이학석사학위논문

Multiple Biomarker Panels to Predict Response to Tocilizumab (anti-IL6R) in Rheumatoid Arthritis Patients Using High-precision Proteomics Approach

고정밀 단백체 분석 기술을 통한 류마티스 관절염 환자의 Tocilizumab 반응 예측 다중 바이오마커 패널 구축에 대한 연구

2019 년 7 월

서울대학교 대학원

분자의학 및 바이오제약학과 전공

정진우

ABSTRACT

Multiple Biomarker Panels to Predict Response to Tocilizumab (anti-IL6R) in Rheumatoid Arthritis Patients Using High-precision Proteomics Approach.

Jinwoo Jung
Department of Molecular Medicine and Biopharmaceutical Sciences
Graduate school of Convergence Science and Technology
Seoul National University

Rheumatoid arthritis (RA) is a chronical systemic autoimmune disease that cause inflammation in synovial lining layer of the joints. Interleukin-6 (IL-6), along with TNF-α and several inflammatory cytokines, acts a vital role in activation of local synovial leukocyte and induction of chronic inflammation. A humanized anti-IL-6 receptor (IL-6R) monoclonal antibody, Tocilizumab (TCZ), has been demonstrated a significant clinical efficacy for RA patients. However, like other inflammatory cytokine blockers such as TNF-α, Interleukin-1 (IL-1), or CD20 inhibitors, some patients are still partially responsive or resistant to the treatment. This study therefore aimed at identifying protein biomarkers that could predict clinical response against TCZ in RA patients by implementing high-precision

proteomics approach. We first identified 54 serum protein biomarker candidates from a large-scale serum proteome profiling of TCZ responder and non-responder groups. Selected protein biomarker candidates combined with known RA biomarkers from the literature data mining were verified by two different targeted quantification methods; multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM) with Triple-quadrupole (QqQ) and Q-Exactive (QE), respectively. Moreover, we validated the results with 47 individual serum samples using MRM and developed as a multi-biomarker panel. The constructed 4-biomarker panel, consisted of APOB, CRP, SERPINA3, and C4A showed 83% discriminate power in average between response and non-response groups with high AUC value of 0.859. The panel also shows 82% sensitivity and 84% specificity. Collectively, our multi-biomarker panel implies that 4 selected proteins were able to serve as diagnostic assessments to predict the TCZ non-responders in RA patients and possible to supplement serum biomarker discovery-validation process in the clinical field based on integrative proteomic approach.

Keywords: Rheumatoid arthritis, Tocilizumab, Interleukin-6, Proteomics, Biomarker, Tandem mass spectrometry, Multiple reaction monitoring, Parallel reaction monitoring

Student Number: 2017-24859

ABBREVIATIONS

The abbreviations are used:

QqQ	Triple-quadruple mass spectrometry
LC-MS/MS	Liquid chromatography and tandem mass spectrometry
Q-Exactive	Quadrupole-orbitrap mass spectrometer
MRM	Multiple reaction monitoring
PRM	Parallel reaction monitoring
SIS	Stable-isotope labeled internal standard
RA	Rheumatoid arthritis
IL-6	Interleukin-6
IL-1	Interleukin-1
TNF-α	Tumor necrosis factor-alpha
UPLC	Ultra performance liquid chromatography
RF	Rheumatoid factor
Anti-CCP	Anti-cyclic citrullinated peptide
IAA	Iodoacetamide
NCE	Normalized collision energy
HCD	High collisional dissociation
CRP	C-reactive protein
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
DAS-28	Disease Activity Score
ACR	American College of Rheumatology
AuDIT	Automated detection of inaccurate transitions
ROC	Receiver operator characteristic
STN	Signal-to-noise ratio

ESI	Electron spray ionization
FDR	False discovery rate

Contents

List of Tables

List of Figures

1. Introduction

Rheumatoid arthritis (RA) is a systemic autoimmune disease that occurs on approximately 1% of the population worldwide [1]. The disease is characterized by joint destruction from synovitis, systemic inflammation, and autoantibodies such as rheumatoid factor (RF) and anti-cyclic citrullinated peptide (anti-CCP) [2]. IL-6, produced from macrophage, T cells, B cells, osteoblasts, and some tumor cells, is one of the dominant pro-inflammatory cytokines which acts a critical role in systemic inflammatory disease pathogenesis [3]. IL-6 stimulates osteoclast differentiation and synthesis of acute phase proteins by binding to membrane bound IL-6 receptor (IL-6R) which allows to form functional homogenous complex with glycoprotein 130 kDa (gp130) and transmits downstream signaling by the Janus kinase/signal transducers and activators of transcription (JAK/STAT) pathway [4]. Current therapeutic agents for RA are disease-modifying anti-rheumatic drugs (DMARDs), and biological DMARDs (bDMARDs), and analgesics [5]. Several pro-inflammatory cytokines or those receptors including IL-6R, tumor necrosis factor α (TNF-α), or interleukin 1 (IL-1) have been specifically inhibited by bDMARDs [6]. Although DMARDs are available as initial drug treatments, adverse effect from extend use of agents must be considered. One of anti-rheumatic agents, Tocilizumab (TCZ), that binds to IL-6R and blocks downstream signal, has been proved to be effective in treatments of RA and a number of immune diseases [7]. However, some patients

still exhibit lack of response to the anti-IL-6R therapies. Therefore, it is important to identify reliable prognostic biomarkers for selecting appropriate patient population for optimal treatments. Here, we have investigated the serum proteome profiling study of TCZ responder and non-responder groups to discover TCZ response prediction biomarkers using MRM and PRM assays. MRM is a targeted quantitative proteome analytical platform in which QqQ coupled with liquid chromatography (LC). It has become a standard analytical platform for protein target validation in a high-throughput manner [8]. PRM which is another targeted quantification platform in which QE coupled with LC has been emerged with its high resolution and ion trapping feature of instrument [9]. Hence, we expected to compensate the quantitative results from two different assays and to select reliable biomarker candidates. For proteome profiling analysis, pooled serum samples of TCZ response and non-response groups from ten individuals were collected and independent cohort of 47 individual patient serum samples were used for validation phase. Based on profiling and quantitation analysis, it is possible to identify 47 biomarker candidates through both of PRM and MRM in common. Moreover, we specifically developed the multi-biomarker panel to overcome limitations of single biomarkers that often indicates incompetent prediction values and poor representation in particular patient samples. The generated 4-biomarker panel by combining validated biomarker candidates, showed area under receiver operating characteristic (ROC) curve value (AUC) of 0.859 and high prediction rates; 88.52% and 76.23% in TCZ responder and non-

responder patients, respectively. Our MRM/PRM driven serum biomarker development platform allows to verify the promising panel to predict TCZ non-response and is expected to contribute to better management of RA patient treatment (Figure 1).

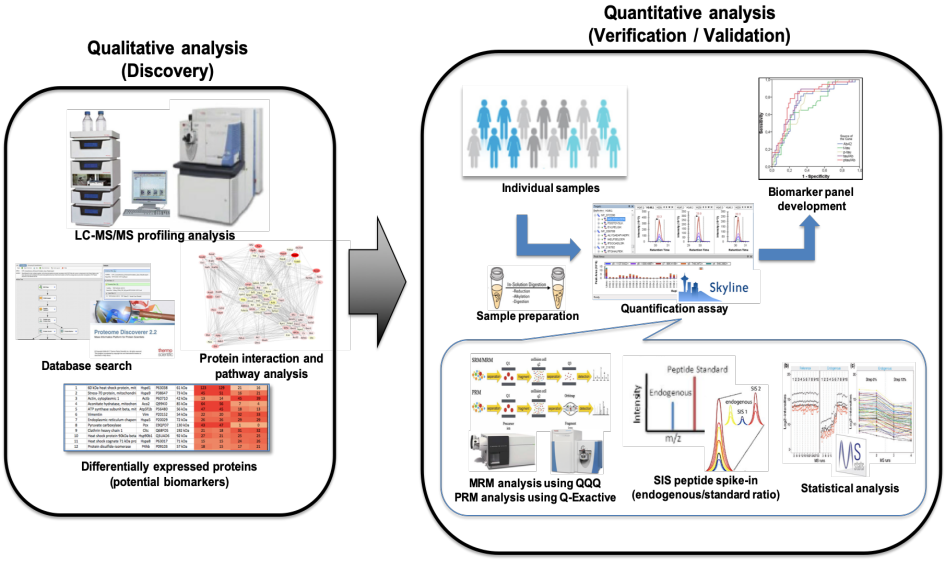


Figure. 1 overall scheme general workflows of biomarker discovery and biomarker panel construction based on proteomic study.

Mass spectrometry coupled with liquid chromatography analysis conducted to identify differentially expressed proteins (DEPs) using bioinformatics with statistical tools and figure out the biological process of those by gene ontology. Selected DEPs were verified and validated with individual patient cohort samples with quantitative analysis using MRM and PRM with statistical analysis.

2. METHODS

2.1 Materials

HPLC grade acetonitrile (ACN), HPLC grade water, formic acid (FA), urea, dithiothreitol (DTT), iodoacetamide (IAA), and ammonium bicarbonate were purchased from Sigma-Aldrich (St. Louis, MO, USA). Sep-pak C18 cartridges were obtained from Waters (Milford, MA, USA). Trypsin protease MS-grade was from Thermo Scientific (Rockford, IL, USA). SIS peptides with a single amino acid labeled with ^{13}C and ^{15}N were synthesized at crude levels from JPT (Berlin, Germany).

2.2 Study population

RA patients were recruited from Rheumatology Clinic, Seoul National University Hospital (Seoul, South Korea) and Sagawa Akira's Clinic (Hokkaido, Japan). Serum samples were obtained from RA patients who were older than 18 at starting TCZ and maintained it during the observation periods; 4, 12, 16, 20 or 24 weeks after the TCZ treatment. Clinical information regarding responsiveness after 6 months of TCZ treatment included tender and swollen joint counts 68 and 66, respectively. C-reactive protein (CRP) level and erythrocyte sedimentation rate (ESR) [10]. Criteria for clinical responsiveness were selected among European League Against Rheumatism (EULAR) response criteria using disease activity

score 28 (DAS-28) and American College of Rheumatology (ACR) criteria that measures 20, 50, or 70% improvements. However, patients who have received other bDMARDs concomitant with TCZ, patients without baseline clinical information, solid or hematologic malignancy, active infection at baseline, or chronic active liver disease were excluded from experiment population. Among samples, ten pooled serums who were either responders (n=10) and non-responders (n=10) to TCZ were included in this study to identify prognostic biomarkers target and global proteome analysis. For larger scale analysis, 88 individual serum samples were divided into two categories as training and test set based on two different centers where the samples were collected. The individual samples in training sets (n= 47) were from Rheumatology Clinic of Seoul National University Hospital and test sets (n=41) were from that of Sagawa Akira's Clinic. All serum samples were immediately aliquoted and frozen at -80°C after centrifugation at 14,000 g for 5 min at 4°C.

2.3 Depletion of high-abundance proteins

Thawed ten pooled serum samples and 47 individual samples in training set were both aliquoted 40 μL total. After 30 seconds brief sonication, all of serum samples were centrifugated at 3,000 g for 5 min to remove remaining debris in serum. The samples were diluted approximately five-folds with the depletion buffer A

(Agilent Technologies, Wilmington, DE, USA). Diluted serum samples were filtered through a 0.22 μm cellulose acetate spin filter (Agilent Technologies, San Diego, CA, USA) and processed to remove the 6 most abundant serum proteins (Albumin, IgG, antitrypsin, IgA, transferrin, and haptoglobin) using Multiple Affinity Removal Spin cartridge (MARS Hu-6 column, Agilent Technologies) following the manufacture protocol with UHPLC-3000 system. After equilibration with the load and wash buffer, 100 μL of diluted serum was loaded on the MARS Hu-6 column at 0.125 mL/min for 16.5 min. The bound proteins were released with elution buffer B (Agilent Technologies, Wilmington, DE, USA) at 1.0 mL/min for 5 min. The column was then washed with the load and wash buffer for 10 min at a flow rate of 1mL/min. Total depletion cycle took 30.1 min. All of depleted fractions were collected and those proteins were concentrated with YM-3 Centricon filter (Millipore, Billerica, MA, USA) and exchange the buffer as 50 mM Tris/HCl (pH = 7.5) for protein elution. Eluted proteins were approximately 150 μL of total and the concentration was measured using a Micro BCA protein Assay Kit (Thermo Scientific, Rockford, IL, USA).

2.4 In-solution tryptic digestion of proteins

Depleted 300 μg serum proteins were incubated in 6 M urea for 1 hour at 37°C. The disulfide bond of denatured proteins was reduced with 10 mM dithiothreitol (DTT) for 1 hour at 37°C, and alkylated with 20 mM iodoacetamide (IAA) for 30

min in the dark at room temperature. Before tryptic digestion, 50 mM ammonium bicarbonate buffer was added to release the urea concentration below 1 M. Trypsin (Promega, Madison, WI, USA) was treated for digestion at 1:50 ratio within enzyme to protein. After incubating at 37°C overnight, digested peptides were desalted using Sep-Pak C18 1cc Vac Cartridge (Waters Corporation, Milford, MA, USA) in accordance of manufacture protocol. The concentration of desalted peptides was measured with Colorimetric peptide Assay kit (Thermo Scientific, Rockford, IL, USA).

