711-0.816)	*0.408*	†	45.1	91.8	88.3	54.9	0.648 (0.593-0.700)
DCP panel (3 Glu-peptides) + AFP level	0.862 (0.822-0.903)	*<0.05*	‡	64.1	90.3	90.1	64.7	0.752 (0.700-0.798)

†DCP panel (3 Glu-peptides) versus AFP level. ‡ DCP panel (3 Glu-peptides) with versus without AFP level. *P*-values less than 0.05 are shown as '<0.05' Abbreviations: CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value

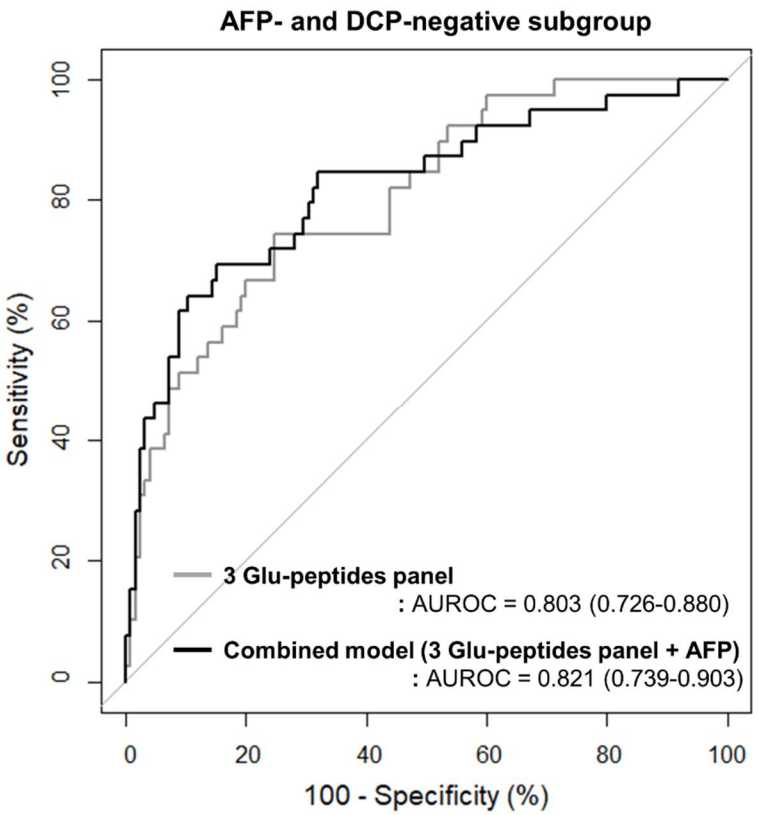


Figure 2-13. The HCC diagnostic abilities of the DCP 3 Glu-peptides panel and the combined model (DCP 3 Glu-peptides panel and serum AFP level) in AFP- and DCP-negative subgroup of the independent validation set.

The ROC curves for the AFP- and DCP-negative subgroup, consisting of 127 at-risk patients and 39 HCC patients with < 20 ng/mL for AFP and < 40 mAU/mL for DCP. All AUROC values are summarized with 95% CI for the ROC curves. The 3 Glu-peptides panel (gray lines) and the combined model (black lines) with serum AFP levels showed reliable performance for HCC patients with serum AFP and DCP values under the corresponding reference values, as indicated by AUROC values over 0.8. Abbreviations: AUROC, area under the ROC curve; CI, confidence interval; ROC, receiver operating characteristic.

2.3.7. HCC Surveillance abilities of the DCP 3 Glu-peptides panel and the combined model (the DCP 3 Glu-peptides panel with the serum AFP level)

To evaluate the surveillance abilities for HCC, the diagnostic performance of the DCP 3 Glu-peptides panel for detecting very-early-stage HCC (BCLC 0, single lesion <2 cm) or early-stage HCC (BCLC stage A, a single 2-5 cm lesion or two to three lesions that are each <3 cm) were analyzed in the independent validation set (**Figure 2-14**).

The subgroup of HCC very-early- or early-stage consisted of 159 cases (36 patients in BCLC 0 and 123 patients in BCLC stage A). According to the ROC analysis results for the detection of very-early- or early-stage patients with HCC from among the 134 at-risk controls, the AUROC value of the three Glu-peptides panel was 0.795 (95% CI, 0.745-0.845). Further, a combined model using both serum AFP levels and the three Glu-peptides panel had significantly greater diagnostic power for under early-stage HCC, with an AUROC of 0.864 (95% CI, 0.822-0.906). According to the DeLong`s test, the diagnostic power of the three Glu-peptides panel and combined model in discriminating very-early- or early-stage HCC was comparable with the overall performance in the total validation set (P = 0.9708).

The very-early-stage HCC subgroup in the independent validation set consisted of 36 cases. According to the ROC analysis results for the detection of very-early-stage patients with HCC from among the 134 at-risk controls, the AUROC value of the three Glu-peptides panel was 0.825 (95% CI, 0.748-0.902). Further, a combined model using both serum AFP levels and the three Glu-peptides panel had

significantly greater diagnostic power for very-early-stage HCC, with an AUROC of 0.896 (95% CI, 0.840-0.953). According to the DeLong`s test, the diagnostic power of the three Glu-peptides panel and combined model in discriminating very-early-stage HCC was comparable with the overall performance in the original validation set (P = 0.4996). Further, no significant differences were observed between the AUROC values of each panel for distinguishing very-early-stage HCC from control subgroups that were stratified by etiology and the entire control group (DeLong`s test, $P > 0.05$), as shown in **Table 2-12**.

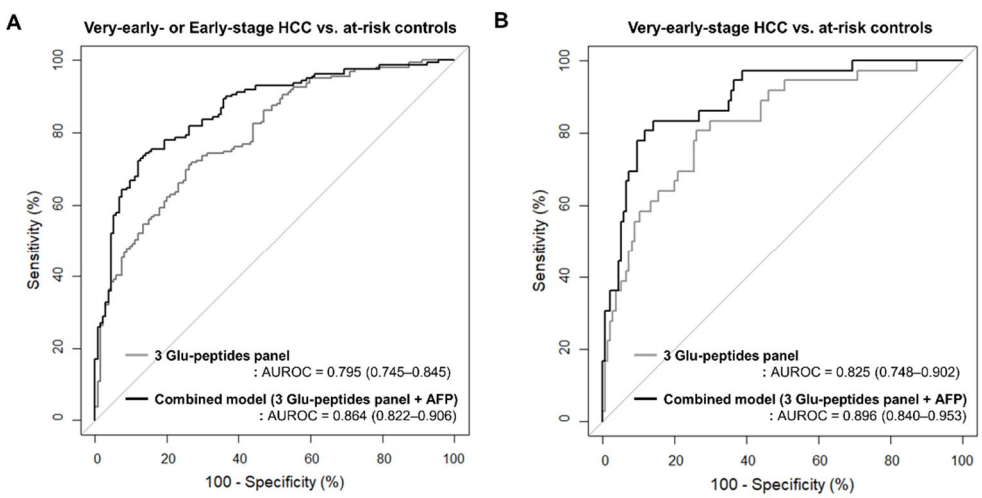


Figure 2-14. The HCC surveillance abilities of the DCP 3 Glu-peptides panel and the combined model (DCP 3 Glu-peptides panel and serum AFP level) in the independent validation set.

(A) The ROC curves for the discrimination of very-early- or early-stage HCC (BCLC 0 or A) cases, consisting of 159 patients from 134 at-risk controls. (B) The ROC curves for the discrimination of very-early-stage HCC (BCLC 0, tumor size < 2 cm) cases, consisting of 36 patients from 134 at-risk controls. All AUROC values

are summarized with 95% CI for the ROC curves. The 3 Glu-peptides panel (gray lines) and the combined model (black lines) with serum AFP levels showed reliable surveillance performance for very-early-stage HCC patients as indicated by AUROC values over 0.8.

Table 2-12. Comparison of diagnostic performance of 3 Glu-peptides panel, AFP level, and combined model for discriminating early-stage HCC from stratified controls.

