Plasma fibronectin levels identified via quantitative proteomics profiling predicts hepatitis B surface antigen seroclearance in chronic hepatitis B

Fen Liu[1], Wai-Kay Seto[1,2,3*], Danny Ka-Ho Wong[1,2], Fung-Yu Huang[1], Ka-Shing Cheung[1], Lung-Yi Mak[1], Rakesh Sharma[4], Saisai Zhang[1], James Fung[1,2], Ching-Lung Lai[1,2], Man-Fung Yuen[1,2*]

[1]Department of Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong

[2]State Key Laboratory for Liver Research, The University of Hong Kong, Hong Kong

[3]Department of Medicine, The University of Hong Kong-Shenzhen Hospital, Shenzhen, China

[4]Proteomics and Metabolomics Core Facility, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

Co-Corresponding authors:

Dr Wai-Kay Seto

Postal address: Department of Medicine, Queen Mary Hospital, The University of Hong Kong, Pokfulam Road 102, Hong Kong

Email address: wkseto@hku.hk

Fax number: +852 28162863

Phone number: +852 22553994

Prof. Man-Fung Yuen

Postal address: Department of Medicine, Queen Mary Hospital, The University of Hong Kong, Pokfulam Road 102, Hong Kong

Email address: mfyuen@hkucc.hku.hk

Fax number: +852 28162863

Phone number: +852 22553984

Main points of the article

This is the first study to reveal the plasma proteomic profiling of chronic hepatitis B patients with and without HBsAg seroclearance. Plasma fibronectin may be associated with HBsAg seroclearance, and potentially a predictor of "functional cure".

Abstract

Background: Seroclearance of hepatitis B surface antigen (HBsAg) is a potentially achievable target of chronic hepatitis B (CHB). Plasma proteins relevant to HBsAg seroclearance remain undetermined.

Methods: We prospectively recruited treatment-naïve CHB patients with spontaneous HBsAg seroclearance and matched HBsAg-positive controls. Plasma protein profiling was performed using isobaric tags for relative and absolute quantitation (iTRAQ)-based proteomics, with the expression of candidate proteins validated in a separate cohort. The predictive value of fibronectin was assessed at 3 years, 1 year (Year -1) before and at the time (Year 0) of HBsAg seroclearance.

Results: 487 plasma proteins were identified via proteomics, with 97 proteins showing altered expression. In verification cohort (n=90), median plasma fibronectin levels in patients with HBsAg seroclearance was higher than in controls (p=0.009). In the longitudinal cohort (n=164), patients with HBsAg seroclearance, when compared to controls, had a higher median fibronectin levels at Year -1 (413.26 vs. 227.95 μg/ml), and Year 0 (349.45 vs. 208.72 μg/ml) (both p<0.001). In patients with an annual HBsAg log reduction >0.5, Year -1 fibronectin level achieved an AUROC of 0.884 in predicting HBsAg seroclearance.

Conclusions: Using proteomics-based technology, plasma fibronectin may be associated with HBsAg seroclearance, and a potential predictor of “functional cure”.

Keywords: biomarker; functional cure; HBV; HBsAg; iTRAQ

Background

Seroclearance of hepatitis B surface antigen (HBsAg) is now regarded as a potentially achievable treatment endpoint of chronic hepatitis B (CHB). Achievement of this target indicates a successful immune control of infection and reflects a "functional cure", with a reduction in liver-related complications[1, 2]. Spontaneous HBsAg seroclearance is uncommon, as its annual rate was only reported in 0.7%[3] and 2.26%[4] according to two studies with average follow-up of 19.6 and 8.04 years, respectively. As the approved antiviral therapies have little effect on the seroclearance of HBsAg[5], exploring the underlying molecular mechanism of HBsAg seroclearance is imperative.

The kinetics of serum HBsAg and HBV DNA levels were previously found to be associated with HBsAg seroclearance[6, 7]. Other associated factors included interferon-inducible protein 10[8], genetic polymorphisms near the HLA-DP regions or interleukin coding genes[9, 10], microRNA-125a-5p[11] and microRNA-581[12], In vivo, sT123N substitution was also shown to accelerate HBsAg seroclearance[13]. Despite all these findings, our current understanding of the clinicopathological mechanisms leading to HBsAg seroclearance remains limited. It is still unknown if there are any "effector proteins" that can promote HBsAg seroclearance.

Proteomics refers to large-scale study to unveil a protein profiling of what is expressed and presented under different biological conditions[14], and may reveal actual metabolic activities and regulatory cascades[15]. Isobaric tags for relative and absolute quantification (iTRAQ) is currently one of the most robust quantitative proteomics techniques[16], allowing identification and quantification of protein abundances simultaneously through peptide labeling. With the availability of

advanced mass spectrometry (MS) technology[17] and bioinformatics tools[18], the application of this technology is expanding in various aspects of clinical research, including cancers[19], viral infections[20] and metabolic diseases[21].

We hypothesized that HBsAg seroclearance may be associated with distinct alterations in plasma protein profiles. We hence aimed to apply iTRAQ-based quantitative proteomics to reveal protein profiles associated with HBsAg seroclearance, followed by validation and prediction of identified proteins in separate patient cohorts, comprising patients achieving spontaneous HBsAg seroclearance and matched controls.

Methods

Patients and sample collection

The present study was designed in three stages: proteomic analysis, followed by prospective validation in a separate cohort of patients, then finally an assessment of predictive values of identified plasma proteins for HBsAg seroclearance using a historical cohort with serially archived plasma samples from 3 years prior to HBsAg seroclearance till HBsAg seroclearance[7] (Figure 1).

For the first two stages, we recruited treatment-naïve CHB patients, all documented with prior HBsAg-positivity for at least 6 months, who were regularly followed up at the Liver Clinic, Department of Medicine, Queen Mary Hospital, Hong Kong. We recruited two groups of patients based on their HBsAg serologic status: spontaneous HBsAg seroclearance (experimental group) occurring between June 2012 and July 2016, and persistent HBsAg-positivity (control group). HBsAg seroclearance was defined as serum HBsAg-negativity on two occasions at least 6 months apart and

remaining undetectable up to the last visit. The two groups were matched for age and gender, with normal serum alanine aminotransferase (ALT) and undetectable serum HBV DNA. Patients with concomitant liver diseases, including chronic hepatitis C and D infection, Wilson's disease, autoimmune hepatitis, primary biliary cholangitis and primary sclerosing cholangitis, and patients with increased alcohol intake (30 grams per day for men, 20 grams per day for women) were excluded. We also excluded patients with history of antiviral therapy, and presence of cirrhosis or hepatocellular carcinoma. For the experimental group, we additionally excluded patients with a positive antibody to HBsAg (anti-HBs) as to avoid its interference with proteomic analysis.

Plasma collection and preparation were based on standard criteria established by Early Detection Research Network[22]. Briefly, fresh blood samples were collected in EDTA tubes, and then centrifuged at 1200 g for 10 minutes within two hours of collection at room temperature. Plasma was apportioned into 0.5 ml aliquots, then put into -80°C fridge immediately. The storage time of all samples prior to usage was less than two months. The detailed protocol is described in the Supplementary Appendix.

The present study was approved by the Institutional Review Board/Ethics Committee of the University of Hong Kong and the Hong Kong West Cluster of Hospital Authority, and conducted according to the Declaration of Helsinki.