2.5 High-pH reverse phase fractionation

In order to increase the protein identification from LC-MS analysis, high-pH reverse fractionation was applied to peptide fractionation of complex samples. A 260 μg digested peptides from TCZ response and non-response groups were dissolved in 100 μL of high pH fractionation buffer A, composed with 10 mM of ammonium formate in water with 0.6% of ammonium hydroxide (pH= 10). An Agilent 1100 series of HPLC system was applied with Accucore XL C-18 (2.1*150 mm, 4 μm, Thermo Scientific, Rockford, IL, USA) fractionation column. A total of 100 μL was injected with releasing the bound peptides with high-pH fractionation buffer B, composed with 10 Mm ammonium formate in 90% ACN with 0.6% of ammonium hydroxide (pH= 10) and the elution gradients are followed; 0-10 minutes with 5%, 10-60 minutes with 5-35%, 60-70 minutes with

35-70%, 70-80 minutes with 70%, and 80-105 minutes with 5% buffer B. The eluted peptides were collected in 96-well RV plate and orthogonally combined in 15 fractions to reduce sample complexity before LC-MS/MS analysis.

2.6 LC-MS/MS analysis

Peptides from ten pooled serum samples were resuspended in 30 μL of Solvent A (0.1%, Formic acid in water) and 3 μL of sample flows through trap column (PepMap^{TM} RSLC C_{18} column 75 μm ID*2 cm 2 μm, Thermo Fisher Scientific) for sample clean up and remove contaminations. After that, sample was loaded onto an analytical column (PepMap^{TM} RSLC, C18 column 75 µm ID*50 cm 2 μm, Thermo Fisher Scientific) and separated with a linear gradient 5-35% Solvent B (0.1% Formic acid in Acetonitrile) for 90 min at a flow rate of 300 nL/min; 0-10 min 2% of solvent B, 11-12 min 5% of solvent B, 13-67 min with 5-35% gradual gradient of solvent B, 68-83 70% of solvent B, and 84-90 min 2% of solvent B for column re-equilibration. MS spectrums were recorded on Q-Exactive (Thermo Fisher Scientific, San Jose, CA) hybrid quadrupole-orbitrap MS coupled with Ultimate 3000 HPLC system (Thermo Fisher Scientific). Standard mass spectrometric condition of the spray voltage was set to 2.0 kV and the temperature of the heated capillary was set to 250°C. The full scans were acquired in range at 350-1400 m/z with 70,000 resolutions and the normalized

collision energy was 27% and 17,500 resolution for high-energy collisional dissociation (HCD) fragmentation. The data-dependent acquisition (DDA) was operated with single survey MS scan followed by ten MS/MS scans in a dynamic exclusion time of 30 seconds.

2.7 Database search for protein identification and functional analysis.

Collected MS/MS raw data was converted into mzXML files through the Trans Proteomic Pipeline (TPP) and searched with SEQUEST® (version v.27, rev. 11) engine-based Proteome Discoverer 1.4 (Thermo Fisher scientific, San Jose, CA) platform. The processed MS/MS peak lists were compared with Uniprot homo sapiens reviewed database containing 20931 entries. Precursor and fragment ion tolerance were set to 10 ppm and 0.8 Da, respectively. The enzyme was selected as trypsin with a maximum allowance of up to two missed cleavages. For post-translational modification, carbamidomethylation (+ 57.0215 Da) for cysteine as fixed modification and oxidation (+ 15.9949 Da) for methionine as dynamic modification. Peptide validator was employed for peptide false discovery rate (FDR) which is between relaxed target FDR 0.05% and strict target FDR 0.01%. Those peptide identifications based on MS/MS spectrum were validated with Scaffold (version 4.6.4, Proteome Software Inc., Portland, OR, USA). The specific threshold was applied which are peptide identification if those were able to be established greater than 95.0% probability of peptides with containing equal

or more than two identified peptides and 99.0% probability of proteins. The classification of cellular component, biological process, and pathways of identified proteins were analyzed using STRING (v 10.5, Swiss Institute of Bioinformatics) [38]. The protein-protein interactions and pathways of differentially expressed proteins (DEPs) were visualized by Cytoscape. (version 3.6.1)

2.8 Relative quantification analysis of identified proteins

We performed relative protein quantification which is label-free quantification along with spectral counts. Scaffold software allows the MS/MS data to compare the spectral counts of identified proteins. The normalized values along duplicate analyses of two different groups were analyzed using Power Law Global Error Model (PLGEM) (http://www.bioconductor.org) package within R program (version 2.15) [11]. After PLGEM analysis, it is possible to distinguish statistically significant DEPs and calculate expression level change by *p*-value and signal-to-noise (STN) [12].

2.9 Selection of PRM/MRM target peptides and transitions

Based on selected DEPs from LC-MS/MS analysis, tryptic peptides whose sequences uniquely represent the parent proteins were preferentially selected as PRM and MRM target peptides. The other criteria of selections are following 1)

charge states of peptides must be considered which exhibit the most detectable ion charge with MS/MS fragmentation, 2) mass-to-charge (m/z) ratio between 400 and 1500, 3) peptide length between 6 and 20 number of amino acids, 4) no missed cleavage during digestion that are unable to represent specific peptides [13]. Based on those criteria, we selected three peptides per target proteins. Unique peptide sequences that were covered by SEQUEST search were selected for the first and rest of peptides were chosen from SRMAtlas (www.srmatlas.org) which archived the peptides analyzed with various mass devices including QqQ. The stable-isotope-labeled standard peptides analogues (SIS peptides) (JPT peptide Technologies GmbH, Berlin, Germany) of target peptides that were synthesized with heavy labeled isotope of carbon and nitrogen (^{13}C and ^{15}N) at arginine or lysine were spiked-into samples and quantified along with endogenous peptides during PRM/MRM assays. Transitions that are pair of precursor and product ions (Q1/Q3) were automatically selected on Skyline software and determined the seven most abundant transitions after MRM and PRM analysis.

2.10 Quantitative analysis: multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM)

After proteome profiling analysis with tandem MS spectrometry, we validated the results with parallel reaction monitoring (PRM) and multiple reaction monitoring

(MRM). Peptides from serum samples were dissolved in 0.1% formic acid and 4 μg of samples along with 200 fmol of SIS peptides were injected. In case of MRM-MS, Agilent 6490 (Santa Clara, CA) triple-quadrupole mass device, coupled with Agilent 1290 Infinity HPLC system monitors product ions with collision-induced dissociation (CID) fragmented manner. Peptide separation was performed with a ZORBAX Eclipse Plus C18 column (2.1 x 150 mm, 3.5 μm particle size, Agilent) at a flow rate of 0.25 mL/min over a 45 min gradient from 2% to 40% acetonitrile in 0.1% formic acid within total of 60 min run time. Acquisition method was performed with following parameters: positive mode capillary voltage with 3500 V, fragmentor of 130 V, ion source gas temperature of 250°C, nebulizer pressure of 50 psi, sheath gas temperature of 350°C. Dynamic MRM scan type was used with 2 min of delta retention time. Collision energies for each peptide were optimized by Skyline software (v 4.1.0 MacCoss Lab, UW). Peak area integration, ratios, coefficient variance (CV), and retention times were manually adjusted also with Skyline if necessary. PRM analysis was performed with Q-Exactive coupled with Ultimate 3000 HPLC system. We applied same C18 column and linear gradients with same solvent from profiling analysis, but the method was changed with PRM tab coupled with full scan. PRM method was performed with following parameters: the full scan was acquired by 35,000 resolutions with automatic gain control (AGC) target value of $3e^6$ and PRM properties were 17,500 resolutions with AGC target value of $1e^6$. Isolation window was set as 2 m/z and NCE was differentially optimized through Skyline

software depends on peptide sequences. The chromatogram peak width was set to 30 s and the other parameters were same as profiling analysis.

2.11 AuDIT analysis of determining precise MRM and PRM transitions

Both of MRM and PRM were performed with triplicate analysis of serum samples and those endogenous SIS peptides. For data integration and statistical analysis, we performed Automated Detection of Inaccurate and Imprecise transitions (AuDIT) analysis that uses p-value of t-test to eliminate the ineligible peptides by measuring the coefficients of variance (CV) of peak area of analytes on QuaSAR program (http://genepattern.broadinstitute.org). The two thresholds were applied to determine the final MRM transitions; p-value lower than 10^{-5} by t-test and CV $\leq$ 20% peak area during triplicate measurements [14].

2.12 Statistical analysis

In order to confirm statistically significant and meaningful DEPs between TCZ responders and non-responders, we performed statistical analysis of MRM and PRM result with MSstats (version 3.13.2) through Skyline [15]. All of transition intensities were converted into logarithmic values for process. After that, quality control (QC) function was performed to normalize the data by equalizing the median peak ratio intensities of reference SIS peptide transitions. The QC

processed over MRM analysis and modified the bias to signal of SIS peptides. The normalized data was applied on protein level quantification by group comparison analysis between response and non-response groups. Target peptides were filtered at 0.05 FDR and p-value, adjusted to references below 0.05 were considered significant. The stepwise logistic regression analysis was performed on SPSS Statistics (IBM, version 25) and it generated new prediction scores based on selected proteins. The scores applied to construct Receiver Operating Characteristic (ROC) curves to attain AUC value and prediction tables were also generated within SPSS Statistics (IBM, version 25).

2.13 Quantitative linear curves

In order to confirm the quantitative function of MRM and PRM, 10 SIS peptides which showed relatively high abundance were spiked into 4 μg for MRM and 2 μg for PRM of digested pooled serum samples and serially diluted into 8 concentrations (0.5, 1, 2.5, 5, 10, 25, 50, and 100 fmol/μl). Both of MRM and PRM were performed in triplicates and the normalized peak abundances were applied to construct linear curve with regression values. Furthermore, SIS peptides, included in developed multi-biomarker panel also applied to construct linear curve to confirm the quantitative feature within complex sample. The peptides were serially diluted into 10 concentrations (1, 2.5, 5, 10, 25, 50, 100, 250, 500, and 1000 fmol) and conducted in triplicate analysis.

RESULTS

3.1 Overall strategy of proteomic approach based on serum biomarker identification for TCZ non-response.

We aimed to discover prediction biomarkers against specific therapeutic agents. Those biomarkers were used to construct multi-marker panel for TCZ response and non-response from RA patients. As described in Figure 2, we selected 54 potential biomarker candidates by LC-MS/MS proteome analysis and literature data mining of known RA biomarkers. Profiling analysis was performed with ten pooled RA patient serum samples of TCZ response and non-response to determine potential biomarker candidates. After discovery process, MRM/PRM analysis was performed to validate the DEPs from pooled serum samples and select final MRM and PRM target. Those targets were selected based on AuDIT analysis that adopts reliable detection ability with an accurate quantitative manner in LC interfaced mass spectrometry systems. As a result, MRM analysis detected 362 transitions derived from 116 target peptides. However, PRM analysis was able to identify 501 transitions derived from 124 target peptides. Therefore, we determined the final prognostic biomarker target candidates for prediction of TCZ response with independently detected peptides from MRM and PRM analysis. For the validation process, we collected 88 individual RA patient serum samples and divided into two independent cohorts which were training (n= 47) and test set

(n= 41) for precise verification and prediction [16]. Multiplexed MRM and PRM assays with SIS peptide spike-in were performed within training set and validated DEPs among responders (n= 26) and non-responders (n= 21) [17]. As a result, 47 and 50 DEPs from MRM and PRM analysis respectively indicated quantitative differences between two different groups with statistically reliable *p*-value lower than 0.05 by MSstats tool. Those detected DEPs were adequately combined to construct multi-biomarker panel based on logistic regression analysis and C-statistics using SPSS in order to discriminate two different groups. Furthermore, we planned to apply multiplexed-MRM and PRM analysis with large independent cohort, test set to demonstrate the prediction performance of constructed multi-biomarker panel.

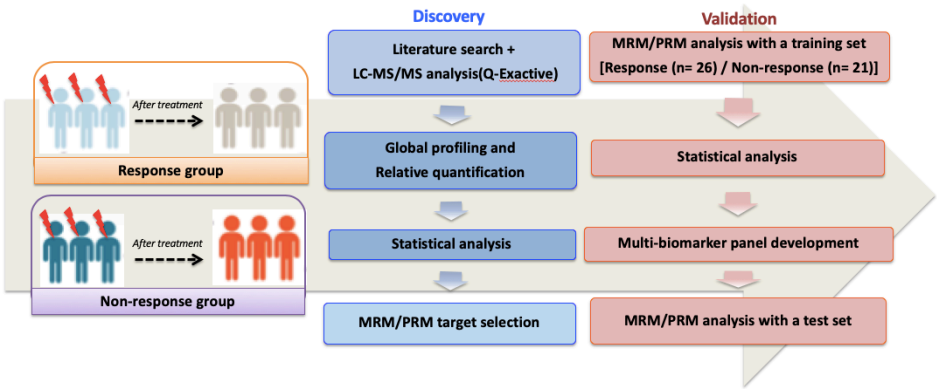


Figure 2. Overview of serum biomarker development for identification of Tocilizumab (TCZ) non-response group.