Diagnostic performance	Very-early-stage HCC (BCLC 0) *(n = 36)*								Very-early- or Early-stage HCC (BCLC 0 and A) *(n = 159)*							
	AUROC (95% CI)			Sens. (%)	Spec. (%)	PPV (%)	NPV (%)	Accuracy (95% CI)	AUROC (95% CI)			Sens. (%)	Spec. (%)	PPV (%)	NPV (%)	Accuracy (95% CI)
Controls (n = 134)																
		P-value								***P-value***						
DCP panel (3 Glu-peptides)	0.825			36.1	95.5	68.4	84.8	0.829	0.795			38.4	95.5	91.0	56.6	0.645
	(0.748-0.902)							(0.764-0.883)	(0.745-0.845)							(0.587-0.700)
AFP value	0.764	*0.376*	†	50.0	91.8	62.1	87.2	0.829	0.754	*0.293*	†	44.7	91.8	86.6	58.3	0.662
	(0.659-0.869)							(0.764-0.883)	(0.698-0.811)							(0.605-0.716)
DCP panel (3 Glu-peptides) + AFP value	0.896	*< 0.05*	‡	69.4	90.3	65.8	91.7	0.859	0.864	*< 0.05*	‡	64.8	90.3	88.8	68.4	0.765
	(0.840-0.953)							(0.797-0.907)	(0.822-0.906)							(0.712-0.812)
CHB with no LC (n = 92)																
		P-value								***P-value***						
DCP panel (3 Glu-peptides)	0.835			36.1	97.8	86.7	79.7	0.805	0.803			38.4	97.8	96.8	47.9	0.602
	(0.755-0.915)							(0.725-0.869)	(0.750-0.857)							(0.538-0.663)
AFP value	0.793	*0.547*	†	50.0	93.5	75.0	82.7	0.813	0.787	*0.686*	†	44.7	93.5	92.2	49.4	0.626
	(0.731-0.844)							(0.734-0.876)	(0.731-0.844)							(0.562-0.686)
DCP panel (3 Glu-peptides) + AFP value	0.915	*< 0.05*	‡	69.4	94.6	83.3	88.8	0.875	0.884	*< 0.05*	‡	64.8	94.6	95.4	60.8	0.757
	(0.861-0.969)							(0.805-0.927)	(0.842-0.926)							(0.699-0.809)

CHC with no LC (n = 13)														
		P-value								***P-value***				
DCP panel (3 Glu-peptides)	0.729		36.1	84.6	86.7	32.4	0.490	0.687		38.4	84.6	96.8	10.1	0.419
	(0.557-0.900)						(0.344-0.637)	(0.536-0.837)						(0.344-0.496)
AFP value	0.682	*0.683* †	50.0	100.0	100.0	41.9	0.813	0.663	*0.799* †	44.7	100.0	100.0	12.9	0.488
	(0.537-0.846)						(0.734-0.876)	(0.574-0.753)						(0.412-0.566)
DCP panel (3 Glu-peptides) + AFP value	0.840	*< 0.05* ‡	69.4	84.6	92.6	50.0	0.735	0.801	*< 0.05* ‡	64.8	84.6	98.1	16.4	0.663
	(0.714-0.965)						(0.589-0.851)	(0.691-0.911)						(0.587-0.733)
LC (n = 29)														
		P-value								***P-value***				
DCP panel (3 Glu-peptides)	0.835		36.1	93.1	86.7	54.0	0.615	0.816		38.4	93.1	96.8	21.6	0.468
	(0.731-0.939)						(0.486-0.734)	(0.723-0.909)						(0.723-0.909)
AFP value	0.707	*0.140* †	50.0	82.8	78.3	57.1	0.646	0.691	*0.074* †	44.7	82.8	93.4	21.4	0.505
	(0.576-0.838)						(0.518-0.761)	(0.591-0.791)						(0.432-0.579)
DCP panel (3 Glu-peptides) + AFP value	0.861	*0.533* ‡	69.4	79.3	80.7	67.7	0.739	0.831	*0.650* ‡	64.8	79.3	94.5	29.1	0.670
	(0.769-0.953)						(0.615-0.840)	(0.749-0.914)						(0.598-0.737)

†DCP panel (3 Glu-peptides) versus AFP level. ‡ DCP panel (3 Glu-peptides) with versus without AFP level. *P*-values less than 0.05 are shown as '<0.05'. No significant differences were found between the AUROC values of each panel for distinguishing very-early-stage or early-stage HCC from stratified control subgroups and the entire control group (DeLong`s test, *P*-value > 0.05).

Abbreviations: CHB, chronic hepatitis B; CHC, chronic hepatitis C; CI, confidence interval; LC, liver cirrhosis; NPV, negative predictive value; PPV, positive predictive value; Sens., sensitivity; Spec., specificity

2.4. DISCUSSION

According to recent studies, the γ-carboxylation of the 10 Glu residues in the N-terminal Gla domain of DCP occurs in a specific order, resulting in blood DCP populations consisting of a heterogeneous mixture of ten possible proteoforms (**Figure 2-15**) (83, 84). To quantify the DCP proteoforms inclusively, I constructed a quantitative assay for DCP measurement to simultaneously monitor three non-carboxylated peptides within the Gla domain using the MRM-MS method. The three monitored Glu-peptides could be obtained from different subgroups of DCP proteoforms (**Table 2-13**). The ANTFLEEVR peptide (referred to as the 'ANT peptide') contains the 9th and 10th carboxylated residues and represents the subgroup of DCP containing more than two Glu residues (2 to 10 Glu residues). Similarly, the ESSTATDVF peptide (referred to as the 'ESS peptide') contains the 8th carboxylated Glu residue and represents the subgroup of DCP with more than three Glu residues (3 to 10 Glu residues). The ERECVEETCSY peptide (referred to as 'ERE peptide') contains the 3rd, 5th, 6th, and 7th carboxylated Glu residues and represents the subgroup of DCP with more than eight Glu residues (8 to 10 Glu residues). Therefore, the quantification assay developed in this study is able to detect both substantially des-carboxylated forms and less des-carboxylated forms within the same batch, requiring lower cost and labor than immunoassay detection methods and with the minimized potential for variations due to batch effects.

In the present study, the ERE peptide quantity significantly increased in the HCC group, implying that DCP proteoforms with more than eight Glu residues (8 to

10 Glu residues) are elevated in HCC patients relative to the control group (fold change = 1.70, $P < 0.0001$). The increasing tendency in DCP proteoforms observed for the HCC group was consistent in the subgroup containing a wider range of DCP proteoforms, as represented by the ESS peptide quantity (3 to 10 Glu residues; fold change = 1.35, $P < 0.0001$). However, the ANT peptide was elevated in control patients (fold change = 0.85, $P < 0.0001$). When considering that the ANT peptide targets the DCP with 2 Glu residues as well as the DCP proteoforms that were targeted by the ESS peptide (**Table 2-13**), the DCP with 2 Glu residues appeared to constitute a higher proportion of the DCP population in the control versus case group. Meanwhile, the DCP variants with 2 Glu residues have approximately half the activity of normal prothrombin (97, 98). Presumably, a larger portion of DCP variants with 2 Glu residues would be beneficial for benign liver diseases than for the HCC group, although the direct impact of prothrombin activity in the progression of HCC remains unknown. The combination of the relatively lower level of ANT peptides and a higher level of ERE or ESS peptide could be used to characterize the DCP proteoforms that are synthesized during HCC rather than benign liver diseases.

However, this study has some limitations. First, the assay was unable to cover the three Glu residues at positions 25, 26, and 29, which are located within either a long tryptic peptide with poor ionization or a short chymotryptic peptide with fewer than 6 amino acids. If alternative proteolytic enzymes are available to generate an appropriate peptide length for stable MS analysis, the quantification and discernment of additional DCP proteoforms might be possible. Second, an investigation of the missed cleavages should be performed, due to the abundance of glutamic acid residues in the Gla domain. Previous studies had discussed the frequent occurrence

of missed cleavage when there is a large amount of glutamic acid residue in the proximity of cleavage sites (99-103). The information theory approach (103) predicted the missed cleavage score for ANTFLEEVR peptide to be 0.2, which was lower than the threshold of 0.25 for predicting missed cleavage sites. However, the fidelity analysis of protease digestion is necessary to increase the credibility of quantification results obtained from MRM-MS assay in further study as illustrated in **Figure 2-16**. Further, the cohorts in this study consisted solely of individuals of Korean ethnicity, primarily with an HBV etiology. Therefore, additional studies are needed to validate the assay using different populations, consisting of other ethnicities and etiologies. Analytical validations to confirm the robustness and reproducibility of the MRM-MS assay should be evaluated using larger cohorts in future studies.