Stage I: iTRAQ coupled with liquid chromatography (LC)-tandem MS analysis (MS/MS) and protein identification

Proteins preparation and iTRAQ labeling

Plasma samples in each group were pooled and high abundant proteins were immunodepleted using Human 14 Multiple Affinity Removal System Column (Agilent Technologies, Santa Clara, CA). Unbound depleted plasma was desalted. The total protein amount was quantitated by Pierce™ BCA assay (Thermo Fisher Scientific). Equal amount of proteins (200 ug) from each group were labeled with iTRAQ reagents according to the manufacturer's instructions (Applied Biosystems, Foster City, CA, USA). Briefly, proteins were subjected to denaturation, reduction and alkylation, following by digesting with trypsin. The tryptic peptides from each group were divided into two equal halves. Peptides from control samples were labeled with reagents containing 114 and 115 reporter tags, and peptides from HBsAg seroclearance samples were labeled with reporter tags 116 and 117, respectively. The methodological details are described in the Supplementary Appendix.

Peptide fractionation and LC-MS/MS analysis

Labeled peptides were fractionated using a strong cation exchange (SCX) column (5 μm, 200 Å, 2.1× 200mm; PolyLC Inc, Columbia, MD, USA). The Waters Acquity Ultra-Performance LC system was coupled with LTQ-Orbitrap Fusion Tribid Lumos mass spectrometer (Thermo Fisher Scientific) for electrospray analysis. Survey scans were acquired from 150 to 2000 with up to 20 most intense precursor ions selected for MS/MS. All measurements were performed in duplicate.

Protein identification and relative quantification

Proteome Discoverer (Version 2.1, Thermo Fisher Scientific) was used to analyze the MS/MS data. The data was searched using the SEQUEST algorithm against the NCBI Human RefSeq database 71 containing 35,985 proteins

(https://www.ncbi.nlm.nih.gov/refseq/). The precursor and fragment mass tolerance were set to 20 ppm and 0.1 Da, respectively. Peptide and protein data were fetched using high peptide confidence (<1% false discovery rate) and top one peptide rank filters. The intensities of the reporter ions of peptides generated from MS/MS were used to determine the relative quantitation of proteins. The false changes ratio < 0.8 or >1.2 were considered as the cutoff value to distinguish the down-regulated or up-regulated changes of protein expression[23, 24].

Bioinformatics analysis

Bioinformatics analysis and annotations of the differentially expressed proteins were performed with multistep methods. First, Geno Ontology (GO) enrichment analysis[25] was used to classify their functions including molecular function, cellular component, and biological process. Second, the signaling pathway, disease function, network, together with regulator effects were analyzed using Ingenuity® Pathway Knowledge (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). IPA® is a powerful web-based search tool that enables analysis, integration and interpretation of huge amounts of data generated from omics experiments. Consistency scores of different regulator effects, i.e. how predicted upstream regulators might alter phenotypic or functional outcomes downstream, were calculated, with a score >20 signifying most of the predicted signaling cascades[26]. Third, deeper annotations of these proteins were complemented by referencing to the Human Protein Reference Database (HPRD) and UniProtKB database. These two databases contain comprehensive protein information and describe multiple features of proteins such as subcellular localization, post-translational modifications and protein-protein

interactions. Proteins associated with the most relevant biological function were selected in the subsequent study.

Stage II: Validation via ELISA

Plasma expression of identified proteins from Stage I was measured with commercially available ELISA kits from Abcam or US Biological companies. All ELISA measurements were performed in duplicates according to the manufacturer's instructions.

Briefly, based on the detection range, samples were diluted with the reagent supplied by ELISA kits. The diluted samples were firstly incubated in 96-well microplates. After incubation and washing, the biotinylated antibody and streptavidin peroxidase conjugate was added, respectively. After a second incubation and washing, a chromogen substrate was added till the appearance of optimal blue color density. Stop solution was then added to terminate the reaction. A microplate spectrophotometer was used to read the absorbance. The data were imputed into the CurveExpert professional 2.3.0 software. The concentration of proteins was calculated with a four-parameter logistic curve and multiplied by the dilution factor.

Stage III: Assessment of predictive values using a historical cohort

We retrieved available archived plasma samples of patients achieving HBsAg seroclearance at 3 time points: 3 years prior to, 1 year prior to and at the time of HBsAg seroclearance[7]. Corresponding samples of matched HBsAg-positive controls were also retrieved. Plasma levels of proteins validated from Stage II were measured for all three time points. Our previous study has demonstrated that the optimal cut-off HBsAg level and HBsAg reduction to predict HBsAg seroclearance

are <200 IU/ml and >0.5 log IU/ml/year, respectively[7]. In the present study, we therefore further investigated whether the addition of plasma fibronectin to HBsAg levels or HBsAg reduction could improve the predictive value of HBsAg seroclearance.

Laboratory tests

Serum HBsAg levels were performed using Elecsys HBsAg II assay (Roche Diagnostics, Gmbh, Mannheim), with a linear range of 0.05 IU/ml. Serum HBeAg, antibody to HBeAg (anti-HBe) and anti-HBs were measured by Abbott Laboratories (Chicago, IL, USA). Serum HBV DNA levels were measured using Cobas Taqman assay (Roche Diagnostics, Branchburg, NJ, USA) with the lower limit of detection of 20 IU/ml.

Statistics analysis

Continuous variables were expressed in mean (± standard) or median (interquartile range, IQR). Serum HBsAg and HBV DNA levels were expressed in logarithm. The continuous variables were compared using Student's t-test or Wilcoxon matched-pairs singed rank test as appropriate. Comparisons of categorical variables were analyzed using Pearson's chi-squared test or Fisher's exact test as appropriate. Receiver operating characteristic (ROC) curves were constructed to predict HBsAg seroclearance, with the optimal cut-off protein level for prediction of HBsAg seroclearance determined using the Youden's index. A two-sided P value of <0.05 was considered statistically significant. Statistical analyses were performed using SPSS version 24.0 (SPSS Inc, Chicago, IL, USA) or MedCalc Software Version 18.10.2.

Results

Stage I: Proteomics analysis

Differentially expressed proteins identified via iTRAQ-based proteomics technology

The clinical characteristics of 20 patients achieving HBsAg seroclearance and 20 matched controls are depicted in Supplementary Table 1. The mean age was 53.2 ± 7.8 years (65% male). For the experimental group, the median duration from the time of HBsAg seroclearance to sample collection was 2.2 (1.51-2.65) years. The median HBsAg level in the control group was 2.67 (1.58-2.97) log IU/ml. There were no significant differences between the two groups with respect to other clinical characteristics and liver parameters ($p>0.05$).

A total of 487 plasma proteins were identified (Supplementary Table 2A), of which 97 proteins showed altered expression. Among them, 88 proteins were found to be upregulated (fold change >1.2), and 9 proteins down-regulated (fold change <0.8) (Supplementary Table 2B). All differentially expressed proteins had at least one unique peptide; 45.36 % (45/97) proteins had two more unique peptides. Six proteins were identified with more than 10 unique peptides including fibronectin, alpha-2-macroglobulin, talin-1, apolipoprotein E, epididymis luminal protein 114, and kallikrein.

Bioinformatics analysis

In GO enrichment analysis, the top two biological processes were cellular processes (27.3%) and immune responses (19.0%) (Figure 2A). Approximately one-third of these proteins (34.6%) were annotated as cell compartments (Supplementary Figure

1A). The top two molecular functions were catalytic activity and binding, and both accounting for 37% (Supplementary Figure 1B).