Initial prognostic biomarker candidates for identification of TCZ, anti-IL6 receptor therapeutic agent, non-response in rheumatoid arthritis patients were selected through literature search and global profiling of RA patient serum samples in the discovery stage. Among identified targets, initial MRM/PRM and AuDIT analysis were performed to conclude MRM/PRM target peptides that show significant quantitative results. In the validation and verification process, multiplexed-MRM/PRM analysis was executed within training set (n= 47) to construct a multi marker panel after statistical confirmation. At last, independent test set was analyzed by multiplexed-MRM/PRM as well to demonstrate predictions of developed multiple biomarker panel.

3.2 Serum sample collection and study population.

The study samples were collected in two different clinical sites, Division of Rheumatology in Seoul National University Hospital (Seoul, South Korea) and Sagawa Akira's Clinic (Hokkaido, Japan). TCZ responders and non-responders were classified according to criteria of American College of Rheumatology, measuring 20% improvements (ACR20), European League Against Rheumatism (EULAR) response using Disease Activity Score in 28 joints (DAS28). The scores were measured between baseline (W0) and 12 weeks after treating TCZ (W12). Among those criteria, 26 samples were categorized as TCZ responders and 21 samples were categorized as TCZ non-responders based on ACR20 response (Table. 1).

3.3 Serum proteome analysis of TCZ responding and non-responding RA patients

To identify TCZ prediction protein biomarker candidates, we selected 10 TCZ responding and non-responding patients (W0) from longitudinal study samples collected from individuals. Clinical data selected for the serum proteome analysis are summarized in Table 2. Each of ten pooled serum sample were depleted to remove high abundant proteins (Figure 3A) and those were in-solution digested with trypsin. Prior to LC-MS/MS analysis, peptides were followed with high pH

fractionation to reduce sample complexity and fractionized peptides were orthogonally integrated as total of 15 fractions. Those peptides were analyzed by Q-Exactive coupled with nano-LC. As described in figure 3B, a total of 711 proteins were identified through Proteome Discoverer software (version 1.4) based on SEQUEST search engine with specific thresholds; peptide probability greater than 95.0%, a protein probability greater than 99.0%, and contained at least 2 minimum numbers of peptides. We further conducted label free quantitation with PLGEM-STN analysis and quantified 78 DEPs in with *p*-value less than 0.05. Among those, 39 proteins are up-regulated and the other 39 proteins are down-regulated in TCZ non-response groups compared with the protein expression levels of TCZ response groups (Figure 3B). The DEPs furtherly applied to STRING interaction analysis to understand protein networks and biological process. It revealed that most of up-regulated proteins in TCZ non-responder group were categorized with inflammatory response, defense response, and coagulation that were mostly located in extracellular space (Figure 4A). We additionally sorted out DEPs with logarithmic scale of fold-change value ($1.2>\log_2 FC$) and specific biological process (Table 3). Collectively, 54 biomarker candidates (Figure 4B) were selected with adding reported TCZ response markers; osteopontin (OPN) [18], interleukin 8 (CXCL8), and C-C motif chemokine 2 (CCL2) [19].

Table. 1 Clinical characteristic of individual serum samples of RA patients.

	* Response group (n= 26)	* Non-response group (n=21)
Age		
Median (range)	52 (28-72)	55 (36-72)
Sex		
Female / Male (Female %)	24/2 (92.3%)	18/3 (85.7%)
American College of Rheumatology (ACR) scoring, non-responder/responder (non-responder %)		
ACR50	14/12 (53.8%)	21/0 (100%)
ACR70	26/0 (100%)	21/0 (100%)
Disease Active Score 28 (DAS-28), median (range)		
w0	6.2 (4.4-7.7)	6.1 (5.0-7.8)
w12	2.6 (1.0-4.9)	4.1 (1.8-5.6)
w12 remission/non-remission (non-remission %)	13/13 (50%)	20/1 (95.2%)
European League Against Rheumatism (EULAR), n (%)		
Good response	20 (76.9%)	3 (14.3%)
Moderate response	6 (23.1%)	15 (71.4%)
No response	0 (0%)	3 (14.3%)

∗ Response and non-response groups were categorized according to American College of Rheumatology criteria with improvement of 20% (ACR20).

Table 2. Clinical characteristic of pooled serum samples of RA patients.

	Response group (n= 10)					Non-response group (n=10)				
Treatments duration, w (weeks)	*W0*	*W12*	*W16*	*W20*	*W24*	*W0*	*W12*	*W16*	*W20*	*W24*
Age										
Median (range)			47 (28-67)					56 (36-67)		
Sex										
Female / Male (Female %)			9/1 (90.0%)					10/0 (100.0%)		
American College of Rheumatology (ACR), non-responder/responder										
ACR20	-	0/10	-	0/10	0/10	-	10/0	-	9/1	10/0
ACR50	-	2/8	-	2/8	1/9	-	10/0	-	10/0	10/0
ACR70	-	10/0	7/3	6/4	8/2	-	10/0	10/0	10/0	10/0
Disease Active Score 28 (DAS-28), n										
Remission	-	8	-	8	9	-	0	-	1	0
No Remission	-	2	-	2	1	-	10	-	9	10
Score (median)	5.80	2.18	-	1.92	1.72	5.99	4.49	-	4.70	4.32
European League Against Rheumatism (EULAR), n										
Good response	-	9	10	10	-	-	0	-	1	1
Moderate response	-	1	0	0	-	-	7	-	5	5
No response	-	0	0	0	-	-	3	-	4	4

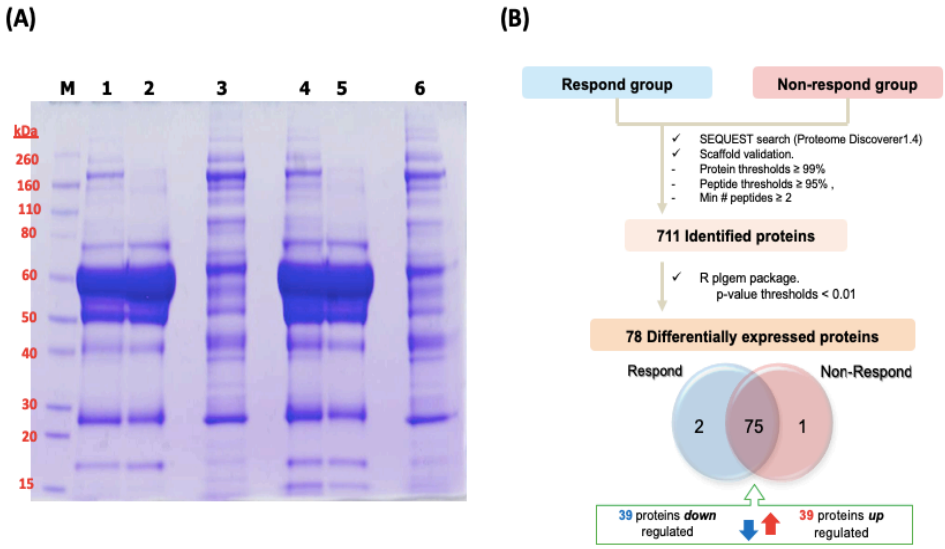


Figure 3. SDS-PAGE fraction of depleted serum proteins and Venn diagram of identified proteins with DEPs by mass spectrometry analysis.

Pooled immune-depleted serum proteins of TCZ response (n=10) and non-response (n=10) were separated on Bolt^{TM} 4-12% Bis-Tris gel. Coomassie brilliant blue was used for staining and able to confirm that most abundant proteins were depleted compared intact or high fractions with low fractions from depletion. Protein markers (lane M), intact serum samples of TCZ treated, high and low fractions of depleted serum samples of response group (each 20 μg in lane 1, 2, and 3, respectively), and intact serum samples of TCZ treated, high and low fractions of depleted serum samples of non-response group (each 20 μg in lane 4, 5, and 6, respectively) (A).

A total of 711 serum proteins and 78 DEPs between TCZ response and non-response group were identified by duplicate LC-MS/MS analysis (B).

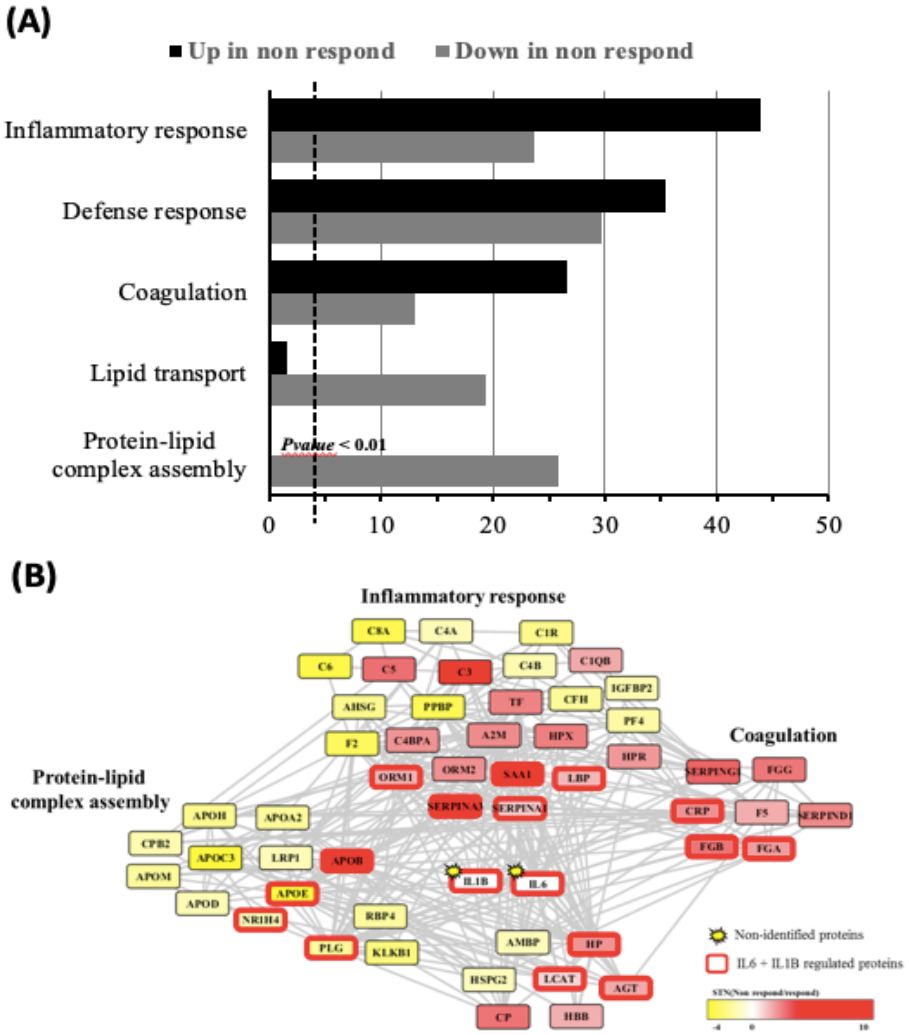


Figure 4. DEPs from pooled serum samples of response and non-response groups and gene ontology (GO)

Among 78 DEPs, we selected 51 serum proteins based on network analysis and were categorized into biological process using STRING database tool. Specific functions were represented in logarithmic scale of p-value and each of response

and non-response groups are shown in grey and black, respectively (A). Those proteins were visualized as color representation with fold-change values of STN. We additionally included 3 known biomarkers of RA by literature search. Therefore, a total of 54 proteins were selected as prediction biomarker candidates (B).

Table 3. Differentially expressed serum proteins between TCZ response and non-response based on network analysis.