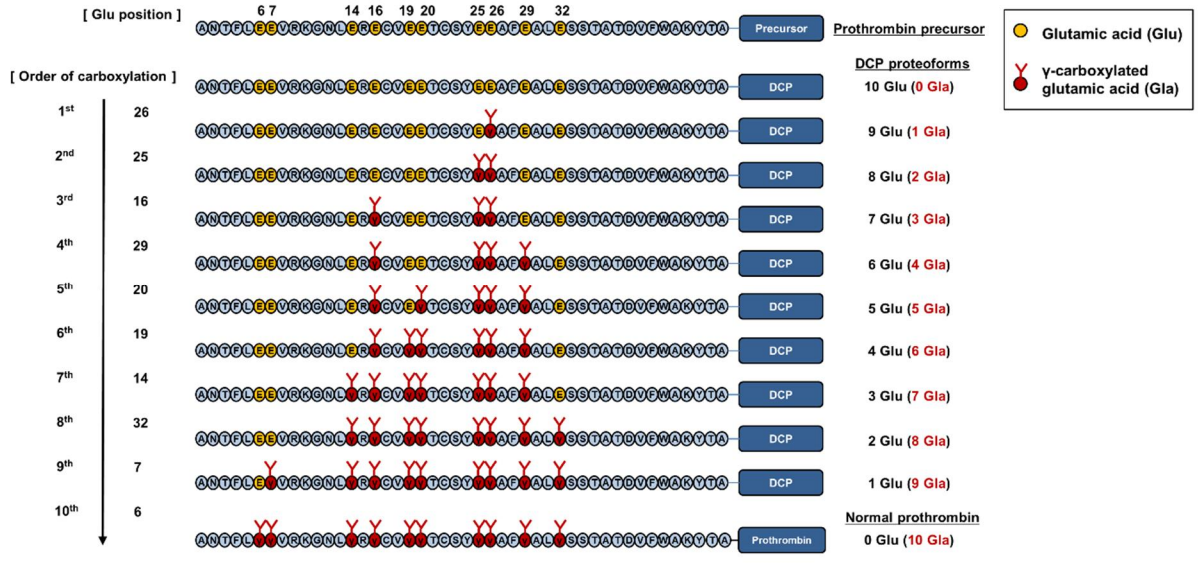


Figure 2-15. The specific order of carboxylation modifications for ten Glu residues resulting in the ten possible intermediate DCP proteoforms.

The ANTFLEEVR peptide contains the 9th and 10th carboxylated Glu residues (positions 7 and 6 of the Gla domain, respectively). The ESSTATDVF peptide contains the 8th carboxylated Glu residue (position 32 of the Gla domain). The ERECVEETCSY peptide contains the 3rd, 5th, 6th, and 7th carboxylated Glu residues (positions 16, 20, 19, and 14 of the Gla domain, respectively). The subgroups of DCP proteoforms, represented by each Glu-peptide, are summarized in **Table 2-13**.

Table 2-13. Prospective combinations of 3 Glu-peptides that represent 10 DCP variants.

		Abnormal Prothrombin (DCP proteoforms)										Normal Prothrombin	Subgroup of DCP proteoforms represented by Glu-peptides
The number of residues in the Gla domain	Glu	10	9	8	7	6	5	4	3	2	1	0	
	Gla	0	1	2	3	4	5	6	7	8	9	10	
Component Glu-peptides	ANT	O	O	O	O	O	O	O	O	O			2-10 Glu residues
	ESS	O	O	O	O	O	O	O	O				3-10 Glu residues
	ERE	O	O	O									8-10 Glu residues
Combination of existing Glu-peptides		ANT + ESS + ERE			ANT + ESS					ANT	None of them		

Abbreviations: Gla, γ-carboxylated glutamic acid; Glu, glutamic acid; ANT, ANTFLEEVR peptide; ESS, ESSTATDVF peptide; ERE, ERECVEETCSY peptide. Note: All of the DCP proteoforms have one of the accumulative combinations of three Glu peptides: one (ANTFLEEVR only), two (ANTFLEEVR and ESSTATDVF), and three (All Glu-peptides) peptides. Therefore, the ANTFLEEVR peptide represents the subgroup of DCP proteoforms with more than two Glu residues (2 to 10 Glu residues). Likewise, the ESSTATDVF peptide represents the subgroup of DCP with more than three Glu residues (3 to 10 Glu residues). The ERECVEETCSY peptide represents the subgroup of DCP with more than eight Glu residues (8 to 10 Glu residues).

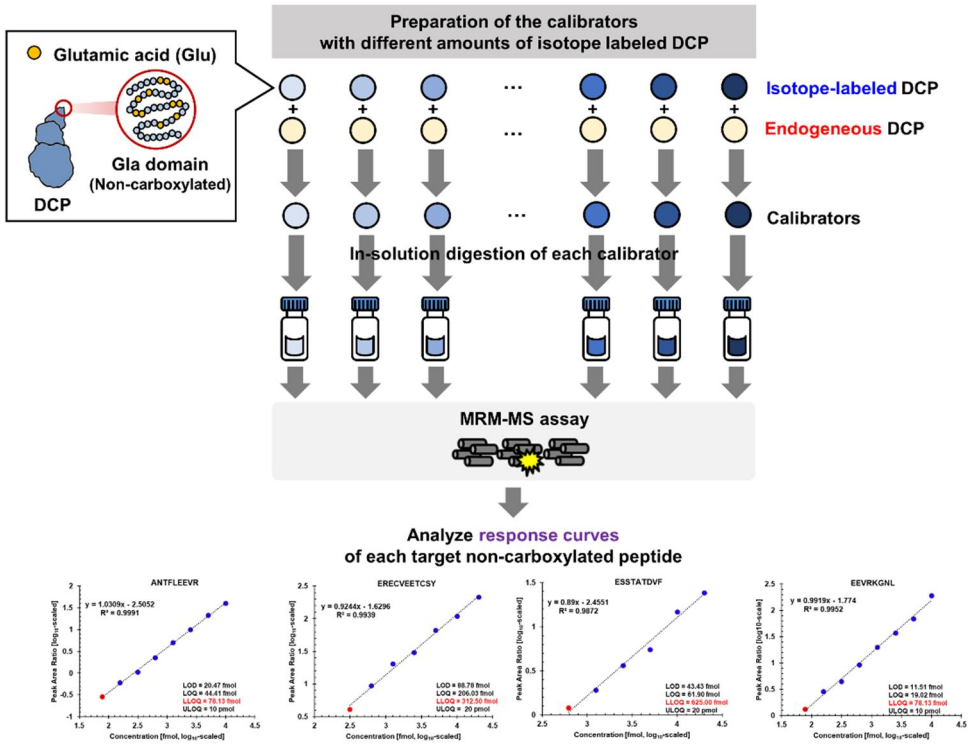


Figure 2-16. Possible experiment scheme for fidelity analysis of protease digestion indirectly to increase the credibility of quantification results from the MRM-MS assay

The fidelity analysis of protease digestion can be determined by response curve analysis with isotope-labeled non-carboxylated DCP proteins (containing 10 Glu residues). Each calibration point is prepared by spiking different amounts of non-carboxylated DCP standard proteins and undergoing digestion separately in triplicates. Then the quantification results of the MRM-MS assay are plotted with expected concentrations for each point to analyze their linearity. Although missed cleavages can occur at the cleavage sites in the Gla domain, it can be used as targets for MRM-MS assay if the quantification results have sufficient linearity and reproducibility. Therefore, the response curves of each peptide obtained from

different concentrations of non-carboxylated DCP can be analyzed to inspect whether the effect of missed cleavage can be deemed negligible.

2.5. CONCLUSION

It is clear that the quantitative MRM-MS assay for DCP measurement that was designed in this study shows equivalent diagnostic performance as the antibody-based DCP immunoassay. The MRM-MS assay to quantify three Glu-peptides enabled not only the extensive detection of DCP proteoforms but also a detailed comparison of the DCP proteoform compositions between HCC and benign liver diseases. This study indicates that the comprehensive measurement of DCP proteoforms using the MRM-MS assay has great potential as a surveillance test for the detection of HCC at the very-early-stage, even among patients with AFP and DCP levels under the corresponding cutoff values. Further, this assay is advantageous compared with the DCP immunoassay because it facilitates the high-throughput analysis of large cohorts with increased diagnostic accuracy, while requiring lower costs and sample volumes. The multiplexing ability of the MS-based quantification approach has the potential to develop an HCC surveillance assay that simultaneously analyzes the DCP proteoforms in combination with hundreds of existing serological biomarkers in a high-throughput format that would be suitable for routine check-ups.