IPA® was performed to retrieve the biological significance of these differentially expressed proteins (Supplementary Figure 1C). The top disease function was immunological disease, achieving a p value of 1.08E-15 via IPA® (Figure 2B and Supplementary Table 3A). The top regulator effect achieved a highest consistency score of 58.69. The detailed information of canonical pathways and networks analysis are provided in Supplementary Table 3B and 3C.

The retrieved information from HPRD and UniProtKB databases is presented in the Supplementary Table 2B.

The four proteins that were most associated with immunological response were fibronectin[27-29], CD44[30-32], aldolase A (ALODA)[33], and S100 calcium-binding protein A9 (S100A9)[34, 35] and were finally identified and subjected to the next Stage II study. Their representative MS/MS spectra with reporter ions of peptides are shown in Figure 3.

Stage II: Validating plasma expression of the four candidate proteins via ELISA

The clinical characteristics of 45 CHB patients achieving HBsAg seroclearance and 45 matched controls involved in the prospective validation are depicted in Table 1. The median duration of follow-up after HBsAg seroclearance was 1.67 (0.94-2.23) years. In the control group, median HBsAg level was 2.84 (2.29 -3.24) log IU/ml. The duration of follow-up and HBeAg seroconversion were comparable between the two groups (both $p>0.05$). Liver biochemistries in the two groups were also similar ($p>0.05$).

Validation ELISA showed median fibronectin level in patients achieving HBsAg seroclearance was significantly higher than that of HBsAg-positive controls (188.82 ug/ml vs. 166.36 ug/ml, p=0.009) (Figure 4A). There was no significant difference in the median levels of CD44, ALODA, and S100A9 between the two validation groups (Figures 4B-D, all p>0.05).

Stage III: Assessment of fibronectin's predictive value in HBsAg seroclearance

One hundred and sixty-four patients (82 patients achieving HBsAg seroclearance and 82 age and gender-matched HBsAg-positive controls) had available archived plasma at three years (Year -3), one year (Year -1) prior to HBsAg seroclearance and at the time of HBsAg seroclearance (Year 0). At Year 0, the mean age of patients achieving HBsAg seroclearance was 51.48±9.98 years (66.2% male); sixty-eight patients (82.9%) had undetectable HBV DNA (<20 IU/mL) and 10 (12.2 %) patients had developed anti-HBs. Median HBsAg and HBV DNA levels in the control group were 2.27 (1.57-3.28) log IU/ml, and 3.34 (2.25-4.30) log IU/ml, respectively.

At Year 0, median plasma fibronectin in patients achieving HBsAg seroclearance was significantly higher than that of controls (349.45 µg/ml vs. 208.72 µg/ml, p <0.001) (Figure 5). A similar pattern was noted at Year -1 (413.26 µg/ml vs. 227.95 µg/ml, p <0.001). There was no significant difference in the plasma levels of fibronectin between the two groups at Year -3 (432.79 µg/ml vs. 400.28 µg/ml, p=0.124). In the HBsAg seroclearance group, there was no difference in median fibronectin levels across all three time points (p=0.667), while in the control group a significant difference was observed (p<0.001) (Supplementary Figure 2).

The predictability of plasma fibronectin for HBsAg seroclearance is depicted in Table 2. The plasma level of fibronectin at Year -1 achieved a modest value to predict HBsAg seroclearance, with an area under the receiving operator characteristic (AUROC) of 0.669 ($p<0.0001$), while fibronectin at Year -3 had a weak AUROC of 0.567 ($p=0.1353$). Analyzing patients with the optimal serum HBsAg levels < 200 IU/ml and HBsAg log reduction >0.5, their AUROC were 0.663 and 0.777, respectively (both $p<0.05$). In the subgroup of patients with HBsAg level < 200 IU/ml (N=87), the AUROC of plasma fibronectin at Year -1 increased to 0.730 (95% confidence interval [CI]: 0.624-819; $p=0.0003$). In the subgroup patients with annual HBsAg log reduction >0.5 (N=60), the addition of plasma fibronectin level at Year -1 achieved the highest predictive value, with the AUROC of 0.884 (CI: 0.775-952; $p=0.0003$). The optimal cut-off fibronectin level was 122.03 μg/ml, with the sensitivity of 98.2% and specificity of 75.0%. The positive predictive value (PPV) reached to 98.2%, while the negative predictive value (NPV) was 75.0%.

Discussion

Studies of proteomics have been increasingly used for biomarker identification in different disease entities. Examples include establishing a panel of serum biomarkers for gynecological malignancies[19], and assessing the treatment response of lung cancer during tyrosine kinase therapy[36]. In the present study, we employed iTRAQ based proteomics technology to reveal the plasma proteomic profiling of CHB patients with and without HBsAg seroclearance and identified a related plasma biomarker. To our knowledge, this was the first study of this kind performed in this disease entity.

To ensure that the proteins identified were highly associated with HBsAg, we established stringent matching criteria of enrolling patients during proteomic analysis. Liver biochemistry and HBV DNA levels were well-matched. Achieving a good sample quality and tackling the wide dynamic range of protein concentration are challenges in the study of proteomics[37]. To circumvent these issues, the procedures of sample processing and preservation adhered strictly to standard procedures[22], ensuring minimal sample contamination and degradation. The most abundantly-found proteins (e.g. albumin, different immunoglobulins) had been depleted for enriching low-abundance proteins[38]. Also, pre-fractionations of peptides was performed to reduce the dynamic range and enhance the coverage of protein identification[39].

Through validation in both a prospective and historical cohort, our present study demonstrated plasma fibronectin can potentially act as a biomarker for predicting HBsAg seroclearance. Fibronectin may be involved in both innate and adaptive immune cell function via toll like receptor 4 (TLR4), enhancing HBV-specific T cells and B cells responses[40], with associated cytokines correlating with serum HBsAg levels[41]. In vitro and in vivo experimental studies have demonstrated that activation of TLR system can initiate antiviral mechanisms via the production of type I interferons[42], pro-inflammatory cytokines[43] and intracellular antiviral pathways[44], resulting in inhibition of HBV replication. As a marker of immune stimulation, this can explain the sustained high levels of fibronectin in patients achieving HBsAg seroclearance, as well as the significant decline in fibronectin levels among patients persistently HBsAg-positive. Further research will be required to validate mechanistically this linkage of fibronectin to HBsAg seroclearance.

The potential future use of fibronectin will be for use in combination with other established viral markers, e.g. quantitative HBsAg levels, which is limited by a low PPV[7]. Among patients achieving a certain degree of HBsAg reduction (e.g. annual decline >0.5 log), fibronectin was able to differentiate patients with and without HBsAg seroclearance, with a sensitivity of 98.2% and specificity of 75.0%. The high PPV of 98.2% signified fibronectin, together with moderate HBsAg reduction, can identify a majority of HBsAg seroclearance patients one year prior to HBsAg seroclearance. If similar results are validated in patients on long-term nucleoside analogue therapy, plasma fibronectin may aid in identifying patients with the possibility of achieving the "functional cure" of HBsAg seroclearance and eventual cessation of therapy.

The merits of this study included the application of advanced technology and strict matching criteria, which facilitated to identification of subtle changes of protein abundance. However, our present study had certain limitations. Although several clinical parameters had been matched between experimental and control groups, other variables such as viral mutants and genotypes may have confounding effects. The predictability of fibronectin was only evident one year prior to HBsAg seroclearance and in patients with a substantial serum HBsAg level decline. Our study results did not provide a true mechanistic link between fibronectin and HBsAg seroclearance, and further studies investigating this linkage will be required. Nonetheless, our present study displayed the real-world "end result" of plasma protein profiles that were associated with HBsAg seroclearance, and identified a potential biomarker for the "functional cure" of CHB.