Uniprot ID	Gene Name	Protein name	STN*	P-value
P01024	C3	Complement C3	12.424	0.000
P01011	SERPINA3	Alpha-1-antichymotrypsin	9.147	0.000
P04114	APOB	Apolipoprotein B-100	7.661	0.000
P0DJI8	SAA1	Serum amyloid A-1 protein	6.530	0.000
P05155	SERPING1	Plasma protease C1 inhibitor	3.487	0.000
P01031	C5	Complement C5	3.200	0.000
P00450	CP	Ceruloplasmin	3.164	0.000
P02675	FGB	Fibrinogen beta chain	3.006	0.000
P02790	HPX	Hemopexin	2.925	0.000
P02679	FGG	Fibrinogen gamma chain	2.860	0.000
P02787	TF	Serotransferrin	2.690	0.001
P01023	A2M	Alpha-2-macroglobulin	2.657	0.001
P05546	SERPIND1	Heparin cofactor 2	2.587	0.001
P00738	HP	Haptoglobin	2.438	0.001
P19652	ORM2	Alpha-1-acid glycoprotein 2	2.420	0.001
P04003	C4BPA	C4b-binding protein alpha chain	2.388	0.001
P02741	CRP	C-reactive protein	2.344	0.001
P00739	HPR	Haptoglobin-related protein	2.275	0.001
P02671	FGA	Fibrinogen alpha chain	2.162	0.002
P01019	AGT	Angiotensinogen	1.876	0.003
P02746	C1QB	Complement C1q subcomponent subunit B	1.872	0.003
P12259	F5	Coagulation factor V	1.788	0.005

P02763	ORM1	Alpha-1-acid glycoprotein 1	1.752	0.005
P68871	HBB	Hemoglobin subunit beta	1.730	0.005
P18428	LBP	Lipopolysaccharide-binding protein	1.593	0.006
P01009	SERPINA1	Alpha-1-antitrypsin	1.164	0.015
P04180	LCAT	Phosphatidylcholine-sterol acyltransferase	0.983	0.022
P98160	HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein	-1.480	0.007
Q96RI1	NR1H4	Bile acid receptor	-1.512	0.006
Q07954	LRP1	Prolow-density lipoprotein receptor-related protein 1	-1.542	0.006
P02760	AMBP	Protein AMBP	-1.587	0.006
P05090	APOD	Apolipoprotein D	-1.658	0.006
P0C0L4	C4A	Complement C4-A	-1.671	0.006
P18065	IGFBP2	Insulin-like growth factor-binding protein 2	-1.736	0.005
Q96IY4	CPB2	Carboxypeptidase B2	-1.758	0.005
P0C0L5	C4B	Complement C4-B	-1.801	0.004
P02776	PF4	Platelet factor 4	-1.997	0.002
P02753	RBP4	Retinol-binding protein 4	-2.181	0.002
O95445	APOM	Apolipoprotein M	-2.214	0.001
P02652	APOA2	Apolipoprotein A-II	-2.309	0.001
P00747	PLG	Plasminogen	-2.340	0.001
P08603	CFH	Complement factor H	-2.346	0.001
P02749	APOH	Beta-2-glycoprotein 1	-2.545	0.001
P02765	AHSG	Alpha-2-HS-glycoprotein	-2.554	0.001
P00736	C1R	Complement C1r subcomponent	-2.612	0.001
P03952	KLKB1	Plasma kallikrein	-2.884	0.000

P00734	F2	Prothrombin	-3.708	0.000
P02775	PPBP	Platelet basic protein	-3.738	0.000
P07357	C8A	Complement component C8 alpha chain	-3.868	0.000
P13671	C6	Complement component C6	-4.014	0.000
P02656	APOC3	Apolipoprotein C-III	-4.235	0.000
P02649	APOE	Apolipoprotein E	-4.317	0.000

*STN: Signal-to-noise ratio obtained from PLGEM analysis

3.4 Final biomarker candidate selection by preliminary PRM/MRM assays with AuDIT analysis.

Among selected 54 potential biomarker candidates, at least 3 unique peptides of target proteins were selected from SRMAtlas (http://www.srmatlas.org) [35], PeptideAtlas (http://www.peptideatlas.org) [36], and National Cancer Institute Office of Cancer Clinical Proteomics (https://proteomics.cancer.gov) [37] that archived experimental data. Those unique peptides followed the general selection criteria such as features to uniquely represent the specific proteins, 6 - 20 amino acid length of peptides, the charge states that can be optimally monitored by specific MS device, and peptides without miss cleavage and post translational modifications [34]. Therefore, 165 peptides were selected and in-silico digested in Skyline software for fragmented ion selection. MRM and PRM detectability were preliminary verified by analysis of 4 μg and 2 μg of pooled serum samples, respectively. Before analysis, SIS peptides were spiked into two groups of serum samples and 200 fmol was injected for peak abundance normalization and precise retention time confirmation. The abundance of fragmented ions, presented as transitions which are pair of precursor and product ions of target peptides represent each target protein abundance. All of transitions were measured via Skyline software and at least 3 transition candidates were selected by intensity order based on qualitative analysis results with criteria described in figure 5. The ion intensity order of selected transitions was referred with the public deposited

transition data within ProteomeXchange Consortium via the PRIDE [20], Institute of System Biology data from PeptideAtlas [21], or NIST libraries [22]. Among detected transitions, inaccurate transition data was eliminated based on inconsistent reproducibility during replicates or interfering signals by coefficient of variance lower than 20% and p-value lower than 10^{-5} by t-test through AuDIT analysis [14]. As chromatograms from two different quantitative analysis described in Figure 6, 116 peptides derived from 47 proteins and 124 peptides derived from 50 proteins were selected from preliminary MRM and PRM analysis, respectively. All of the proteins detected by MRM were adequately detected in PRM analysis and it uniquely monitored 3 additional proteins. All of protein expressions were statistically analyzed with MSstats that equalizes to median of ion intensities and results were normalized with peak area ratios between SIS peptides and corresponding endogenous. The DEPs from TCZ non-response group were represented in color scheme of logarithmic value of fold change ($\log_2$FC) and specific functions of DEPs were also explained in Figure 7. At last, the followed DEPs were considered as final prediction biomarker candidates (Table 4). Among them, C-reactive protein (CRP) that has already been used as clinical standard for osteoclastogenesis in RA patients [23] and haptoglobin (HP) reported as its expression signal was increased about 5.1-fold in RA patients who were resistant to Methotrexate, one of the DMARDs [24]. Moreover, it was recently reported that fibrinogen gamma chain (FGG) was overexpressed in

polyarticular group of juvenile idiopathic arthritis disease [25], also known as juvenile rheumatoid arthritis.

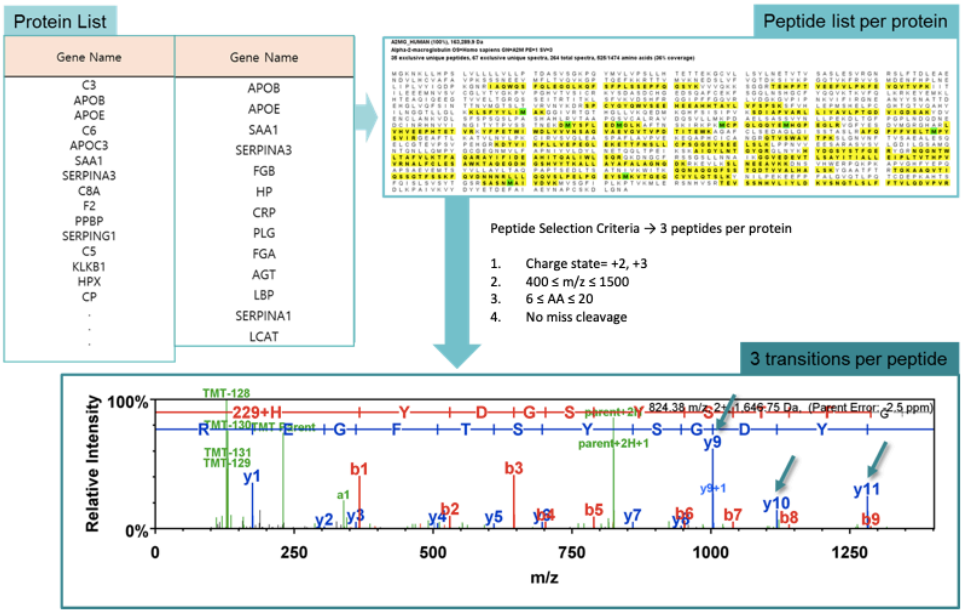


Figure 5. Selection of PRM and MRM target peptides and transitions of biomarker candidates.

Basically, MRM/PRM target peptides were selected from detected unique sequences from our profiling search results using PD 1.4 with including peptides from SRMAtlas and National Cancer Institute Office of Cancer Clinical Proteomics. After peptide selections, the transitions of each peptides were selected followed by those order of ion intensities. The order of transitions was confirmed with the mass spectrometry proteomics data from PRIDE, ISB, and NIST.

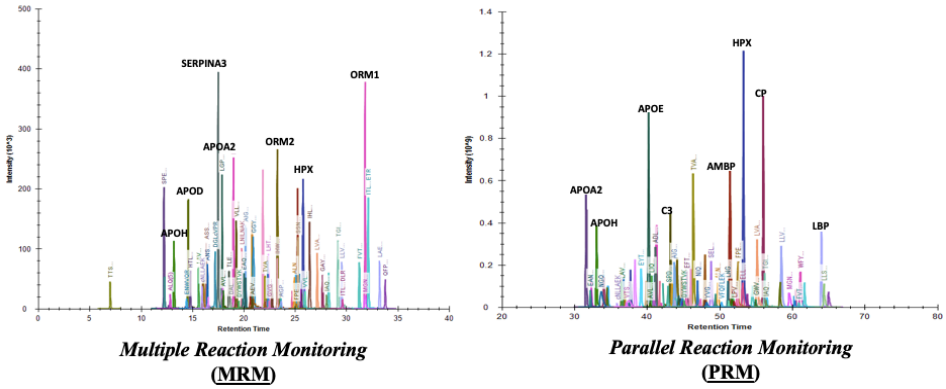


Figure 6. Chromatograms of preliminary MRM/PRM analysis of candidate serum protein biomarkers for prediction of TCZ response groups in RA patients

Among 171 target unique peptides from profiling analysis and public data source such as SRMAtlas, PeptideAtlas, and National Cancer Institute Office of Cancer Clinical Proteomics, MRM could identify 116 peptides and PRM identified 124 peptides. Injected amount of two different groups of serum samples were 5ug and 2ug for MRM and PRM respectively, with constant 200 fmol of SIS peptides. The detected transitions from MRM and PRM were 724 and 1024 transitions with including transitions of SIS peptides as well.

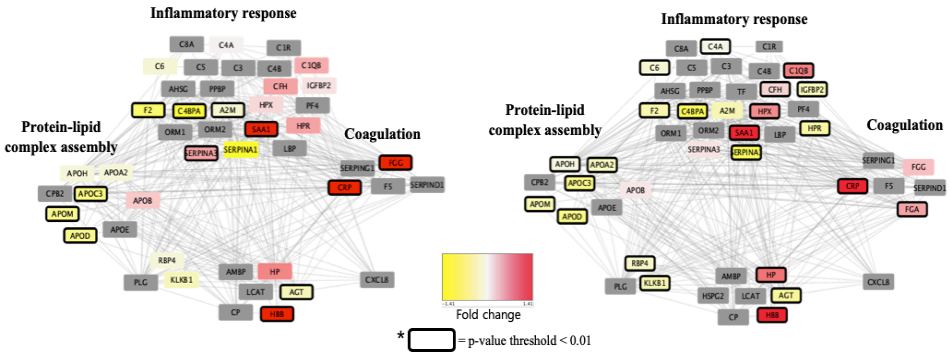


Figure 7. Protein expressions of target proteins from pooled serum samples after AuDIT analysis by MRM/PRM analysis.

The protein expressions between TCZ response and non-response group by MRM and PRM analysis. In case of PRM, it was possible to detect 50 proteins that includes all of 47 proteins monitored from MRM analysis. However, both of methods could not detect 4 proteins in common; FGB, NR1H4, LRP1, and SPP1. The expressions were represented as logarithmic fold-change values (log_2FC) in color scheme and proteins with p-value lower than 0.01 were specifically marked. Most of up-regulated proteins were related with inflammatory response and coagulation functions among total of DEPs. Those monitored proteins were set to final candidates for prediction biomarkers and validated with individual samples.