GENERAL CONCLUSION

Clinical proteomics is a promising approach for the discovery or validation of biomarkers through the systematic analysis of proteomes in clinical samples. The advent of the era of artificial intelligence increases the need to achieve datasets of sufficient size in large-scale clinical proteomics. The quantitative MRM-MS assays are the best alternative to enable the acquisition of large-scale proteomic data including quantification results for proteoforms with high analytical sensitivity, accuracy, precision, and reproducibility. However, the central obstacle for reproducible MRM-MS assays in large-scale cohorts is the sample preparation.

The purpose of this research was to make the MRM-MS assays better-suited for large-scale clinical proteomic analysis by: (1) evaluating of feasibility for the automated workflow of sample preparation for high throughput MRM-MS analysis in the large-scale proteomics experiments; (2) establishing the reproducible MRM-MS assay for DCP proteoforms to increase HCC surveillance performances.

In chapter 1, I developed an automated workflow for serum sample preparation with the robotic liquid handling system which is compatible with 96-well plate. Further increment of throughput can be achieved by pipette channels with 384-well format with little modifications. The automated workflow was assessed by preparing the pooled serum samples and quantifying 52 peptides. The results from this study proved that the automated workflow ensures stable serum sample preparation (an average of total CVs of 15.3%). Further study is needed to evaluate the probable variability in the automated preparation of heterogeneous clinical samples for the MRM-MS analysis on the expanded target peptides. The automated preparation

workflow can be adopted by other liquid-handling systems similar to an Agilent Bravo apparatus. This automated preparation workflow is expected to perform well to cope with a vast number of various sample types, such as cyst fluids, tissue or FFPE samples, with some adaptation. Overall, the automated workflow developed in this study is a significant step toward facilitating robust clinical MRM-MS assays in large cohorts with high throughput.

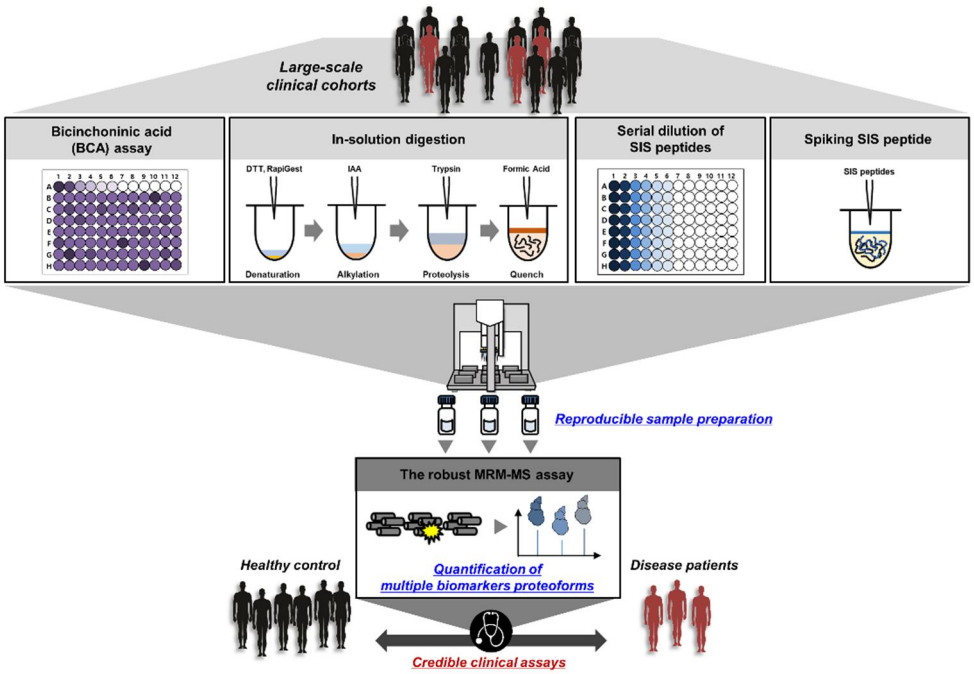
In chapter 2, I developed a reproducible MRM-MS assay to inclusively quantify DCP proteoforms. The assay quantifies the three non-carboxylated peptides obtained from the Gla domain of the precursor of prothrombin. This MS-based quantitative assay for DCP proteoforms enables the reproducible analysis of large cohorts with enhanced diagnostic performances than the immunoassay. The great potential of the MRM-MS assay for DCP proteoforms as HCC surveillance tests has been demonstrated by its diagnostic power for the HCC at the very-early-stage or early-stage. Current findings suggest that the comprehensive measurement of DCP proteoforms can make a detailed comparison of the DCP proteoform compositions between HCC and benign liver diseases with high throughput and reproducibility. Future exploration into the difference in the composition of DCP proteoforms in the well-characterized large-scale clinical cohorts can investigate the role of DCP proteoforms in HCC progress by the herein MRM-MS assay for DCP. Furthermore, the MRM-MS assay for DCP developed in this study has the potential for developing a powerful HCC surveillance assay simultaneously analyzing the DCP proteoforms with hundreds of existing serological biomarkers with high throughput.

Based on the analysis conveyed, it can be concluded that (1) development of the automated workflow for serum sample preparation for MRM-MS assay can

reduce the human labor and errors while increasing the throughput of the clinical assay; (2) development of a single quantitative MRM-MS assay for multiple proteoforms of DCP can improve HCC surveillance performances. Consequently, I expect that the automated sample preparation workflow and the quantitative MRM-MS assay for inclusive measurement of biomarker proteoforms will contribute to routine implementations of MS-based assays in clinical practice.

However, several challenges remain unaddressed; in order to implement the developed automated workflow of sample preparation for MRM-MS assay in clinical practice, the analytical method validation of the system should be conducted in a future study: not only for in-depth inspection of the robustness of the assay, but also for determining the limitation of the quantification results from the automated workflow. It is not feasible to apply the novel workflow into practice without addressing the credibility of quantification results from the system. Meanwhile, the MRM-MS assay of DCP proteoforms requires a more simplified and automated sample preparation workflow to implement the assay in clinical practice.

Despite these limitations, the research on the MRM-MS for large-scale clinical proteomics demonstrated that a reproducible MRM-MS assay has potential to be a powerful technique for high throughput biomarker quantification to increase diagnostic performance. Further research on the application of the MRM-MS assay through the automated sample preparation workflow for quantification of existing or candidate biomarkers in large cohorts will enable credible clinical assays for various diseases, as described below.


Large-scale clinical cohorts
Bicinchoninic acid (BCA) assay
In-solution digestion
DTT, RapiGest
IAA
Trypsin
Formic Acid
Denaturation
Alkylation
Proteolysis
Quench
Serial dilution of SIS peptides
Spiking SIS peptide
SIS peptides
Reproducible sample preparation
The robust MRM-MS assay
Quantification of multiple biomarkers proteoforms
Healthy control
Disease patients
Credible clinical assays

REFERENCES

1. Lee, J.; Kim, H.; Sohn, A.; Yeo, I.; Kim, Y., Cost-Effective Automated Preparation of Serum Samples for Reproducible Quantitative Clinical Proteomics. *J Proteome Res* **2019,** 18, (5), 2337-2345.

2. Lee, J.; Lim, Y. S.; Lee, J. H.; Gwak, G. Y.; Do, M.; Yeo, I.; Shin, D.; Han, D.; Park, T.; Kim, Y., Inclusive Quantification Assay of Serum Des-gamma-Carboxyprothrombin Proteoforms for Hepatocellular Carcinoma Surveillance by Targeted Mass Spectrometry. *Hepatol Commun* **2021,** 5, (10), 1767-1783.

3. Su, M.; Zhang, Z.; Zhou, L.; Han, C.; Huang, C.; Nice, E. C., Proteomics, Personalized Medicine and Cancer. *Cancers (Basel)* **2021,** 13, (11).

4. Amiri-Dashatan, N.; Koushki, M.; Abbaszadeh, H. A.; Rostami-Nejad, M.; Rezaei-Tavirani, M., Proteomics Applications in Health: Biomarker and Drug Discovery and Food Industry. *Iran J Pharm Res* **2018,** 17, (4), 1523-1536.

5. Kuras, M.; Betancourt, L. H.; Rezeli, M.; Rodriguez, J.; Szasz, M.; Zhou, Q.; Miliotis, T.; Andersson, R.; Marko-Varga, G., Assessing Automated Sample Preparation Technologies for High-Throughput Proteomics of Frozen Well Characterized Tissues from Swedish Biobanks. *J Proteome Res* **2019,** 18, (1), 548-556.

6. Switzar, L.; van Angeren, J.; Pinkse, M.; Kool, J.; Niessen, W. M., A high-throughput sample preparation method for cellular proteomics using 96-well filter plates. *Proteomics* **2013,** 13, (20), 2980-3.