In conclusion, plasma fibronectin level was identified to be associated with HBsAg seroclearance, and could be a potential predictor of "functional cure". Future studies should explore the mechanistic nature between fibronectin and HBsAg seroclearance, as well as its application in clinical prognostication of CHB.

Acknowledgement: The authors will like to thank the kind assistance of the Proteomics and Metabolomics Core Facility, Li Ka Shing Faculty of Medicine Faculty, The University of Hong Kong.

Conflicts of interest: WK Seto received speaker fees from Alfa-Wassermann, Astrazeneca, is an advisory board member of Celltrion, received speaker fees and is an advisory board member of AbbVie, and received speaker fees, research funding and is an advisory board member of Gilead Sciences. MF Yuen is an advisory board member and received speaker's fees from AbbVie, Janssen, Biocartis NV, Bristol Myers Squibb, Fujirebio Incorporation, Gilead Sciences, Merck Sharp and Dohme, and Sysmex Corporation, and received research funding from Bristol Myers Squibb and Gilead Sciences.

Study Funding: This work was supported by Seed Fund for Basic Research and Incentive Scheme for Grant Application, Department of Medicine, The University of Hong Kong.

Author contributions: F Liu were involved in study concept and design, data acquisition, performing of laboratory measurements, analysis and interpretation of data, and drafting of the manuscript. DK Wong was involved in study concept and design, data acquisition and performing of laboratory measurements. FY Huang was

involved in data acquisition and performing of laboratory measurements. KS Cheung, LY Mak, SS Zhang and J Fung were involved in data acquisition. Rakesh Sharma was involved in performing of laboratory measurements. CL Lai was involved in critical revision of the manuscript. WK Seto and MF Yuen were involved in study concept and design, analysis and interpretation of data, critical revision of manuscript, and overall study supervision. All authors have approved the final version of the manuscript.

Dr. WK Seto is the article's guarantor.

References

1. Yuen MF, Wong DK, Fung J, et al. HBsAg Seroclearance in chronic hepatitis B in Asian patients: replicative level and risk of hepatocellular carcinoma. Gastroenterology **2008**; 135:1192-9.

2. Revill P, Testoni B, Locarnini S, Zoulim F. Global strategies are required to cure and eliminate HBV infection. Nature reviews Gastroenterology & hepatology **2016**; 13:239-48.

3. Simonetti J, Bulkow L, McMahon BJ, et al. Clearance of hepatitis B surface antigen and risk of hepatocellular carcinoma in a cohort chronically infected with hepatitis B virus. Hepatology **2010**; 51:1531-7.

4. Liu J, Yang HI, Lee MH, et al. Incidence and determinants of spontaneous hepatitis B surface antigen seroclearance: a community-based follow-up study. Gastroenterology **2010**; 139:474-82.

5. Kim GA, Lim YS, An J, et al. HBsAg seroclearance after nucleoside analogue therapy in patients with chronic hepatitis B: clinical outcomes and durability. Gut **2014**; 63:1325-32.

6. Tseng TC, Liu CJ, Su TH, et al. Serum hepatitis B surface antigen levels predict surface antigen loss in hepatitis B e antigen seroconverters. Gastroenterology **2011**; 141:517-25, 25.e1-2.

7. Seto WK, Wong DK, Fung J, et al. A large case-control study on the predictability of hepatitis B surface antigen levels three years before hepatitis B surface antigen seroclearance. Hepatology (Baltimore, Md) **2012**; 56:812-9.

8. Wong GL, Chan HL, Chan HY, et al. Serum interferon-inducible protein 10 levels predict hepatitis B s antigen seroclearance in patients with chronic hepatitis B. Aliment Pharmacol Ther **2016**; 43:145-53.

9. Seto WK, Wong DK, Kopaniszen M, et al. HLA-DP and IL28B polymorphisms: influence of host genome on hepatitis B surface antigen seroclearance in chronic hepatitis B. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2013**; 56:1695-703.

10. Galmozzi E, Facchetti F, Grossi G, et al. IFNL4 rs368234815 and rs117648444 variants predict off-treatment HBsAg seroclearance in IFN-treated HBeAg-negative chronic hepatitis B patients. Liver Int **2018**; 38:417-23.

11. Potenza N, Papa U, Mosca N, Zerbini F, Nobile V, Russo A. Human microRNA hsa-miR-125a-5p interferes with expression of hepatitis B virus surface antigen. Nucleic Acids Res **2011**; 39:5157-63.

12. Wang YQ, Ren YF, Song YJ, et al. MicroRNA-581 promotes hepatitis B virus surface antigen expression by targeting Dicer and EDEM1. Carcinogenesis **2014**; 35:2127-33.

13. Li S, Zhao K, Liu S, et al. HBsAg sT123N mutation induces stronger antibody responses to HBsAg and HBcAg and accelerates in vivo HBsAg clearance. Virus Res **2015**; 210:119-25.

14. Larance M, Lamond AI. Multidimensional proteomics for cell biology. Nature reviews Molecular cell biology **2015**; 16:269-80.

15. Graveley BR. Alternative splicing: increasing diversity in the proteomic world. Trends in genetics : TIG **2001**; 17:100-7.

16. Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics **2007**; 7:340-50.

17. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng **2009**; 11:49-79.

18. Kumar C, Mann M. Bioinformatics analysis of mass spectrometry-based proteomics data sets. FEBS Lett **2009**; 583:1703-12.

19. Zhang Z, Bast RC, Jr., Yu Y, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer research **2004**; 64:5882-90.

20. Viswanathan K, Fruh K. Viral proteomics: global evaluation of viruses and their interaction with the host. Expert review of proteomics **2007**; 4:815-29.

21. Bell LN, Theodorakis JL, Vuppalanchi R, et al. Serum proteomics and biomarker discovery across the spectrum of nonalcoholic fatty liver disease. Hepatology **2010**; 51:111-20.

22. Tuck MK, Chan DW, Chia D, et al. Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. J Proteome Res **2009**; 8:113-7.

23. Huang C, Wang Y, Liu S, et al. Quantitative proteomic analysis identified paraoxonase 1 as a novel serum biomarker for microvascular invasion in hepatocellular carcinoma. J Proteome Res **2013**; 12:1838-46.

24. Zhu G, Cai G, Liu Y, et al. Quantitative iTRAQ LC-MS/MS Proteomics Reveals Transcription Factor Crosstalk and Regulatory Networks in Hypopharyngeal Squamous Cell Carcinoma. Journal of Cancer **2014**; 5:525-36.

25. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics **2000**; 25:25-9.

26. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics (Oxford, England) **2014**; 30:523-30.

27. Lasarte JJ, Casares N, Gorraiz M, et al. The extra domain A from fibronectin targets antigens to TLR4-expressing cells and induces cytotoxic T cell responses in vivo. Journal of immunology (Baltimore, Md : 1950) **2007**; 178:748-56.

28. Yang J, Wang F, Tian L, et al. Fibronectin and asialoglyprotein receptor mediate hepatitis B surface antigen binding to the cell surface. Archives of virology **2010**; 155:881-8.

29. Julier Z, de Titta A, Grimm AJ, Simeoni E, Swartz MA, Hubbell JA. Fibronectin EDA and CpG synergize to enhance antigen-specific Th1 and cytotoxic responses. Vaccine **2016**; 34:2453-9.

30. Guidotti LG, Inverso D, Sironi L, et al. Immunosurveillance of the liver by intravascular effector CD8(+) T cells. Cell **2015**; 161:486-500.