Table. 4 Final MRM target transitions after AuDIT analysis

Uniprot ID	Gene Name	Peptide sequence	Precursor Ion (m/z)	Product Ion (m/z)	Ion Name
P02763	ORM1	SDVVYTDWK	556.77	811.40	y6
		YVGGQEHFAHLLILR	584.99	560.34	y9
		WFYIASAFR	580.80	974.51	y8
P19652	ORM2	EHVAHLLFLR	412.24	661.44	y5
		TEDTIFLR	497.76	548.36	y4
P01011	SERPINA3	LYGSEAFATDFQDSAAAK	946.44	1124.52	y11
		ITLLSALVETR	608.37	1001.60	y9
		ADLSGITGAR	480.76	774.45	y8
P01009	SERPINA1	AVLTIDEK	444.76	718.40	y6
		DTEEEDFHVDQVTTVK	631.29	889.50	y8
P02765	AHSG	CNLLAEK	424.22	687.40	y6
		FSVVYAK	407.23	666.38	y6
		HTLNQIDEVK	399.55	731.39	y6
P01023	A2M	LHTEAQIQEGTVVELTGR	704.03	674.38	y6
		IAQWQSFQLEGGLK	802.93	1292.66	y11
		AIGYLNTGYQR	628.33	1071.52	y9
P01019	AGT	SLDFTELDVAAEK	719.36	1122.57	y10
		ALQDQLVLVAAK	634.88	713.49	y7
P02652	APOA2	SPELQAEAK	486.75	885.47	y8
		EQLTPLIK	471.29	684.47	y6
P04114	APOB	FPEVDVLTK	524.29	361.24	y3
		TGISPLALIK	506.82	654.45	y6
P02656	APOC3	GWVTDGFSSLK	598.80	953.49	y9
		DYWSTVK	449.72	620.34	y5

		DALSSVQESQVAQQAR	572.96	887.47	y8
P05090	APOD	NPNLPPETVDSLK	712.38	1098.60	y10
		VLNQELR	436.25	772.43	y6
		IPTTFENGR	517.77	723.34	y6
P02649	APOE	SELEEQLTPVAEETR	865.93	801.41	y7
		LAVYQAGAR	474.77	835.44	y8
O95445	APOM	DGLCVPR	408.71	644.35	y5
		WIYHLTEGSTDLR	530.94	288.20	y2
		AFLLTPR	409.25	599.39	y5
P02749	APOH	LGNWSAMPSCK	625.79	491.23	y4
		ATVVYQGER	511.77	850.44	y7
		EHSSLAFWK	368.85	480.26	y3
P04003	C4BPA	YTCLPGYVR	564.78	864.44	y7
		FSAICQGDGTWSPR	791.36	875.40	y8
		GVGWSHPLPQCEIVK	569.63	747.38	y12
Q96IY4	CPB2	QVHFFVNASDVDNVK	573.62	689.35	y6
		YSFTIELR	514.77	778.45	y6
		AVASFLR	382.23	593.34	y5
P00450	CP	EYTDASFTNR	602.27	911.42	y8
		GAYPLSIEPIGVR	686.39	1080.64	y10
P12259	F5	AEVDDVIQVR	572.30	729.43	y6
		FTVNNLAEPQK	630.83	913.47	y8
P02746	C1QB	IAFSATR	383.22	581.30	y5
P00736	C1R	MGNFPWQVFTNIHGR	601.96	943.51	y8
P01024	C3	IHWESASLLR	404.56	488.32	y4
		LVAYYTLIGASGQR	756.41	902.51	y9

		VLLDGVQNPR	555.82	898.47	y8
P0C0L4	C4A	ANSFLGEK	433.22	680.36	y6
		DSSTWLTAFVLK	684.36	977.58	y8
P0C0L5	C4B	ASSFLGEK	419.72	767.39	y7
P01031	C5	FQNSAILTIQPK	680.39	1084.64	y10
		VFQFLEK	455.76	664.37	y5
P13671	C6	ALQEYAAK	447.24	709.35	y6
		ALNHLPLEYNSALYSR	620.99	810.41	y7
		SEYGAALAWEK	612.80	845.45	y8
P07357	C8A	HTSLGPLEAK	351.53	557.33	y5
		LGSLGAACEQTQTEGAK	860.91	991.47	y9
		LYYGDDEK	501.72	889.36	y7
P08603	CFH	SPDVINGSPISQK	671.35	716.39	y7
		SSNLIILEEHLK	465.93	768.43	y6
		NGQWSEPPK	521.75	871.43	y7
P02741	CRP	GYSIFSYATK	568.78	916.48	y8
		ESDTSYVSLK	564.77	912.47	y8
P02679	FGG	IHLISTQSAIPYALR	561.66	890.51	y8
P00738	HP	DIAPTLTLYVGK	645.87	1062.62	y10
		VVLHPNYSQVDIGLIK	599.01	543.39	y5
		VGYVSGWGR	490.75	661.34	y6
P00739	HPR	SCAVAEYGVYVK	673.33	928.48	y8
		VTSIQHWVQK	409.23	563.80	y9
		VGYVSGWGQSDNFK	772.36	1125.50	y10
P68871	HBB	EFTPPVQAAYQK	689.85	904.49	y8

		LLVVYPWTQR	637.87	1048.56	y8
P02790	HPX	GGYTLVSGYPK	571.30	763.43	y7
		NFPSPVDAAFR	610.81	959.49	y9
P05546	SERPIND1	QFPILLDFK	560.82	992.58	y8
		TLEAQLTPR	514.79	814.44	y7
		LNILNAK	393.25	558.36	y5
P18065	IGFBP2	LEGEACGVYTPR	676.32	272.17	y2
		GPLEHLYSLHIPNCDK	631.65	442.21	y7
P10145	CXCL8	ELCLDPK	437.72	359.19	y3
		ENWVQR	416.21	303.18	y2
P18428	LBP	LAEGFPLPLLK	599.37	680.47	y6
		VQLYDLGLQIHK	476.27	600.34	y10
		ITLPDFTGDLR	624.34	1033.53	y9
P04180	LCAT	MAWPEDHVFISTPSFNYTGR	785.70	941.45	y8
		TYSVEYLDSSK	646.31	841.39	y7
		SSGLVSNAPGVQIR	692.88	941.52	y9
P03952	KLKB1	GVNFNVSK	432.73	594.32	y5
		IYSGILNLSDITK	718.90	903.51	y8
P05155	SERPING1	VLSNNSDANLELINTWVAK	701.04	831.47	y7
		GVTSVSQIFHSPDLAIR	609.66	835.95	y15
		TNLESILSYPK	632.84	807.46	y7
P00747	PLG	FVTWIEGVMR	619.32	890.46	y7
		CTTPPPSSGPTYQCLK	897.41	1237.59	y11
		EAQLPVIENK	570.82	940.55	y8
P02775	PPBP	EESLDSDLYAELR	770.36	1081.52	y9

		TTSGIHPK	280.82	638.36	y6
		NIQSLEVIGK	550.82	873.50	y8
P02776	PF4	AGPHCPTAQLIATLK	526.62	786.51	y7
		ICLDLQAPLYK	667.36	591.35	y5
		HITSLEVIK	520.31	789.47	y7
P02760	AMBP	AFIQLWAFDAVK	704.88	1190.66	y10
		ETLLQDFR	511.27	892.49	y7
		TVAACNLPIVR	607.34	1013.56	y9
P00734	F2	ELLESYIDGR	597.80	839.39	y7
		ETWTANVGK	503.25	488.28	y5
P02753	RBP4	DPNGLPPEAQK	583.30	839.46	y8
P0DJI8	SAA1	EANYIGSDK	498.74	796.38	y7
		SFFSFLGEAFDGAR	775.87	1169.56	y11
		GPGGVWAAEAISDAR	728.86	903.45	y9

3.5 Evaluation and comparison between MRM and PRM targeted quantification.

In accordance of quantification analysis from two different targeted quantification methods, we evaluated and compared the results and characteristics between those. In the MRM workflow, the first quadrupole (Q1) functions as a mass filter to select precursor ion of target peptides and fragments those in second quadrupole (q2) with collision-induced dissociation (CID) manner. The predefined product ions are selected in the last quadrupole (Q3) and intensity of transitions were measured on detector. This technique is highly sensitive with pre-selected product ions and high reproducibility during replications. The other assay, PRM includes same process until q2 but, fragments precursor ions with high-energy collisional dissociation (HCD) and applies orbitrap instead of Q3. The high resolution of Q-Exactive allows HCD fragmentation that results highly accurate results. In this study, we employed these two quantification assays and as a results, PRM could detect total of 50 proteins including all of MRM detected proteins with 3 more targets; TF, HSPG2, FGA. Both analyses monitored 108 unique peptides in common but, 17 and 9 peptides were uniquely identified by PRM and MRM, respectively. Furthermore, we focused on a quantitative feature between those. The top 10 peptides which showed the highest abundances were selected and measured the selectivity and reproducibility from MRM and PRM analysis. First of all, selectivity which is the quality of selecting precise targets

was measured with STN between total peak area and background peak area from Skyline output [27]. On Skyline software, the peak intensities which over 10 are considered as signal or detected peptides [28]. Although both analyses specifically detect target peptides, PRM results the relatively high STN among 10 peptides than MRM analysis as shown in figure 8B. It implies that PRM assay is much selective with lower interferences in measurements compared with MRM. We also computed reproducibility between two different analysis with those peptides. Both of analysis were performed with triplicate analysis and lower than 20% CV of peak areas are considered as reproducible during multiple replications [29]. The results represented that both of methods were reproducible with lower than 20% CV, however MRM assay showed relatively stable variances with lower than almost 10% in average (Figure 8B). At last, sensitivity and quantification ability were evaluated by the linear curve of peak abundance with a range of analyte's concentrations (0.5, 1, 2.5, 5, 10, 25, 50, and, 100 fmol). As described in figure 8C, we were able to conclude that both of methods indicated eligible quantitative ability with high coefficient of determination (R^2 value); 0.9754 and 0.9659 for MRM and PRM, respectively. Moreover, R^2 value of MRM was close to 1 compared with that of PRM. It means that MRM was relatively sensitive in detection of abundance components even with low concentrations.

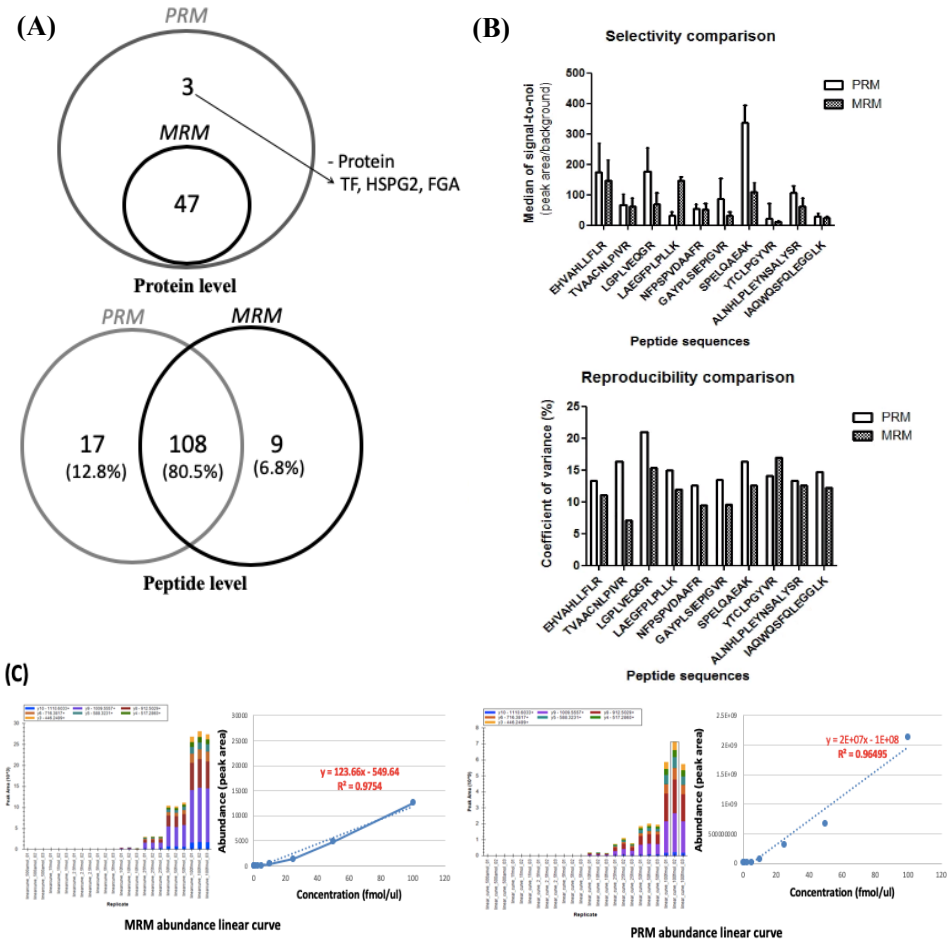


Figure 8. Comparison between two different quantification methods with aspects; sensitivity, selectivity, reproducibility, and quantification ability.

The quantitative features of MRM and PRM methods were compared based on our results. Figure 8A shows the identified proteins and peptides between MRM and PRM assays. 47 proteins were identified via MRM and PRM was possible to detect all of proteins from MRM with including 3 additional proteins; TF, HSPG2,

FGA. In figure 8B, we compared the selectivity by signal-to-noise ratio (STN) of each target peptides. Each STN was calculated by Skyline output which is total area divided by background area. The reproducibility of both methods was compared by coefficient of variance of each replicate. We also compared the quantification ability, represented in figure 8C, by linear curve. The range of concentrations for linear curve was 0.5, 1, 2.5, 5, 10, 25, 50, and, 100 fmol in both of MRM and PRM analysis.