7. Burns, A. P.; Zhang, Y. Q.; Xu, T.; Wei, Z.; Yao, Q.; Fang, Y.; Cebotaru, V.; Xia, M.; Hall, M. D.; Huang, R.; Simeonov, A.; LeClair, C. A.; Tao, D., A Universal and High-Throughput Proteomics Sample Preparation Platform. *Anal Chem* **2021,** 93, (24), 8423-8431.

8. Chen, Y.; Guenther, J. M.; Gin, J. W.; Chan, L. J. G.; Costello, Z.; Ogorzalek, T. L.; Tran, H. M.; Blake-Hedges, J. M.; Keasling, J. D.; Adams, P. D.; Garcia Martin, H.; Hillson, N. J.; Petzold, C. J., Automated "Cells-To-Peptides" Sample Preparation Workflow for High-Throughput, Quantitative Proteomic Assays of Microbes. *J Proteome Res* **2019,** 18, (10), 3752-3761.

9. Klatt, J. N.; Dinh, T. J.; Schilling, O.; Zengerle, R.; Schmidt, F.; Hutzenlaub, T.; Paust, N., Automation of peptide desalting for proteomic liquid chromatography - tandem mass spectrometry by centrifugal microfluidics. *Lab Chip* **2021,** 21, (11), 2255-2264.

10. Dayon, L.; Nunez Galindo, A.; Corthesy, J.; Cominetti, O.; Kussmann, M., Comprehensive and Scalable Highly Automated MS-Based Proteomic Workflow for Clinical Biomarker Discovery in Human Plasma. *J Proteome Res* **2014,** 13, (8), 3837-3845.

11. Liang, Y.; Acor, H.; McCown, M. A.; Nwosu, A. J.; Boekweg, H.; Axtell, N. B.; Truong, T.; Cong, Y.; Payne, S. H.; Kelly, R. T., Fully Automated Sample Processing and Analysis Workflow for Low-Input Proteome Profiling. *Anal Chem* **2021,** 93, (3), 1658-1666.

12. Muller, T.; Kalxdorf, M.; Longuespee, R.; Kazdal, D. N.; Stenzinger, A.; Krijgsveld, J., Automated sample preparation with SP3 for low-input clinical proteomics. *Mol Syst Biol* **2020,** 16, (1), e9111.

13. Loroch, S.; Kopczynski, D.; Schneider, A. C.; Schumbrutzki, C.; Feldmann, I.; Panagiotidis, E.; Reinders, Y.; Sakson, R.; Solari, F. A.; Vening, A.; Swieringa, F.; Heemskerk, J. W. M.; Grandoch, M.; Dandekar, T.; Sickmann, A., Toward Zero Variance in Proteomics Sample Preparation: Positive-Pressure FASP in 96-Well Format (PF96) Enables Highly Reproducible, Time- and Cost-Efficient Analysis of Sample Cohorts. *J Proteome Res* **2022,** 21, (4), 1181-1188.

14. Ruelcke, J. E.; Loo, D.; Hill, M. M., Reducing the cost of semi-automated in-gel tryptic digestion and GeLC sample preparation for high-throughput proteomics. *J Proteomics* **2016,** 149, 3-6.

15. Alexovic, M.; Urban, P. L.; Tabani, H.; Sabo, J., Recent advances in robotic protein sample preparation for clinical analysis and other biomedical applications. *Clin Chim Acta* **2020,** 507, 104-116.

16. Schaffer, L. V.; Millikin, R. J.; Miller, R. M.; Anderson, L. C.; Fellers, R. T.; Ge, Y.; Kelleher, N. L.; LeDuc, R. D.; Liu, X.; Payne, S. H.; Sun, L.; Thomas, P. M.; Tucholski, T.; Wang, Z.; Wu, S.; Wu, Z.; Yu, D.; Shortreed, M. R.; Smith, L. M., Identification and Quantification of Proteoforms by Mass Spectrometry. *Proteomics* **2019,** 19, (10), e1800361.

17. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **2008,** 4, 222.

18. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **2012,** 9, (6), 555-66.

19. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **2005,** 1, (5), 252-62.

20. Shi, T.; Fillmore, T. L.; Sun, X.; Zhao, R.; Schepmoes, A. A.; Hossain, M.; Xie, F.; Wu, S.; Kim, J. S.; Jones, N.; Moore, R. J.; Pasa-Tolic, L.; Kagan, J.; Rodland, K. D.; Liu, T.; Tang, K.; Camp, D. G., 2nd; Smith, R. D.; Qian, W. J., Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc Natl Acad Sci U S A* **2012,** 109, (38), 15395-400.

21. Zhang, F.; Bartels, M. J.; Stott, W. T., Quantitation of human glutathione S-transferases in complex matrices by liquid chromatography/tandem mass spectrometry with

signature peptides. *Rapid Commun Mass Spectrom* **2004,** 18, (4), 491-8.

22. Yocum, A. K.; Gratsch, T. E.; Leff, N.; Strahler, J. R.; Hunter, C. L.; Walker, A. K.; Michailidis, G.; Omenn, G. S.; O'Shea, K. S.; Andrews, P. C., Coupled global and targeted proteomics of human embryonic stem cells during induced differentiation. *Mol Cell Proteomics* **2008,** 7, (4), 750-67.

23. Anderson, L.; Hunter, C. L., Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* **2006,** 5, (4), 573-88.

24. Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A., Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **2007,** 6, (12), 2212-29.

25. Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A. J.; Keshishian, H.; Hall, S. C.; Allen, S.; Blackman, R. K.; Borchers, C. H.; Buck, C.; Cardasis, H. L.; Cusack, M. P.; Dodder, N. G.; Gibson, B. W.; Held, J. M.; Hiltke, T.; Jackson, A.; Johansen, E. B.; Kinsinger, C. R.; Li, J.; Mesri, M.; Neubert, T. A.; Niles, R. K.; Pulsipher, T. C.; Ransohoff, D.; Rodriguez, H.; Rudnick, P. A.; Smith, D.; Tabb, D. L.; Tegeler, T. J.; Variyath, A. M.; Vega-Montoto, L. J.; Wahlander, A.; Waldemarson, S.; Wang, M.; Whiteaker, J. R.; Zhao, L.; Anderson, N. L.; Fisher, S. J.; Liebler, D. C.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **2009,** 27, (7), 633-41.

26. Shi, T.; Su, D.; Liu, T.; Tang, K.; Camp, D. G., 2nd; Qian, W. J.; Smith, R. D., Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics* **2012,** 12, (8), 1074-92.

27. Prakash, A.; Rezai, T.; Krastins, B.; Sarracino, D.; Athanas, M.; Russo, P.; Ross, M. M.; Zhang, H.; Tian, Y.; Kulasingam, V.; Drabovich, A. P.; Smith, C.; Batruch, I.; Liotta, L.; Petricoin, E.; Diamandis, E. P.; Chan, D. W.; Lopez, M. F., Platform for establishing interlaboratory reproducibility of selected reaction monitoring-based mass spectrometry peptide assays. *J Proteome Res* **2010,** 9, (12), 6678-88.

28. Kim, H.; Yu, S. J.; Yeo, I.; Cho, Y. Y.; Lee, D. H.; Cho, Y.; Cho, E. J.; Lee, J. H.; Kim, Y. J.; Lee, S.; Jun, J.; Park, T.; Yoon, J. H.; Kim, Y., Prediction of Response to Sorafenib in Hepatocellular Carcinoma: A Putative Marker Panel by Multiple Reaction Monitoring-Mass Spectrometry (MRM-MS). *Mol Cell Proteomics* **2017,** 16, (7), 1312-1323.

29. Kim, H.; Park, J.; Kim, Y.; Sohn, A.; Yeo, I.; Jong Yu, S.; Yoon, J. H.; Park, T.; Kim, Y., Serum fibronectin distinguishes the early stages of hepatocellular carcinoma. *Sci Rep* **2017,** 7, (1), 9449.

30. Yu, S. J.; Kim, H.; Min, H.; Sohn, A.; Cho, Y. Y.; Yoo, J. J.; Lee, D. H.; Cho, E. J.; Lee, J. H.; Gim, J.; Park, T.; Kim, Y. J.; Kim, C. Y.; Yoon, J. H.; Kim, Y., Targeted Proteomics Predicts a Sustained Complete-Response after Transarterial Chemoembolization and Clinical Outcomes in Patients with Hepatocellular Carcinoma: A Prospective Cohort Study. *J Proteome Res* **2017,** 16, (3), 1239-1248.