31. Utoh R, Tateno C, Yamasaki C, et al. Susceptibility of chimeric mice with livers repopulated by serially subcultured human hepatocytes to hepatitis B virus. Hepatology **2008**; 47:435-46.

32. Li F, Shao Q, Ji D, Li B, Chen G. Genetic association between CD44 polymorphisms and chronic hepatitis B virus infection in a Chinese Han population. International journal of clinical and experimental pathology **2015**; 8:11675-9.

33. Pan JS, Zhou F, Xie CX, et al. Aldolase A-HBsAg interaction and its effect on ultraviolet radiation induced apoptosis in 293FT cells. J Gastroenterol Hepatol **2010**; 25:1702-9.

34. Tsai SY, Segovia JA, Chang TH, et al. DAMP molecule S100A9 acts as a molecular pattern to enhance inflammation during influenza A virus infection: role of DDX21-TRIF-TLR4-MyD88 pathway. PLoS Pathog **2014**; 10:e1003848.

35. Wu R, Zhang Y, Xiang Y, et al. Association between serum S100A9 levels and liver necroinflammation in chronic hepatitis B. Journal of translational medicine **2018**; 16:83.

36. Gregorc V, Novello S, Lazzari C, et al. Predictive value of a proteomic signature in patients with non-small-cell lung cancer treated with second-line erlotinib or chemotherapy (PROSE): a biomarker-stratified, randomised phase 3 trial. The Lancet Oncology **2014**; 15:713-21.

37. Zubarev RA. The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics **2013**; 13:723-6.

38. Tu C, Rudnick PA, Martinez MY, et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. J Proteome Res **2010**; 9:4982-91.

39. Pernemalm M, Lewensohn R, Lehtio J. Affinity prefractionation for MS-based plasma proteomics. Proteomics **2009**; 9:1420-7.

40. Liu J, Jiang M, Ma Z, et al. TLR1/2 ligand-stimulated mouse liver endothelial cells secrete IL-12 and trigger CD8+ T cell immunity in vitro. Journal of immunology (Baltimore, Md : 1950) **2013**; 191:6178-90.

41. Chen Z, Cheng Y, Xu Y, et al. Expression profiles and function of Toll-like receptors 2 and 4 in peripheral blood mononuclear cells of chronic hepatitis B patients. Clinical immunology (Orlando, Fla) **2008**; 128:400-8.

42. Wu J, Lu M, Meng Z, et al. Toll-like receptor-mediated control of HBV replication by nonparenchymal liver cells in mice. Hepatology **2007**; 46:1769-78.

43. Isogawa M, Robek MD, Furuichi Y, Chisari FV. Toll-like receptor signaling inhibits hepatitis B virus replication in vivo. Journal of virology **2005**; 79:7269-72.

44. Guo H, Jiang D, Ma D, et al. Activation of pattern recognition receptor-mediated innate immunity inhibits the replication of hepatitis B virus in human hepatocyte-derived cells. Journal of virology **2009**; 83:847-58.

Table 1. Baseline characteristics of 45 paired chronic hepatitis B patients with and without spontaneous HBsAg seroclearance involved in stage II: prospective validation.

	HBsAg seroclearance (N=45)	HBsAg positive* (N=45)	P value
Age, years	53.76±11.86	53.73±11.88	0.697
Sex, male%	21/45 (46.7%)	21/45 (46.7%)	1
Quantitative HBsAg, log IU/ml	/	2.84(2.29-3.24)	–
Duration of HBsAg seroclearance, years	1.67(0.94-2.23)	/	–
Duration of follow up, years	9.91(6.97-12.63)	9.60(7.03-10.26)	0.487

Duration of HBeAg seroconversion, years	7.16(5.15-8.07)	7.07(5.89-8.18)	0.326
Albumin, g/L	45.00(43.00-46.50)	44.00(43.00-46.00)	0.322
ALT, U/L	22.20(15.50-25.00)	22.00(18.00-34.50)	0.099
AST, U/L	24.00(18.00-27.00)	24.00(22.50-29.00)	0.053
ALP, U/L	60.00(54.00-75.50)	68.00(53.50-76.00)	0.355
GGT, U/L	19.00(15.00-26.50)	18.00(15.25-29.75)	0.645
Bilirubin, umol/L	8.00(7.00-12.00)	8.00(5.50-12.00)	0.716
WBC, x 10^9/L	6.32(4.84-7.44)	5.51(4.42-6.60)	0.138
Platelet, x 10^9/L	229.00(189.00-258.50)	223.00(186.5-258.00)	0.888
Fibronectin, ug/ml	188.82 (156.39-252.66)	166.36 (107.73-219.24)	0.009
CD44, ng/ml	44.30(27.56-62.34)	52.92(33.95-64.62)	0.364
ALDOA, ng/ml	19.40(12.11-23.74)	15.30(11.45-21.30)	0.203
S100A9, pg/ml	19.56(11.63-41.67)	21.70(9.30-38.21)	0.955

Continuous variables expressed as mean ± standard deviation or median (interquartile range) when appropriate. *HBV DNA in all patients was undetectable. HBsAg: hepatitis B surface antigen, HBeAg: hepatitis B e antigen, ALP: alkaline phosphatase, ALT: alanine aminotransferase, AST: aspartate aminotransferase , GGT: glutamyl transferase, WBC: white blood cells. ALDOA: Aldolase A. S100A9: S100 calcium-binding protein A9. A Wilcoxon matched-pairs singed rank test was used for analysis.

Table 2. Accuracy of plasma fibronectin levels in predicting HBsAg seroclearance in stage III: predictive study

Parameters	AUROC	P	SE	95%CI	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Number of patients
Year -3, fibronectin	0.567	0.1353	0.045	0.487-0.644	78.1	35.4	54.7	61.7	164
Year -1, fibronectin	0.688	<0.0001	0.041	0.611-0.758	80.5	54.9	64.1	73.8	164
HBsAg<200 IU/ml*	0.663	0.0155	0.067	0.554-0.761	60.9	69.6	84.8	39.0	87
Year -1 fibronectin in HBsAg <200 IU/ml	0.730	0.0003	0.063	0.624-0.819	79.7	60.9	85.0	51.9	87
Annual HBsAg reduction >0.5 log[#]	0.777	0.0262	0.124	0.651-0.874	71.4	75.0	97.6	15.8	60
Year -1 fibronectin in annual HBsAg reduction >0.5 log	0.884	0.0003	0.106	0.775-0.952	98.2	75.0	98.2	75.0	60

* HBsAg at Year -3 selected. [#]Annual HBsAg log reductions based on time points Year -3 to Year -1 and then divided by two.

AUROC: area under receiver operator curve. SE: standard error. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value.

Figure legends

Figure 1. Overview of the study design for differentially expressed proteins associated with spontaneous hepatitis B surface (HBsAg) seroclearance in chronic hepatitis B (CHB) patients. Stage I: proteins identification via quantitative proteomics and functional analysis of identified proteins via bioinformatics analysis. Stage II: validation of candidate proteins using a separate cohort of patients. Stage III: evaluation of predictive values of identified proteins for HBsAg seroclearance using a historical cohort.

Figure 2. Bioinformatics analysis of the 97 differentially expressed proteins. A. Biological process using Geno Ontology (GO) enrichment analysis. Each annotated protein is assigned at least one GO term. Numbers refer to the frequency of assigned proteins in each category. B. The disease and function generated through Ingenuity Pathway Analysis.