3.6 MRM validation of serum biomarker candidates of TCZ response in individual sample of RA patients.

After final biomarker candidate selection from serum proteome profiling and preliminary-PRM/MRM analysis of RA patients, we validated the target candidates within individual patient serum samples. The 47 individual serum samples, categorized as training set were randomized prior to MRM analysis to avoid bias from technical MS performance variation. In the validation stage, we compared the protein expressions of 54 potential serum biomarkers between TCZ responders (n= 26) and TCZ non-responders (n= 21) by 4 replicates of MRM analysis in training set. One example of training set chromatograms is described in figure 9. The quantitative data was imported into Skyline with normalized peak area ratio based on corresponding each SIS peptides. After MSstats analysis, we detected all of 114 peptides from 47 target proteins as DEPs. The measured abundances of proteins in training set were utilized to build ROC curve and calculated AUC values for prediction specificity and sensitivity. The validated proteins were applied to construct multi-biomarker panel through statistical analysis and the specific proteins included in panel would further verified with independent large cohort samples, classified as test set (Figure 10) [26].

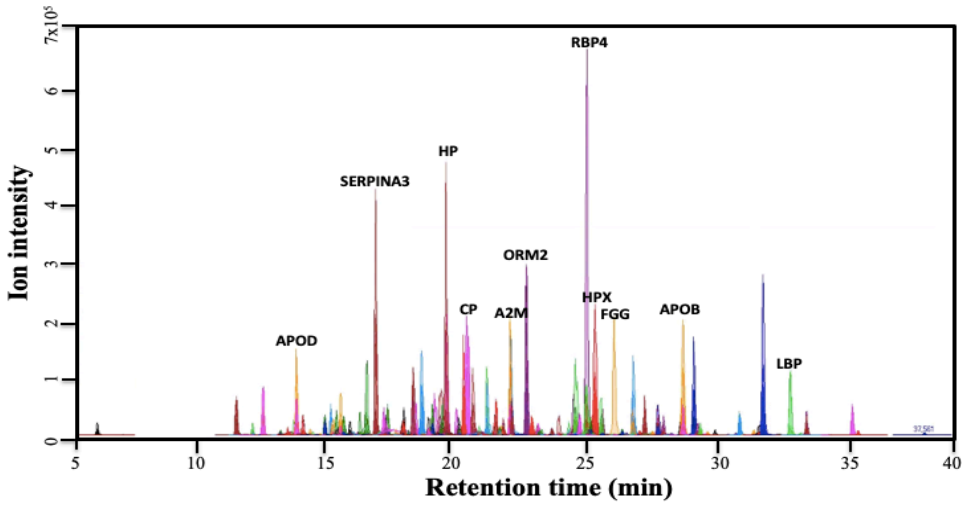


Figure 9. Transition chromatograms of MRM analysis of potential serum biomarker for TCZ non-response predictions in RA patients.

MRM analysis chromatograms of 114 peptides with corresponding SIS peptides attained from 54 target proteins. Ion intensities were measured with observed 724 transitions from 114 target peptides including SIS peptides. Some peaks with major intensities were marked with those protein name. The chromatograms were represented in Agilent MassHunter Qualitative Analysis B.06.00.

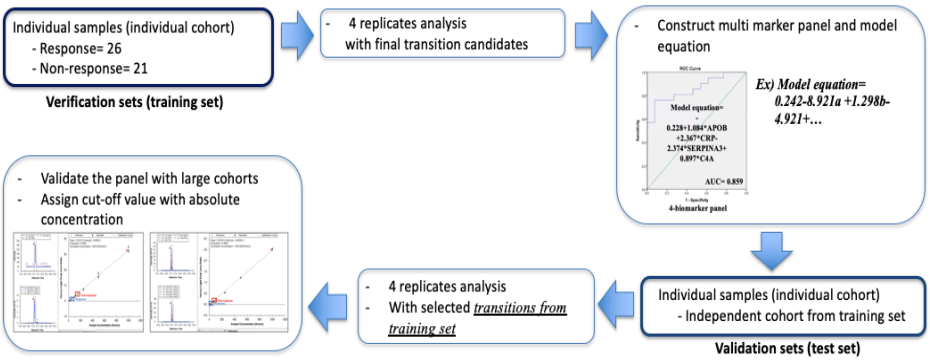


Figure 10. Overall scheme of validation stage with dividing individual patient samples as training and test set.

Among individual samples, we divided them into two independent cohorts for precise validation. First set is training set with 47 patients serum samples with 26 TCZ response and 21 TCZ non-response. The MRM analysis of training set was performed with final biomarker candidates with 4 replicate analysis. Based on the results of analysis, we generated ROC curve of each target and constructed multi biomarker panel for TCZ treatment prediction. The multi-marker panel was also verified with the results from MSstats.

3.7 Multiple biomarker panel construction for prediction of TCZ response and non-response RA patients.

The final purpose of our study was to develop multi-biomarker panel that could predict TCZ response or non-response before treatment to RA patients. As described in serum proteome profiling and quantitative analysis, individual target proteins significantly suggested different expressions between two different groups. However, each individual protein had limitation to predict TCZ response as single protein marker regarding its sensitivity and specificity. Therefore, we developed multi-biomarker panel with combining proteins that shows best prediction power among target proteins. Before multi marker production, we refined the results with adopting only up-regulated proteins from MSstats for much precise prediction over TCZ non-response sample group. The equation of multi-biomarker panel was constructed by stepwise method logistic regression on SPSS and best protein combination was developed including apolipoprotein B-100 (APOB), c-reactive protein (CRP), alpha 1-antichymotrypsin (SERPINA3), and complement C4-A (C4A). Although individual protein among 4-multi marker panel indicated insufficient AUC values for prediction as a single biomarker (AUC values= 0.676, 0.672, 0.625, and 0.452 of CRP, APOB, C4A, and SERPINA3, respectively) (Figure 11), the combination of those in multi-biomarker panel showed much high and sufficient AUC value with 0.859 as described in figure 12. Among all of 4 selected proteins, APOB, CRP, and

SERPINA3 were also statistically meaningful with *p*-value lower than 0.05 (*p*-value= 0.029, 0.004, and 0.007, respectively) but, C4A showed a very slight missed with significant level (*p*-value= 0.063). Most of selected proteins in panel were highly up-regulated in TCZ non-response group by the results of MSstats. It implies that our 4-biomarker panel was statistically and technically proper multi-biomarker panel (figure 12). Furthermore, the constructed 4-biomarker panel was able to discriminate two different groups before TCZ treatments with high prediction rates. The table 5 describes the prediction rate of 4-biomarker panel that could identify 23 of 26 in TCZ response group with 88.52% and 16 of 21 in TCZ non-response group with 76.23%. Those prediction rates were also with high sensitivity and specificity rate with 82.14% and 84.21%, respectively. In addition, hierarchical clustering analysis was implemented for distribution patterns of normalized peak area's ratio of DEPs using Multi Experiment Viewer (version 4.9). The figure 13 displayed the clustering of total 47 biomarker candidates (Figure 13A) and only 4 proteins that are selected in multi-biomarker panel (Figure 13B). The clustering map of 4-biomarker panel exhibited relatively clear distribution rather than total candidates clustering among TCZ response and non-response sample groups. Collectively, we measured absolute concentration of 4 biomarker proteins on multi-marker panel as shown in table 6. It was possible to calculate the concentrations of 4 markers with the known amounts of SIS peptides which were spiked-in. The concentrations were calculated with the ratio

between average transition intensities of 4 target SIS peptides and those endogenous samples among 47 individual samples in training set.

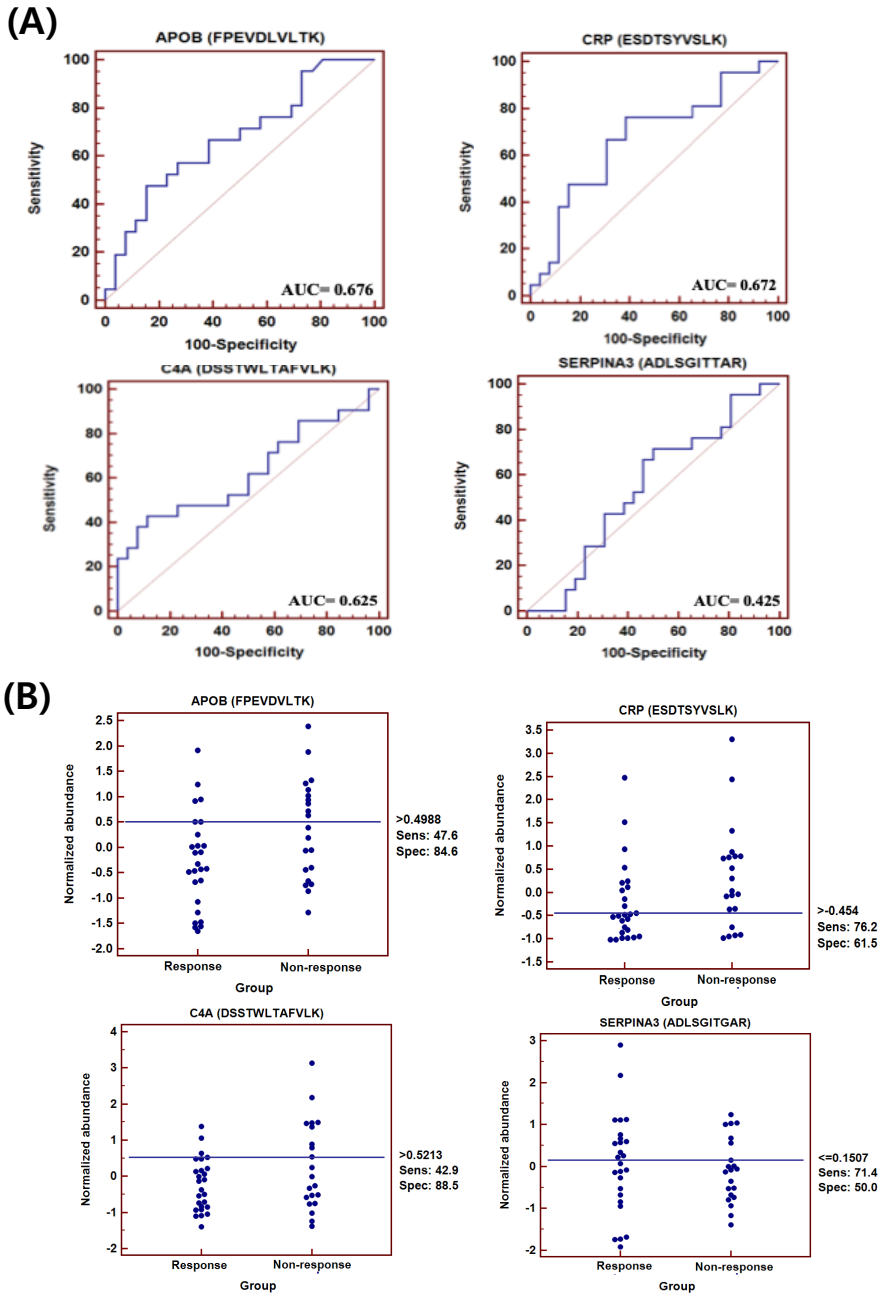
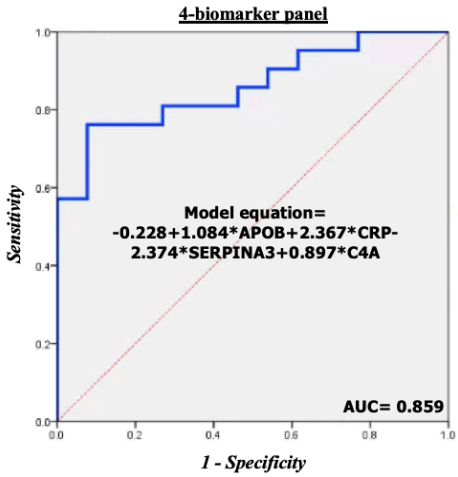


Figure 11. ROC curves with AUC value and scatter plots of each four multi biomarker porteins in a training set.

The normalized peak area of each transition was compared between TCZ responders (n= 21) and non-responders (n= 26) in training set (n= 41). The AUC values of proteins were 0.676, 0.672, 0.625, and 0.452 in APOB, CRP, C4A, and SERPINA3 (A). The scatter plots were represented in ratio of normalized area of 4 proteins between each endogenous and SIS peptides (B). All of the plots and graphs were generated by MedCalc software (version 12.2.1.0).



MSstats results

Uniprot ID	Gene symbol	Log2FC
P02749	CRP	0.85
P0C0L5	C4B	0.77
P05155	SERPING1	0.73
P19652	ORM2	0.65
P02679	FGG	0.58
P01019	AGT	0.56
P02753	RBP4	0.55
P05546	SERPIND1	0.43
Q96IY4	CPB2	0.43
M0QXZ3	C3	0.40
P02763	ORM1	0.40
P12259	F5	0.39
P0C0L4	C4A	0.39
P01011	SPERINA3	0.31
P01301	C5	0.27
P01023	A2M	0.26
P00738	HP	0.26
P02790	HPX	0.25
P05090	APOD	0.25
P00734	F2	0.23
P13671	C6	0.22
P02765	AHSG	0.20
P68871	HBB	0.16
P00747	PLG	0.16
P04114	APOB	0.15
P04003	C4BPA	0.14
P08603	CFH	0.13
P18065	IGFBP2	0.13
P04180	LCAT	0.07
P07357	C8A	0.07
P18428	LBP	0.07
P02652	APOA2	0.06

Uniprot ID	Gene symbol	log2FC
O95455	APOM	-0.01
P00450	AMBP	-0.02
P02760	HPR	-0.07
P00739	CP	-0.08
P00736	C1R	-0.09
P98160	APOH	-0.10
P02656	APOC3	-0.10
P02649	APOE	-0.12
P02776	PF4	-0.15
P02775	PPBP	-0.16
P01009	SERPINA1	-0.17
P10145	CXCL8	-0.19
P02746	C1QB	-0.23
P03952	KLKB1	-0.26
P0DJI8	SAA1	-1.20

Figure 12. AUC value of the constructed 4-multi marker panel under ROC curve from a training set and correlation with MSstats.