31. Whiteaker, J. R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R. M.; Zhao, L.; Voytovich, U. J.; Kelly-Spratt, K. S.; Krasnoselsky, A.; Gafken, P. R.; Hogan, J. M.; Jones, L. A.; Wang, P.; Amon, L.; Chodosh, L. A.; Nelson, P. S.; McIntosh, M. W.; Kemp, C. J.; Paulovich, A. G., A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol* **2011,** 29, (7), 625-34.

32. Bros, P.; Vialaret, J.; Barthelemy, N.; Delatour, V.; Gabelle, A.; Lehmann, S.; Hirtz, C., Antibody-free quantification of seven tau peptides in human CSF using targeted mass spectrometry. *Front Neurosci* **2015,** 9, 302.

33. Percy, A. J.; Chambers, A. G.; Yang, J.; Hardie, D. B.; Borchers, C. H., Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. *Biochim Biophys Acta* **2014,** 1844, (5), 917-26.

34. Domanski, D.; Percy, A. J.; Yang, J.; Chambers, A. G.; Hill, J. S.; Freue, G. V.; Borchers, C. H., MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics* **2012,** 12, (8), 1222-43.

35. Kim, H.; Sohn, A.; Yeo, I.; Yu, S. J.; Yoon, J. H.; Kim, Y., Clinical Assay for AFP-L3 by Using Multiple Reaction Monitoring-Mass Spectrometry for Diagnosing Hepatocellular Carcinoma. *Clin Chem* **2018,** 64, (8), 1230-1238.

36. Sohn, A.; Kim, H.; Yeo, I.; Kim, Y.; Son, M.; Yu, S. J.; Yoon, J. H.; Kim, Y., 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, 142-146.

37. Lowenthal, M. S.; Liang, Y.; Phinney, K. W.; Stein, S. E., Quantitative bottom-up proteomics depends on digestion conditions. *Anal Chem* **2014,** 86, (1), 551-8.

38. Ma, J.; Shi, J.; Le, H.; Cho, R.; Huang, J. C.; Miao, S.; Wong, B. K., A fully automated plasma protein precipitation sample preparation method for LC-MS/MS bioanalysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008,** 862, (1-2), 219-26.

39. Jeon, J.; Yang, J.; Park, J. M.; Han, N. Y.; Lee, Y. B.; Lee, H., Development of an automated high-throughput sample preparation protocol for LC-MS/MS analysis of glycated peptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018,** 1092, 88-94.

40. Chelius, D.; Xiao, G.; Nichols, A. C.; Vizel, A.; He, B.; Dillon, T. M.; Rehder, D. S.; Pipes, G. D.; Kraft, E.; Oroska, A.; Treuheit, M. J.; Bondarenko, P. V., Automated tryptic

digestion procedure for HPLC/MS/MS peptide mapping of immunoglobulin gamma antibodies in pharmaceutics. *J Pharm Biomed Anal* **2008,** 47, (2), 285-94.

41. Zhu, M.; Zhang, P.; Geng-Spyropoulos, M.; Moaddel, R.; Semba, R. D.; Ferrucci, L., A robotic protocol for high-throughput processing of samples for selected reaction monitoring assays. *Proteomics* **2017,** 17, (6).

42. Fu, Q.; Kowalski, M. P.; Mastali, M.; Parker, S. J.; Sobhani, K.; van den Broek, I.; Hunter, C. L.; Van Eyk, J. E., Highly Reproducible Automated Proteomics Sample Preparation Workflow for Quantitative Mass Spectrometry. *J Proteome Res* **2017,** 17, (1), 420-428.

43. Razavi, M.; Leigh Anderson, N.; Pope, M. E.; Yip, R.; Pearson, T. W., High precision quantification of human plasma proteins using the automated SISCAPA Immuno-MS workflow. *N Biotechnol* **2016,** 33, (5 Pt A), 494-502.

44. Zhang, Y. V.; Rockwood, A., Impact of automation on mass spectrometry. *Clin Chim Acta* **2015,** 450, 298-303.

45. Cominetti, O.; Nunez Galindo, A.; Corthesy, J.; Oller Moreno, S.; Irincheeva, I.; Valsesia, A.; Astrup, A.; Saris, W. H.; Hager, J.; Kussmann, M.; Dayon, L., Proteomic Biomarker Discovery in 1000 Human Plasma Samples with Mass Spectrometry. *J Proteome Res* **2016,** 15, (2), 389-99.

46. Popp, R.; Malmstrom, D.; Chambers, A. G.; Lin, D.; Camenzind, A. G.; van der Gugten, J. G.; Holmes, D. T.; Pugia, M.; Jaremek, M.; Cornett, S.; Suckau, D.; Borchers, C. H., An automated assay for the clinical measurement of plasma renin activity by immuno-MALDI (iMALDI). *Biochim Biophys Acta* **2015,** 1854, (6), 547-58.

47. Choi, E.; Loo, D.; Dennis, J. W.; O'Leary, C. A.; Hill, M. M., High-throughput lectin magnetic bead array-coupled tandem mass spectrometry for glycoprotein biomarker discovery. *Electrophoresis* **2011,** 32, (24), 3564-75.

48. Woo, J.; Williams, S. M.; Markillie, L. M.; Feng, S.; Tsai, C. F.; Aguilera-Vazquez, V.; Sontag, R. L.; Moore, R. J.; Hu, D.; Mehta, H. S.; Cantlon-Bruce, J.; Liu, T.; Adkins, J. N.; Smith, R. D.; Clair, G. C.; Pasa-Tolic, L.; Zhu, Y., Author Correction: High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nat Commun* **2021,** 12, (1), 7075.

49. Alexovic, M.; Sabo, J.; Longuespee, R., Automation of single-cell proteomic sample preparation. *Proteomics* **2021,** 21, (23-24).

50. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S., UniProt: the Universal Protein knowledgebase. *Nucleic Acids*

Res **2004,** 32, (Database issue), D115-9.

51. Kusebauch, U.; Campbell, D. S.; Deutsch, E. W.; Chu, C. S.; Spicer, D. A.; Brusniak, M. Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; Shteynberg, D.; Hoopmann, M. R.; Blattmann, P.; Ratushny, A. V.; Rinner, O.; Picotti, P.; Carapito, C.; Huang, C. Y.; Kapousouz, M.; Lam, H.; Tran, T.; Demir, E.; Aitchison, J. D.; Sander, C.; Hood, L.; Aebersold, R.; Moritz, R. L., Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* **2016,** 166, (3), 766-778.

52. Whiteaker, J. R.; Halusa, G. N.; Hoofnagle, A. N.; Sharma, V.; MacLean, B.; Yan, P.; Wrobel, J. A.; Kennedy, J.; Mani, D. R.; Zimmerman, L. J.; Meyer, M. R.; Mesri, M.; Rodriguez, H.; Clinical Proteomic Tumor Analysis, C.; Paulovich, A. G., CPTAC Assay Portal: a repository of targeted proteomic assays. *Nat Methods* **2014,** 11, (7), 703-4.

53. Abbatiello, S. E.; Mani, D. R.; Keshishian, H.; Carr, S. A., Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin Chem* **2010,** 56, (2), 291-305.

54. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** 26, (7), 966-8.

55. Hrncir, E.; Rosina, J., Surface tension of blood. *Physiol Res* **1997,** 46, (4), 319-21.

56. Lewin, S., Blood serum surface tension and its potential. *Br J Haematol* **1972,** 22, (5), 561-6.

57. Villanueva, A., Hepatocellular Carcinoma. *New England Journal of Medicine* **2019,** 380, (15), 1450-1462.

58. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **2018,** 68, (6), 394-424.

59. Forner, A.; Reig, M.; Bruix, J., Hepatocellular carcinoma. *Lancet* **2018,** 391, (10127), 1301-1314.

60. Petrick, J. L.; Florio, A. A.; Znaor, A.; Ruggieri, D.; Laversanne, M.; Alvarez, C. S.; Ferlay, J.; Valery, P. C.; Bray, F.; McGlynn, K. A., International trends in hepatocellular carcinoma incidence, 1978-2012. *Int J Cancer* **2020,** 147, (2), 317-330.

61. McGlynn, K. A.; Petrick, J. L.; El-Serag, H. B., Epidemiology of Hepatocellular Carcinoma. *Hepatology* **2020**.

62. Sartorius, K.; Sartorius, B.; Aldous, C.; Govender, P. S.; Madiba, T. E., Global and country underestimation of hepatocellular carcinoma (HCC) in 2012 and its implications.