Figure 3. Representative tandem mass spectrometry (MS/MS) spectra of peptides with the reporter ions used for identification and quantitation of the four selected differentially expressed proteins, with 114 and 115 tags labelled HBsAg positive patients (control group), and 116 and 117 tags labelled HBsAg seroclearance patients (experimental group). The lower mass spectrum was the product ion scan (MS/MS) of the fragmented isobaric tags for relative and absolute quantification (iTRAQ)-labeled peptides. The b and y ions were indicative of the peptide sequence. The zoomed spectrum represented the quantity of iTRAQ reporter ions generated from the peptides, indicating the relative levels of peptide in the two groups. Lower left in each subfigure represented the MS/MS spectra of precursor ion for identified fibronectin, ALDOA, CD44 and S100A9, respectively; upper right indicated the relative

quantitative of these four proteins according to the intensity ratio for reporters 116:114 and 117/115.

Figure 4. Validation of the four selected candidate proteins in 45 paired patients who did and did not have HBsAg seroclearance using ELISA in Stage II study. The p values were calculated via Wilcoxon matched pairs tests. The horizontal line indicates the median value for each group. The error bar represents 95% confidence interval.

Figure 5. Serial plasma fibronectin levels in patients with spontaneous HBsAg seroclearance and HBsAg positive (controls) at Year -3, Year -1 and Year 0 in Stage III predictive study. The p values were calculated via Wilcoxon matched pairs tests. The horizontal line indicates the median value for each group. The error bar represents 95% confidence interval.

Figure 1. Overview of the study design for differentially expressed proteins discovery, validation and prediction.