A stepwise logistic regression was applied for selecting proteins for multi-biomarker panel. The above step combined four proteins and set the equation in figure. The logistic regression process calculated prediction value of TCZ response from selected markers and drew ROC curve with the actual value as dependent. As shown above, 4-multi biomarker panel indicated much high AUC value (= 0.859) compared with that of single marker. The data was generated from SPSS software (version 25). The correlations between proteins included in multi-panel and MSstats were also considered.

Table. 5 Prediction rate table of 4-biomarker panel to discriminate the TCZ response and non-response group.

Observed		Predicted		
		ACR20		Percentage correct
		Response	Non-response	
ACR20	Response	23	3	88.52%
	Non-response	5	16	76.23%
Overall Percentage				83.00%

- Sensitivity= 82.14%, specificity= 84.21%

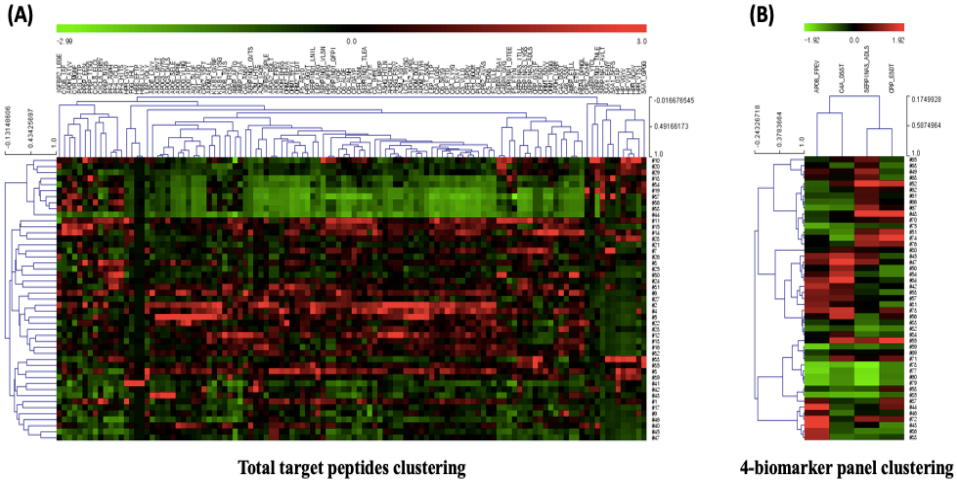


Figure 13. Hierarchical clustering analysis of 54 biomarker candidates and 4 selected biomarkers on multi-marker panel.

The hierarchical clustering analysis was expressed as heat map with colorimetric scheme by Multi Experiment Viewer (version 4.9). Figure 13A consisted with total of 54 biomarker candidates and figure 13B described only proteins in 4-biomarker panel. The x-axis represented the target peptides from MRM analysis and y-axis represented individual serum samples of RA patients in training set.

Table. 6 Absolute concentration of 4 proteins in multi-biomarker panel in two different groups.

4-multi biomarker panel	Average concentration (μg/mL)	
	Response	Non-response
CRP	0.525	2.119
SERPINA3	124.890	137.900
APOB	0.007	1.926
C4A	0.397	0.980

3.8 Quantitative performance verification of 4-biomarker panel in MRM analysis.

We verified the reasonable quantitative performance of the constructed 4-biomarker panel in complex serum matrix by generating quantitative response curve using 4 SIS peptides of those. The SIS peptides were serially diluted as 10 different concentrations (1, 2.5, 5, 10, 25, 50, 100, 250, 500, and 1000 fmol) and spiked into TCZ non-response serum samples to measure MRM signals. The quantitative linear calibration curves showed good regression values with 0.9803, 0.9965, 0.9989, and 0.8528 in FPEVDVLTK of APOB, ESDTSYVSLK of CRP, ADLSGITGAR of SERPINA3, and ANSFLGEK of C4A, respectively (Figure 14). All of the results were reproducible with CV lower than 20% in triplicate analysis. It demonstrated that the proteins included in 4-biomarker panel precisely measured abundances with a strong and reproducible quantitative manner in complex serum.

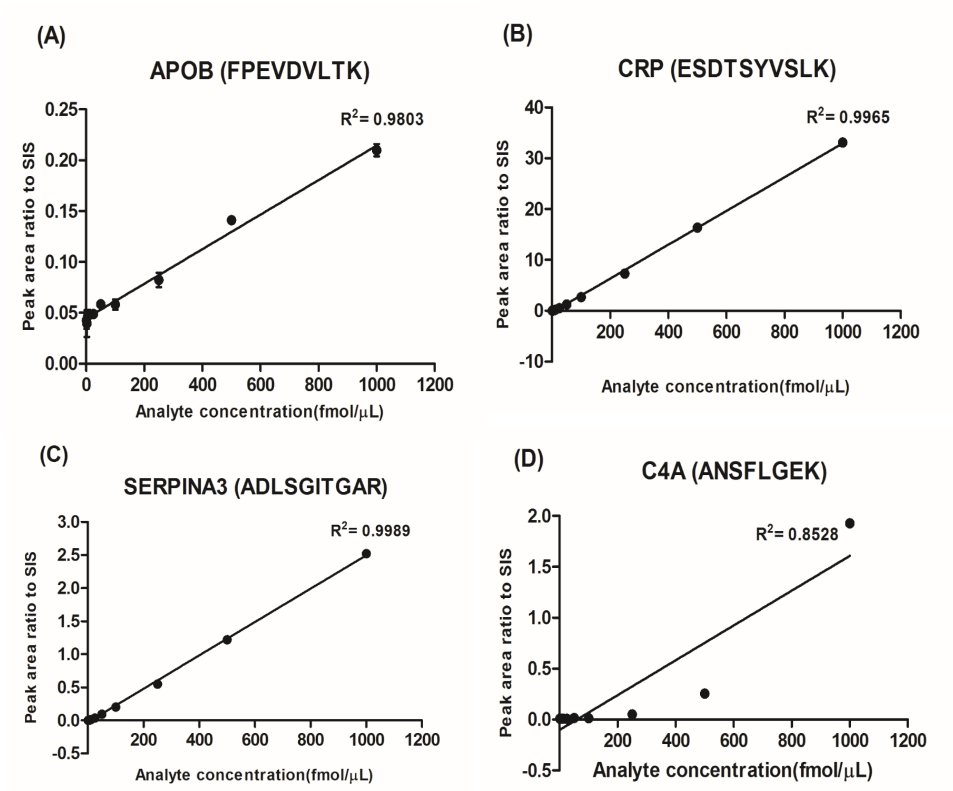


Figure 14. Quantitative calibration curves of 4 proteins in multi-biomarker panel by spiking in SIS peptides.

The MRM analysis was conducted in triplicate with 10 serially diluted concentrations; 1, 2.5, 5, 10, 25, 50, 100, 250, 500, 1000 fmol/μL of 4 SIS peptides (FPEVDVLTK, ESDTSYVSLK, ADLSGITGAR, and ANSFLGEK for APOB, CRP, SERPINA3, and C4A, respectively) spiked into TCZ non-response patient serum samples. The regression value of each curve showed 0.9803 for APOB (A), 0.9965 for CRP (B), 0.9989 for SERPINA3 (C), and 0.8528 for C4A (D).

4. Discussion

In the process of multi-biomarker construction for prediction of TCZ non-response in RA patients, we divided whole procedure as two parts with discovery and validation. Both of research applied quantitative proteomic approach by using MRM and PRM analysis. On discovery stage, we selected 54 initial biomarker candidates by integrative proteomic approach of pooled serum profiling analysis and including reported TCZ markers by literature data mining. The preliminary MRM and PRM analysis were conducted to confirm the detectability of 54 targets with spiked SIS peptides in Q-Exactive and QqQ coupled LC system. After AuDIT analysis to select peptides with credible and reproducible abundances, 116 and 124 peptides that corresponds to 47 and 50 target proteins, respectively and those were established as final biomarker candidates for validation. In addition, we investigate the correlations between two different quantification methods based on the results. PRM could detect 3 additional proteins including all of 47 proteins identified by MRM analysis. However, there were a few uniquely detected peptides depends on LC-MS devices. Moreover, specific mass spectrometry dependent properties such as sensitivity, selectivity, and reproducibility were evaluated. Although the

accurate limit of detection (LOD) or limit of quantification (LOQ) was not calculated in this study, we compared the sensitivity with linear graph of serially diluted SIS peptides. Both of methods showed eligible quantitative performance with high regression value but, it was a little high within MRM (R^2= 0.9754) than that of PRM (R^2= 0.9650) from 0.5 to 100 fmol concentrations. Based on the result that MRM assay shows a little high regression value within same concentration range, it indicated that MRM was slightly more sensitive than PRM analysis. We estimated that MRM process constituted with two selection filter, precursor ion selection at Q1 and product ion selection at Q3 allows sensitive and consistent peak area in wide range of protein concentrations [39]. The selectivity, another important property, were measured with STN between signal and background peak area on Skyline software. Among measured 10 peptides, most of STN from PRM demonstrated the higher quantities compared with that of MRM. The difference between selectivity is due to mass spectrometry devices. PRM analysis utilized Q-Exactive high-resolution orbitrap machine allows high selectivity within complexity of proteomic sample which is less disturbed by interferences in measurements, compared with relatively low-resolution machine in QqQ [40]. In case of reproducibility, both of measurements resulted stable abundances in

multiplicate analysis with lower than 20% CV. In specific, results of MRM were slightly more stable than those from PRM with lower than 13% average CV.

In validation process, we conducted the measurements of 47 final biomarker candidates from preliminary MRM analysis within 47 individual RA patient serum samples of TCZ responders (n= 26) and TCZ non-responders (n= 21) as a training set. The peak area ratio between endogenous and SIS peptides were applied to MSstats analysis and resulted with DEPs between two different groups. The peak area ratio also applied to stepwise regression analysis to select statistically reasonable proteins that could discriminate TCZ non-responders from responders. As a result, the 4 proteins were selected as final biomarkers and developed 4-biomarker panel comprising APOB, CRP, SERPINA3, and C4A. The constructed multi panel consisted with high AUC value of 0.859 and proved that the value is much higher than the single marker panels. Its prediction rates of discriminating two groups were also very high with 88.52% and 76.23% for TCZ response and non-response, respectively. Among those, APOB, CRP, and SERPINA3 in multi panel were statistically significant *p*-values lower than 0.05 between two different groups, but only *p*-value from C4A was slightly missed and close to being

statistically significant (*p*-value= 0.065). Our 4-biomarekr panel was related with RA and auto-immune or inflammatory diseases. Several studies in recent reported about the selected proteins in our 4-biomarker panel. One research indicated that immune cells from RA patients express more enolase-1 (ENO1) on surface compared with healthy subjects and APOB in synovial fluid of RA patients was identified as a specific ligand to ENO1. The identification of ENO1 ligand was performed with mass spectrometry and it also revealed that the production of proinflammatory cytokines and exaggerated arthritis severity [30]. A SERPINA3 was also reported as candidate urine biomarkers of lupus nephritis which is one of the autoimmune diseases. The paper uses unbiased proteomic approach with using isobaric tags for relative and absolute quantification (iTRAQ) analysis [31]. A CRP is already well-known markers and widely used in the clinical field as RA biomarkers as mentioned above [10, 23]. The high copy number of C4A, one of the complement proteins is reported that it confers the risk of Behçet disease which is classified as vasculitis and closely related rheumatologic disease [32, 33]. A common feature of Behçet syndrome is presented with an inflammatory arthritis and most studies stated to multi-system inflammatory disease as well [32]. In the process of multi-biomarker construction, we detected the limitations of

logit regression analysis. The process of logit regression selected discriminatory biomarkers among all of target proteins and generated new logit variables. The variables were saved as prediction value to produce ROC curve. However, the process selected unreasonable proteins that were not statistically significant or eliminated already known TCZ prediction markers. The main reason for the issue is due to high correlation coefficients between target proteins and also low sample size (n= 47) might affect unstable results. The proportion between control groups (TCZ responders, n= 26) and experimental groups (TCZ non-responders, n= 21) were limited numbers to confirm adequate statistical strength in this analysis. Therefore, we concentrated on technical factors of MRM analysis to confidentially measure the transition peak abundances. Moreover, we implemented the result from MSstats and sorted the target proteins with only up-regulated in TCZ non-responders before logit regression to increase the statistical confidence. As a result, 4 proteins were finally included in constructed multi marker panel and accomplished stable CV value less than 20% with a substantially high AUC value with 0.859. In advance, it is necessary to perform further analysis by applying our 4-biomarker panel into large patient cohorts as a test set and build multi marker panel by PRM analysis of individual sample as well. The further

research expected that the correlation between independently constructed multi marker panel from MRM and PRM assays would possible to compensate and improve the TCZ response prediction ability.