Cancer Epidemiol **2015,** 39, (3), 284-90.

63. de Martel, C.; Maucort-Boulch, D.; Plummer, M.; Franceschi, S., World-wide relative contribution of hepatitis B and C viruses in hepatocellular carcinoma. *Hepatology* **2015,** 62, (4), 1190-200.

64. Yin, Z.; Dong, C.; Jiang, K.; Xu, Z.; Li, R.; Guo, K.; Shao, S.; Wang, L., Heterogeneity of cancer-associated fibroblasts and roles in the progression, prognosis, and therapy of hepatocellular carcinoma. *J Hematol Oncol* **2019,** 12, (1), 101.

65. Erstad, D. J.; Tanabe, K. K., Prognostic and Therapeutic Implications of Microvascular Invasion in Hepatocellular Carcinoma. *Ann Surg Oncol* **2019,** 26, (5), 1474-1493.

66. Sarveazad, A.; Agah, S.; Babahajian, A.; Amini, N.; Bahardoust, M., Predictors of 5 year survival rate in hepatocellular carcinoma patients. *J Res Med Sci* **2019,** 24, 86.

67. Njei, B.; Rotman, Y.; Ditah, I.; Lim, J. K., Emerging trends in hepatocellular carcinoma incidence and mortality. *Hepatology* **2015,** 61, (1), 191-9.

68. Singal, A. G.; Pillai, A.; Tiro, J., Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. *PLoS Med* **2014,** 11, (4), e1001624.

69. Sangiovanni, A.; Del Ninno, E.; Fasani, P.; De Fazio, C.; Ronchi, G.; Romeo, R.; Morabito, A.; De Franchis, R.; Colombo, M., Increased survival of cirrhotic patients with a hepatocellular carcinoma detected during surveillance. *Gastroenterology* **2004,** 126, (4), 1005-14.

70. Ryder, S. D.; British Society of, G., Guidelines for the diagnosis and treatment of hepatocellular carcinoma (HCC) in adults. *Gut* **2003,** 52 Suppl 3, iii1-8.

71. Kansagara, D.; Papak, J.; Pasha, A. S.; O'Neil, M.; Freeman, M.; Relevo, R.; Quinones, A.; Motu'apuaka, M.; Jou, J. H., Screening for hepatocellular carcinoma in chronic liver disease: a systematic review. *Ann Intern Med* **2014,** 161, (4), 261-9.

72. Tzartzeva, K.; Obi, J.; Rich, N. E.; Parikh, N. D.; Marrero, J. A.; Yopp, A.; Waljee, A. K.; Singal, A. G., Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta-analysis. *Gastroenterology* **2018,** 154, (6), 1706-1718 e1.

73. European Association For The Study Of The, L.; European Organisation For, R.; Treatment Of, C., EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* **2012,** 56, (4), 908-43.

74. Di Bisceglie, A. M.; Hoofnagle, J. H., Elevations in serum alpha-fetoprotein levels in patients with chronic hepatitis B. *Cancer* **1989,** 64, (10), 2117-20.

75. Di Bisceglie, A. M.; Sterling, R. K.; Chung, R. T.; Everhart, J. E.; Dienstag, J. L.; Bonkovsky, H. L.; Wright, E. C.; Everson, G. T.; Lindsay, K. L.; Lok, A. S.; Lee, W. M.; Morgan, T. R.; Ghany, M. G.; Gretch, D. R.; Group, H.-C. T., Serum alpha-fetoprotein levels in patients with advanced hepatitis C: results from the HALT-C Trial. *J Hepatol* **2005,** 43, (3), 434-41.

76. Marrero, J. A.; Feng, Z.; Wang, Y.; Nguyen, M. H.; Befeler, A. S.; Roberts, L. R.; Reddy, K. R.; Harnois, D.; Llovet, J. M.; Normolle, D.; Dalhgren, J.; Chia, D.; Lok, A. S.; Wagner, P. D.; Srivastava, S.; Schwartz, M., Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* **2009,** 137, (1), 110-8.

77. Lok, A. S.; Sterling, R. K.; Everhart, J. E.; Wright, E. C.; Hoefs, J. C.; Di Bisceglie, A. M.; Morgan, T. R.; Kim, H. Y.; Lee, W. M.; Bonkovsky, H. L.; Dienstag, J. L.; Group, H.-C. T., Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology* **2010,** 138, (2), 493-502.

78. Wang, T.; Zhang, K. H., New Blood Biomarkers for the Diagnosis of AFP-Negative Hepatocellular Carcinoma. *Front Oncol* **2020,** 10, 1316.

79. Liebman, H. A.; Furie, B. C.; Tong, M. J.; Blanchard, R. A.; Lo, K. J.; Lee, S. D.; Coleman, M. S.; Furie, B., Des-gamma-carboxy (abnormal) prothrombin as a serum marker of primary hepatocellular carcinoma. *N Engl J Med* **1984,** 310, (22), 1427-31.

80. Weitz, I. C.; Liebman, H. A., Des-gamma-carboxy (abnormal) prothrombin and hepatocellular carcinoma: a critical review. *Hepatology* **1993,** 18, (4), 990-7.

81. Tsai, S. L.; Huang, G. T.; Yang, P. M.; Sheu, J. C.; Sung, J. L.; Chen, D. S., Plasma des-gamma-carboxyprothrombin in the early stage of hepatocellular carcinoma. *Hepatology* **1990,** 11, (3), 481-8.

82. Borowski, M.; Furie, B. C.; Furie, B., Distribution of gamma-carboxyglutamic acid residues in partially carboxylated human prothrombins. *J Biol Chem* **1986,** 261, (4), 1624-8.

83. Uehara, S.; Gotoh, K.; Handa, H.; Honjo, K.; Hirayama, A., Process of carboxylation of glutamic acid residues in the gla domain of human des-gamma-carboxyprothrombin. *Clin Chim Acta* **1999,** 289, (1-2), 33-44.

84. Liska, D. J.; Suttie, J. W., Location of gamma-carboxyglutamyl residues in partially carboxylated prothrombin preparations. *Biochemistry* **1988,** 27, (23), 8636-41.

85. Huisse, M. G.; Leclercq, M.; Belghiti, J.; Flejou, J. F.; Suttie, J. W.; Bezeaud, A.; Stafford, D. W.; Guillin, M. C., Mechanism of the abnormal vitamin K-dependent gamma-carboxylation process in human hepatocellular carcinomas. *Cancer* **1994,** 74, (5), 1533-41.

86. Friedman, P. A.; Przysiecki, C. T., Vitamin K-dependent carboxylation. *Int J*

Biochem **1987,** 19, (1), 1-7.

87. Uehara, S.; Gotoh, K.; Handa, H.; Tomita, H.; Senshuu, M., Distribution of the heterogeneity of des-gamma-carboxyprothrombin in patients with hepatocellular carcinoma. *J Gastroenterol Hepatol* **2005,** 20, (10), 1545-52.

88. Naraki, T.; Kohno, N.; Saito, H.; Fujimoto, Y.; Ohhira, M.; Morita, T.; Kohgo, Y., gamma-Carboxyglutamic acid content of hepatocellular carcinoma-associated des-gamma-carboxy prothrombin. *Biochim Biophys Acta* **2002,** 1586, (3), 287-98.

89. Sugo, T.; Watanabe, K.; Naraki, T.; Matsuda, M., Chemical modification of gamma-carboxyglutamic acid residues in prothrombin elicits a conformation similar to that of abnormal (des-gamma-carboxy)prothrombin. *J Biochem* **1990,** 108, (3), 382-7.

90. Tameda, M.; Shiraki, K.; Sugimoto, K.; Ogura, S.; Inagaki, Y.; Yamamoto, N.; Ikejiri, M.; Takei, Y.; Ito, M.; Nobori, T., Des-gamma-carboxy prothrombin ratio measured by P-11 and P-16 antibodies is a novel biomarker for hepatocellular carcinoma. *Cancer Sci* **2013,** 104, (6), 725-31.

91. Nagaoka, S.; Yatsuhashi, H.; Hamada, H.; Yano, K.; Matsumoto, T.; Daikoku, M.; Arisawa, K.; Ishibashi, H.; Koga, M.; Sata, M.; Yano, M., The des-gamma-carboxy prothrombin index is a new prognostic indicator for hepatocellular carcinoma. *Cancer* **2003,** 98, (12), 2671-7.