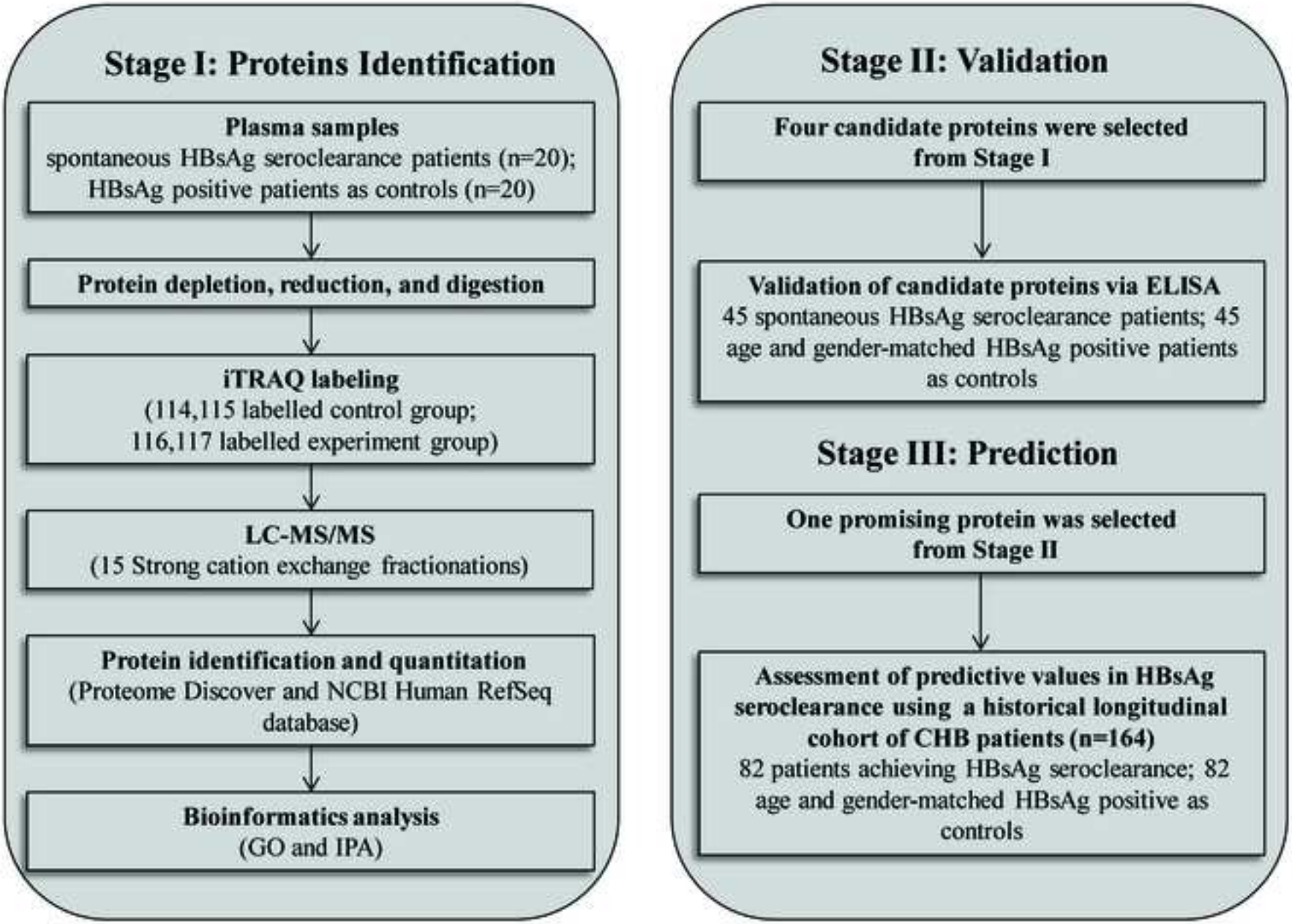

Figure 2A. Biological process via GO

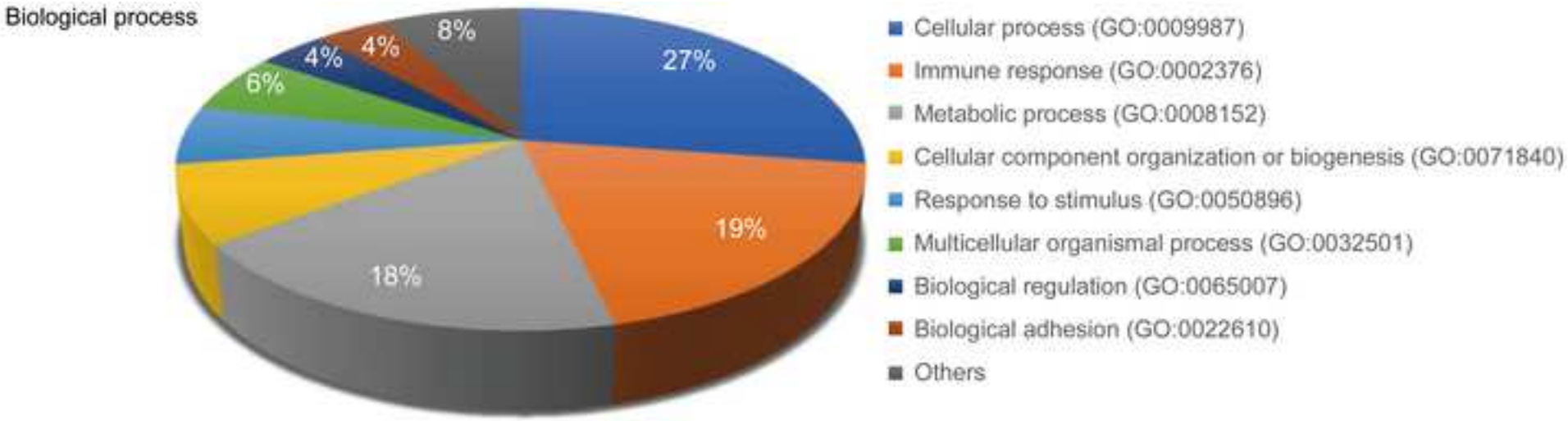

Figure 2B. Disease and function via IPA

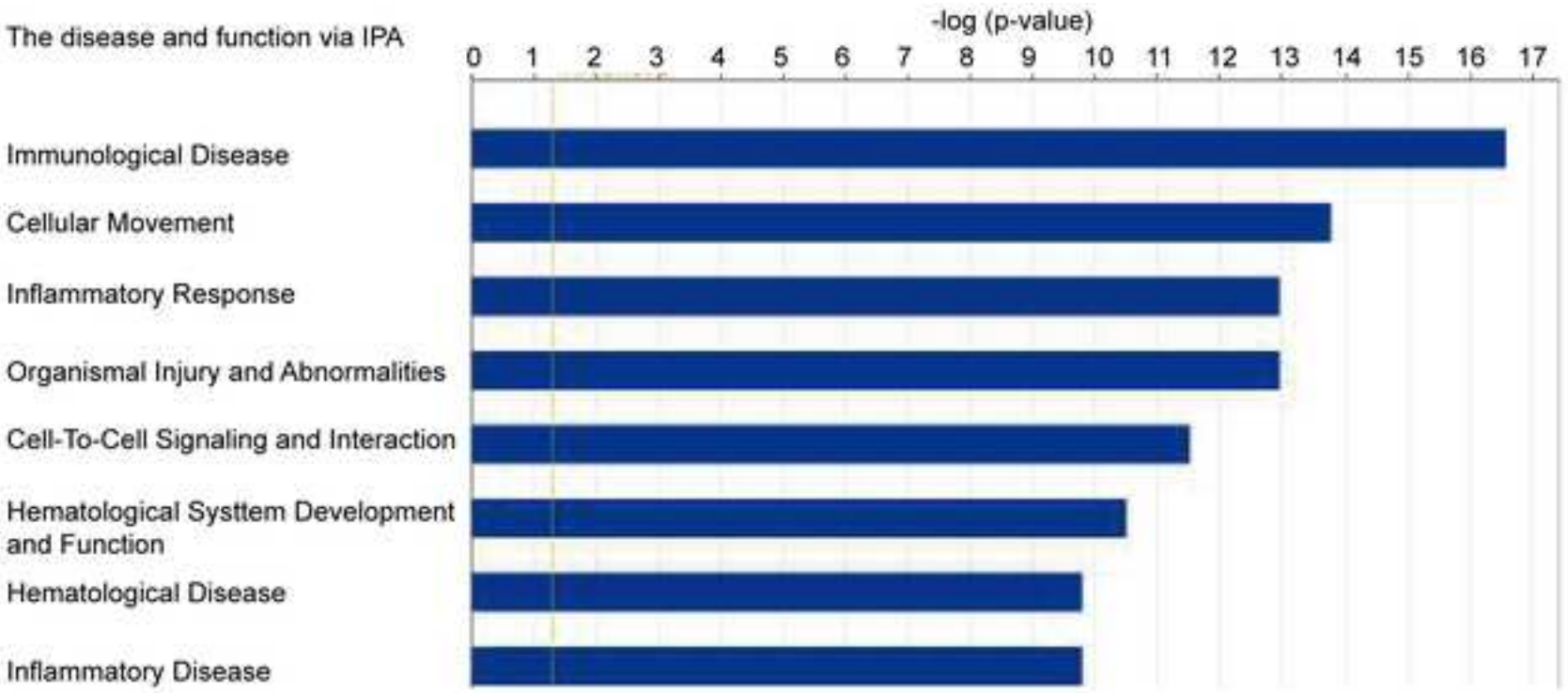

Figure 3A. Fibronectin MS proteomics