5. Conclusion

We constituted a quantitative MRM/PRM based serum biomarker identification process with highly accurate and reproducible manner that were possible to be utilized as a valuable tool in the biomarker validation mechanism as well as discovery process. In this study, we were able to develop the multi-marker panel for prediction of TCZ responders and non-responders in RA patients. Moreover, we evaluated and investigated the correlation between MRM and PRM assays. Our results proposed that there were uniquely detected target peptides in two different mass spectrometry and the properties such as sensitivity, selectivity, and reproducibility also showed differences. Therefore, two different quantitative analysis could compensate each other for better detection in target proteomics approach. Although the further analysis might be necessary for validation with much larger cohorts in MRM assay as test set and individual sample validations for PRM analysis, our 4-biomarker panel suggested that it was possible to contribute to identify the TCZ non-response patients who have been suffered in RA disease. In conclusion, we expected that multi-marker panel development platform based on

integrative proteomics approach, presents the precise predictions against specific therapeutic agents and could be widely applied in clinical fields.

6. References

[1] Lee, D. M. and M. E. Weinblatt (2001). "Rheumatoid arthritis." The Lancet 358(9285): 903-911.

[2] Firestein, G. S. (2003). "Evolving concepts of rheumatoid arthritis." Nature 423: 356.

[3] Srirangan, S. and E. H. Choy (2010). "The role of Interleukin 6 in the pathophysiology of rheumatoid arthritis." Therapeutic Advances in Musculoskeletal Disease 2(5): 247-256.

[4] Kim GW, Lee NR, Pi RH, Lim YS, Lee YM, Lee JM, et al. IL-6 inhibitors for treatment of rheumatoid arthritis: past, present, and future. Archives of Pharmacal Research. 2015;38(5):575-84.

[5] Al-Shakarchi I, Gullick NJ, Scott DL. Current perspectives on tocilizumab for the treatment of rheumatoid arthritis: a review. Patient preference and adherence. 2013; 7:653-66.

[6] Emery P, Keystone E, Tony HP, Cantagrel A, van Vollenhoven R, Sanchez A, et al. IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with rheumatoid arthritis refractory to anti-tumour necrosis factor biologicals: results from a 24-week multicentre randomised placebo-controlled trial. Annals of the Rheumatic Diseases. 2008;67(11):1516.

[7] Okuda Y. Review of tocilizumab in the treatment of rheumatoid arthritis. Biologics: Targets & Therapy. 2008;2(1):75-82.

[8] Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Molecular Systems Biology. 2008; 4:222-.

[9] Bourmaud A, Gallien S, Domon B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. PROTEOMICS. 2016;16(15-16):2146-59.

[10] Colombet I, Pouchot J, Kronz V, Hanras X, Capron L, Durieux P, et al. Agreement between Erythrocyte Sedimentation Rate and C-Reactive Protein in Hospital Practice. The American Journal of Medicine. 2010;123(9): 863.e7-.e13.

[11] Pavelka N, Pelizzola M, Vizzardelli C, Capozzoli M, Splendiani A, Granucci F, et al. A power law global error model for the identification of differentially expressed genes in microarray data. BMC Bioinformatics. 2004;5(1):203.

[12] Pavelka N, Fournier ML, Swanson SK, Pelizzola M, Ricciardi-Castagnoli P, Florens L, et al. Statistical Similarities between Transcriptomics and Quantitative Shotgun Proteomics Data. Molecular & Cellular Proteomics. 2008;7(4):631.

[13] Liebler DC, Zimmerman LJ. Targeted Quantitation of Proteins by Mass Spectrometry. Biochemistry. 2013;52(22):3797-806.

[14] Abbatiello SE, Mani DR, Keshishian H, Carr SA. Automated Detection of Inaccurate and Imprecise Transitions in Peptide Quantification by Multiple Reaction Monitoring Mass Spectrometry. Clinical Chemistry. 2010;56(2):291.

[15] Choi M, Chang C-Y, Clough T, Broudy D, Killeen T, MacLean B, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics. 2014;30(17):2524-6

[16] Wang J-R, Gao W-N, Grimm R, Jiang S, Liang Y, Ye H, et al. A

method to identify trace sulfated IgG N-glycans as biomarkers for rheumatoid arthritis. Nature Communications. 2017;8(1):631.

[17] LeBlanc A, Michaud SA, Percy AJ, Hardie DB, Yang J, Sinclair NJ, et al. Multiplexed MRM-Based Protein Quantitation Using Two Different Stable Isotope-Labeled Peptide Isotopologues for Calibration. Journal of Proteome Research. 2017;16(7):2527-36.

[18] Izumi K, Kaneko Y, Hashizume M, Yoshimoto K, Takeuchi T. Baseline Serum Osteopontin Levels Predict the Clinical Effectiveness of Tocilizumab but Not Infliximab in Biologic-Naïve Patients with Rheumatoid Arthritis: A Single-Center Prospective Study at 1 Year (the Keio First-Bio Cohort Study). PLOS ONE. 2015;10(12): e0145468.

[19] Kasama T, Isojima S, Umemura M, Tsukamoto H, Tokunaga T, Furuya H, et al. Serum macrophage migration inhibitory factor levels are correlated with response to tocilizumab therapy in patients with rheumatoid arthritis. Rheumatology International. 2014;34(3):429-33.

[20] Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Research. 2016;44(D1): D447-D56.

[21] Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, et al. A High-Confidence Human Plasma Proteome Reference Set with Estimated Concentrations in PeptideAtlas. Molecular & Cellular Proteomics. 2011;10(9):M110.006353.

[22] Siderius, D.W., Shen, V.K., Johnson III, R.D., and van Zee, R.D., eds. (2013), NIST/ARPA-E Database of Novel and

Emerging Adsorbent Materials, NIST Standard Reference Database 205, National Institute of Standards and Technology.

[23] Kim K-W, Kim B-M, Moon H-W, Lee S-H, Kim H-R. Role of C-reactive protein in osteoclastogenesis in rheumatoid arthritis. Arthritis Research & Therapy. 2015;17(1):41.

[24] Tan W, Wang F, Guo D, Ke Y, Shen Y, Lv C, et al. High serum level of haptoglobin is associated with the response of 12 weeks methotrexate therapy in recent-onset rheumatoid arthritis patients. International Journal of Rheumatic Diseases. 2014;19(5):482-9

[25] Finnegan S, Robson J, Scaife C, McAllister C, Pennington SR, Gibson DS, et al. Synovial membrane protein expression differs between juvenile idiopathic arthritis subtypes in early disease. Arthritis research & therapy. 2014;16(1): R8-R.

[26] Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nature Biotechnology. 2006; 24:971.

[27] Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Molecular & Cellular Proteomics. 2012;11(6).

[28] Henderson CM, Shulman NJ, MacLean B, MacCoss MJ, Hoofnagle AN. Skyline Performs as Well as Vendor Software in the Quantitative Analysis of Serum 25-Hydroxy Vitamin D and Vitamin D Binding Globulin. Clinical Chemistry. 2018;64(2):408.

[29] Hanne Bjerre Christensen, Mette Erecius Poulsen, Peter Have Rasmussen, Danilo Christen. Development of an LC-MS/MS method for the determination of pesticides and patulin in apples. Food Additives and Contaminants, 2009, 26 (07), pp.1013-1023.

[30] Lee JY, Kang MJ, Choi JY, Park JS, Park JK, Lee EY, et al. Apolipoprotein B binds to enolase-1 and aggravates inflammation in rheumatoid arthritis. Annals of the Rheumatic Diseases. 2018;77(10):1480.

[31] Turnier JL, Brunner HI, Bennett M, Aleed A, Gulati G, Haffey WD, et al. Discovery of SERPINA3 as a candidate urinary biomarker of lupus nephritis activity. Rheumatology. 2018:key301-key.

[32] Barnes CG, Yazici H. Behçet's syndrome. Rheumatology. 1999;38(12):1171-4.

[33] Li Z-g. A new look at rheumatology in China—opportunities and challenges. Nature Reviews Rheumatology. 2015;11:313.

[34] Ozcan S, Cooper JD, Lago SG, Kenny D, Rustogi N, Stocki P, et al. Towards reproducible MRM based biomarker discovery using dried blood spots. Scientific Reports. 2017;7:45178.

[35] Kusebauch U, Campbell DS, Deutsch EW, Chu CS, Spicer DA, Brusniak M-Y, et al. Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. Cell. 2016;166(3):766-78.

[36] Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. Nucleic Acids Research. 2006;34(suppl_1):D655-D8.

[37] Whiteaker JR et al.,(2014) Data used in this publication were generated by the National Cancer Institute, Clinical Proteomic Tumor Analysis Consortium (CPTAC) Assay Portal (assays.cancer.gov). Nat Methods Jul;11(7):703-4. doi: 10.1038/nmeth.3002)

[38] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research. 2015;43(D1):D447-D52.

[39] Shi T, Song E, Nie S, Rodland KD, Liu T, Qian W-J, et al. Advances in targeted proteomics and applications to biomedical research. Proteomics. 2016;16(15-16):2160-82.

[40] Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. Molecular & Cellular Proteomics. 2012;11(11):1475.

[41] Santangelo A, Imbrucè P, Gardenghi B, Belli L, Agushi R, Tamanini A, et al. A microRNA signature from serum exosomes of patients with glioma as complementary diagnostic biomarker2017. 1-12 p.

7. 국문 초록

류마티스 관절염(RA)은 관절 조직의 염증을 유발하는 만성 전신성 자가 면역 질환이다. RA 는 주로 복잡한 시토카인 네트워크의 불균형이 원인으로 알려져 있고, 이 중 염증성 시토카인 중 하나인 인터루킨-6 (IL-6)는 국소적 윤활막 부위의 백혈구 활성화와 만성 염증 유도에 중요한 역할을 하고 있다. 따라서 IL-6 는 RA 환자들에 대해 중요한 치료 표적이며, IL-6 수용체 항체인 tocilizumab (TCZ)는 RA 환자들에게 높은 치료효과를 보인다. 그러나 여전히 일부 RA 환자들에서 해당 항체에 부분적으로 반응하거나 저항 반응을 나타내고 있다. 본 연구에서는, 통합 프로테오믹스 분석법을 기반으로 RA 환자들의 TCZ 반응과 비 반응을 예측하는 혈청 바이오마커 발굴에 목적을 두고 있다. 1 차적인 바이오마커 후보군은 TCZ 처리 전 RA 환자 혈청 시료를 수집하고 TCZ 처리 12 주 후 반응과 비 반응을 보이는 시료에 대한 프로테오믹스 정성 분석과 이미 보고된 TCZ 특이적 바이오마커에 대한 문헌 조사를 통해 54 개의 후보를 선정하였다. 선택된 후보군을 대상으로 선행 다중반응탐지법(MRM)과 동시반응탐지법(PRM)을 통해 최종적으로 각각 47 개와 50 개의 혈액 바이오마커 후보군을 선정하였다. 해당 결과를 기반으로 다중 바이오마커 패널 구축을 위해 47 명의 RA 환자들을 training set 으로 선정하였고, 이를 치료 반응군 (n= 26)과 비 반응군(n= 21)으로 분류하여 다중반응탑지법 분석을 시행하였다. 정량 분석 결과를 기반으로

통계 분석을 수행한 결과 최종적으로 4 개의 단백질(APOB, CRP, SERPINA3, 그리고 C4A)이 다중 바이오마커 패널에 포함되었다. 구축된 다중 바이오마커 패널은 개별 단백질의 AUC 값보다 상대적으로 높은 0.859 의 AUC 값을 보였고 개별 시료들 간의 높은 예측도(TCZ 반응군= 88.52%, TCZ 비 반응군= 76.23%)를 보였다. 결론적으로, 본 연구를 통해 구축된 4 개의 단백질들이 RA 환자들 중 TCZ 에 대해 저항 반응을 예측하는 지표로 사용 될 수 있으며, 동시반응탐지법과 다중반응탐지법 기반의 혈액 바이오마커 발굴 프로세스 또한 임상적인 검증 과정에서 유용하게 활용될 수 있을 것이라고 사료된다.

주요어: 류마티스성 관절염, Tocilizumab, 인터루킨-6, 프로테오믹스, 바이오마커, 다중반응탐지법, 동시반응탐지법.

학번: 2017-24859