92. Sohn, A.; Kim, H.; Yu, S. J.; Yoon, J. H.; Kim, Y., A quantitative analytical method for PIVKA-II using multiple reaction monitoring-mass spectrometry for early diagnosis of hepatocellular carcinoma. *Anal Bioanal Chem* **2017,** 409, (11), 2829-2838.

93. Heimbach, J. K.; Kulik, L. M.; Finn, R. S.; Sirlin, C. B.; Abecassis, M. M.; Roberts, L. R.; Zhu, A. X.; Murad, M. H.; Marrero, J. A., AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* **2018,** 67, (1), 358-380.

94. Kratz, A.; Ferraro, M.; Sluss, P. M.; Lewandrowski, K. B., Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Laboratory reference values. *N Engl J Med* **2004,** 351, (15), 1548-63.

95. Basit, A.; Prasad, B.; Estergreen, J. K.; Sabath, D. E.; Alade, N.; Veenstra, D. L.; Rettie, A. E.; Thummel, K. E., A Novel LC-MS/MS Assay for Quantification of Des-carboxy Prothrombin and Characterization of Warfarin-Induced Changes. *Clin Transl Sci* **2020**.

96. Kuhn, M., Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **2008,** 28, (5), 1-26.

97. Malhotra, O. P.; Nesheim, M. E.; Mann, K. G., The kinetics of activation of normal and gamma-carboxyglutamic acid-deficient prothrombins. *J Biol Chem* **1985,** 260, (1), 279-87.

98. Ratcliffe, J. V.; Furie, B.; Furie, B. C., The importance of specific gamma-carboxyglutamic acid residues in prothrombin. Evaluation by site-specific mutagenesis. *J Biol Chem* **1993,** 268, (32), 24339-45.

99. Slechtova, T.; Gilar, M.; Kalikova, K.; Tesarova, E., Insight into Trypsin Miscleavage: Comparison of Kinetic Constants of Problematic Peptide Sequences. *Analytical Chemistry* **2015,** 87, (15), 7636-7643.

100. Brownridge, P.; Beynon, R. J., The importance of the digest: Proteolysis and absolute quantification in proteomics. *Methods* **2011,** 54, (4), 351-360.

101. Gershon, P. D., Cleaved and Missed Sites for Trypsin, Lys-C, and Lys-N Can Be Predicted with High Confidence on the Basis of Sequence Context. *Journal of Proteome Research* **2014,** 13, (2), 702-709.

102. Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P. R., Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Communications in Mass Spectrometry* **2000,** 14, (6), 496-502.

103. Siepen, J. A.; Keevil, E. J.; Knight, D.; Hubbard, S. J., Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *Journal of Proteome Research* **2007,** 6, (1), 399-408.

ABSTRACT IN KOREAN

국문 초록

서론: 질량분석법 기반 단백체학은 수천개의 단백질을 동시에 탐지 및 정량 할 수 있어, 대량시료를 높은 처리량으로 분석할 수 있도록 한다. 높은 민감도를 가진 다중화 분석법의 개발이 가능한 다중 반응 검지 질량 분석법을 이용하여 대량 임상시료의 분석하기 위해, 시료 전처리 과정의 재현성을 확보하는 것은 주요한 해결 과제로 남아 있다. 자동 용액 분주 플랫폼 (automated liquid-handling platforms)이 시료 전처리 과정의 재현성 문제를 해결하는 데 있어서 매우 큰 잠재력을 가지고 있지만, 해당 플랫폼에서 사용되는 소모품의 높은 비용으로 인한 전체 자동화 운영 비용의 증가는 전처리 과정에서의 자동화 시스템의 일상적인 사용을 어렵게 한다. 한편, 단백체학에서 질량분석법 기반 접근법은 항체 기반 분석법에 비해, 단백질의 이성질체 (isoforms) 또는 번역 후 수식 (PTM, posttranslational modifications) 등을 포함하는 다양한 단백질형 (proteoforms)을 구별하여 동시에 정량 할 수 있다는 면에서, 상당한 장점을 가진다. 번역 후 수식을 가지는 단백질 바이오마커의 전형적 예시는 간세포암 감시 진단을 위한 혈청학적 지표인 des-γ-carboxyprothrombin (DCP)이다. DCP 는 N-말단에 존재하는 글루탐산 잔기 (Glu)에 일어나는 손상된 카르복실화 과정으로 만들어진 다양한 단백질형의 혼합된 형태로 혈액 내에 존재한다. 보편적으로, DCP 수준은 면역 분석법을 통해 결정되는데 이러한 면역 분석법은 DCP 분자에 존재하는 항원 결정기와 항체

사이의 결합력에 의존하기 때문에, DCP 의 이질성이 정량의 정확도에 영향을 미친다.

방법: 1 장에서, 자동화된 시료 전처리 과정에 의해 준비된 혈청 샘플에서 52 개의 펩타이드에 대한 MRM-MS 분석 정량 결과의 재현성을 평가했다. 이와 더불어, 액체 처리 자동화 플랫폼에서 비용을 절감한 전처리 자동화 과정의 가능성을 체계적으로 평가했다. 2 장에서, 질량분석기 기반의 DCP 정량법을 활용하여, 다양한 Gla domain 상태를 가지는 DCP 단백질형들을 포괄적으로 탐지하여 동시 정량하고자 하였다. 이를 위해, 다양한 DCP 단백질형에서 유래될 수 있는 4 개의 비카르복실화(non-carboxylation) 펩타이드를 동시 정량하는 다중 반응 모니터링 (MRM-MS) 정량법을 구축하였다. 본 MRM-MS 분석법을 이용하여, 300 명의 간세포암 또는 간세포암 고위험군 (간염 및 간경화) 환자로 구성된 코호트에서 얻은 혈청 시료에서 4 개의 비카르복실화 펩타이드를 동시 정량하였으며, 정량 결과를 머신러닝 기법인 로지스틱 회귀 방법으로 분석하였다. 결과적으로, DCP 유래 3 개의 비카르복실화 펩타이드의 정량 결과를 이용해 간세포암 고위험군으로부터 간세포암 환자를 구분할 수 있는 견고한 감시 진단 모델을 구축하였다.

결과: 1 장에서, 나는 자동화된 시료 전처리 과정이 총 CV 의 평균값(15.3%)으로 증명된 바와 같이 안정적인 혈청 샘플 준비를 보장한다는 것을 입증했다. 또한, 표준 절차와 비교했을 때 비용 최적화된 방법으로 자동화된 시료 전처리 과정을 이용해 총 실험 비용의 37%를 절약하는 동시에 거의 동등한 재현성을 유지할 수 있는 가능성을 확인했다. 2 장에서, 향상된 질량분석기 기반의 DCP 정량

법은, 훈련 세트와 평가 세트에서의 수신자 조작 특성 곡선의 아래 면적 (AUROC, area under the receiver operating characteristic curve) 값이 각각 0.874 와 0.844 로 확인 되어, 매우 신뢰할 만한 감시 진단 성능을 가지는 것을 확인하였다. 이는 기존 항체 기반의 전통적인 DCP 정량 법이 훈련 세트 및 평가 세트에서 각각 0.743 와 0.704 의 AUROC 값을 가지는 것과 비교 동등한 성과였다. 그뿐만 아니라, DCP 정량 기반 간세포암 감시 진단 모델의 성능을 318 례로 구성된 외부 독립 검증 세트에서 검증하였을 때, 그 AUROC 값이 0.793 정도로 확인되었으며, 이는 평가 세트에서의 성능과 비교 동등한 수준이었다.

결론: 1 장에서, 이러한 비용 효율적인 자동 용액 분주 플랫폼을 일상적으로 운용함으로써, 단백체 정량 분석을 위한 대규모 시료의 전처리 과정을 고속 대량으로 수행하면서도, 실험자로부터 기인하는 오류를 줄여 분석 결과의 신뢰성을 높일 수 있을 것으로 기대한다. 2 장에서, 연구를 통해 구축한 MRM-MS 기반의 포괄적 DCP 단백질형 정량 법은 높은 재현성을 가진 비용 효율적인 방법으로, 기존 항체 기반의 DCP 정량 법과 비교하여 더 높은 성능으로 간세포암 감시 진단을 수행할 수 있다는 점에서 더욱 우수하다.

주요어: 임상 단백체학; 질량분석법; 다중반응검지법; 바이오마커; 자동화; 분석법 개발

학 번: 2017-24628

* 본 내용은 학술지에 게재된 두 논문, Journal of Proteome Research (2019, 18:5, 2337-2345), 그리고 Hepatology Communications (2021, 5:10, 1767-1783)을 바탕으로 작성하였음.