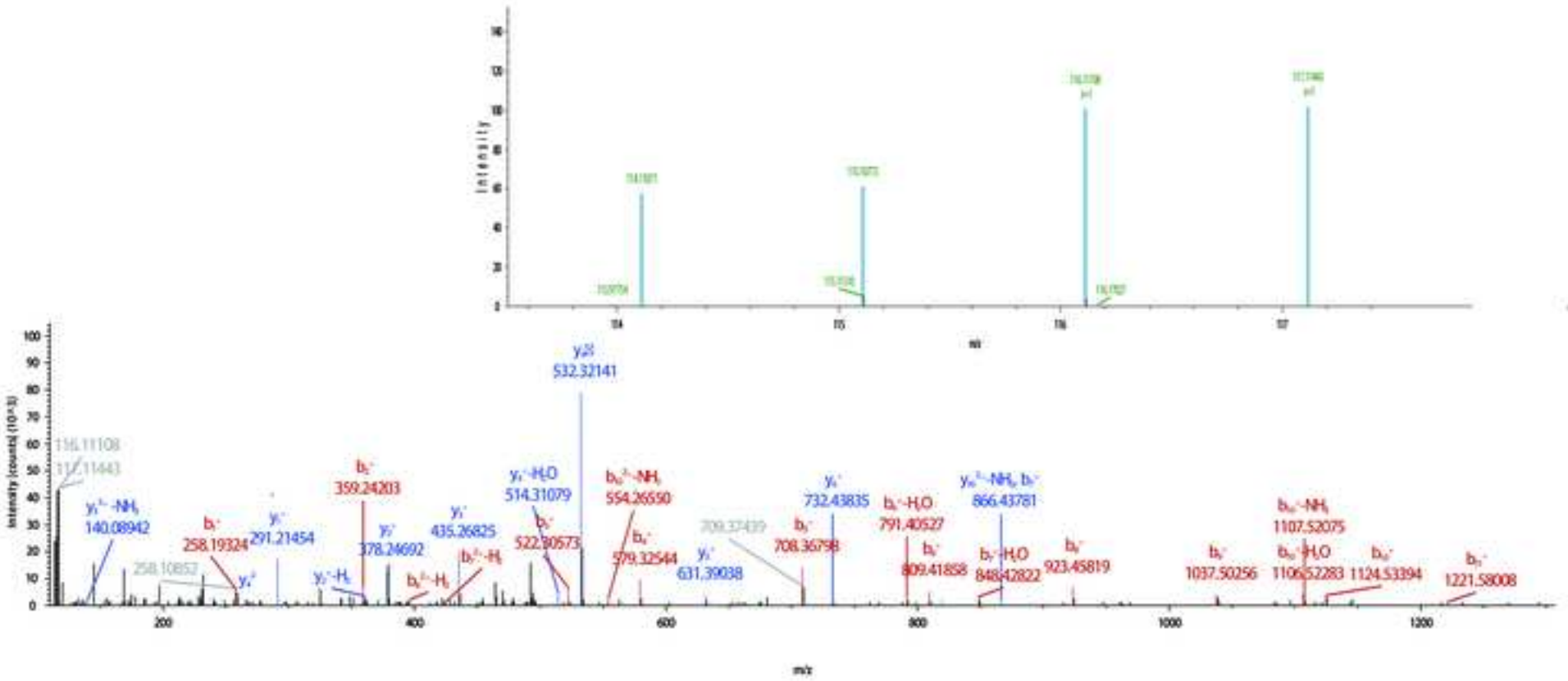

Figure 3B. ALDOA MS proteomics

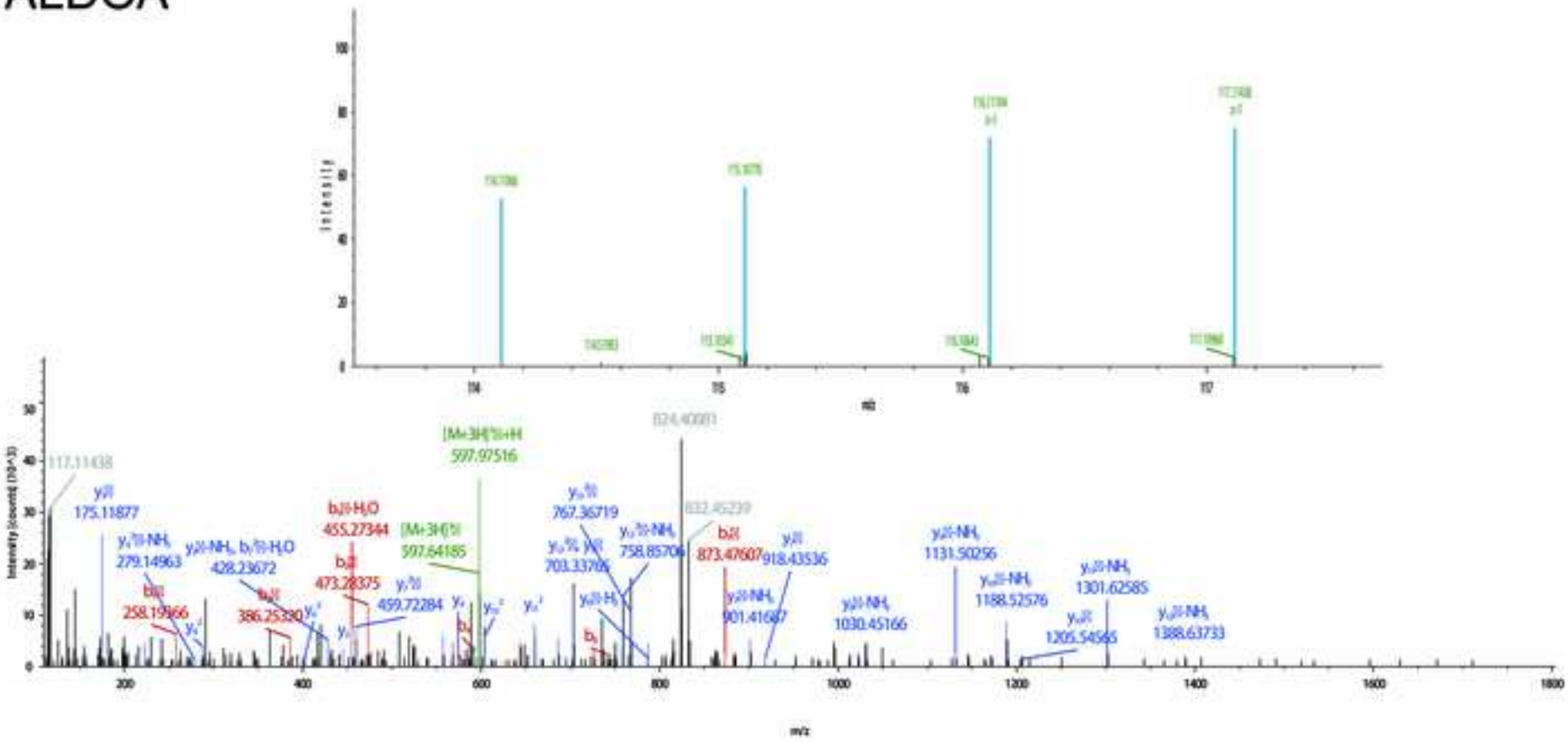

Figure 3C. CD44 MS proteomics

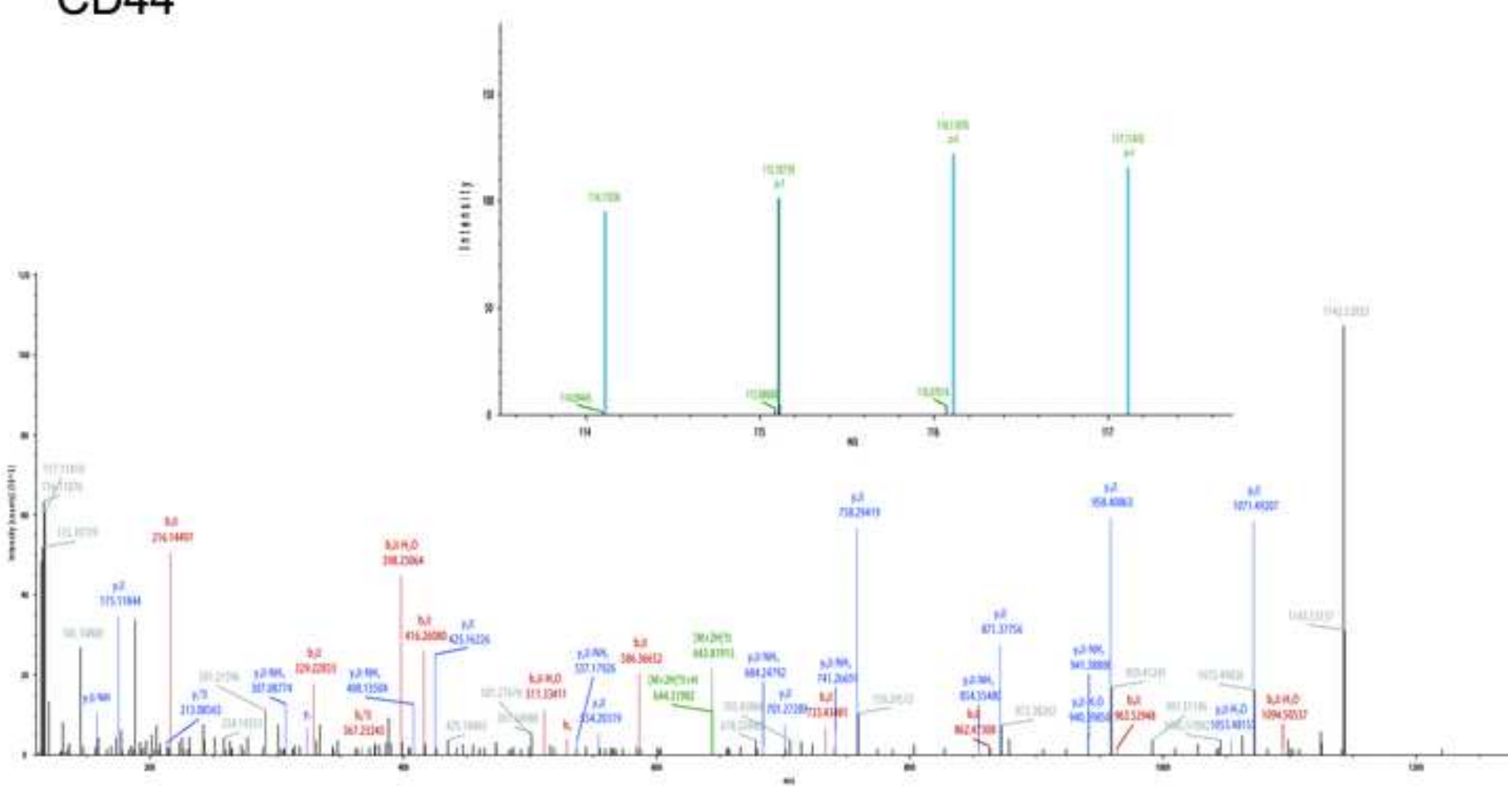

Figure 3D. S100A9 MS proteomics

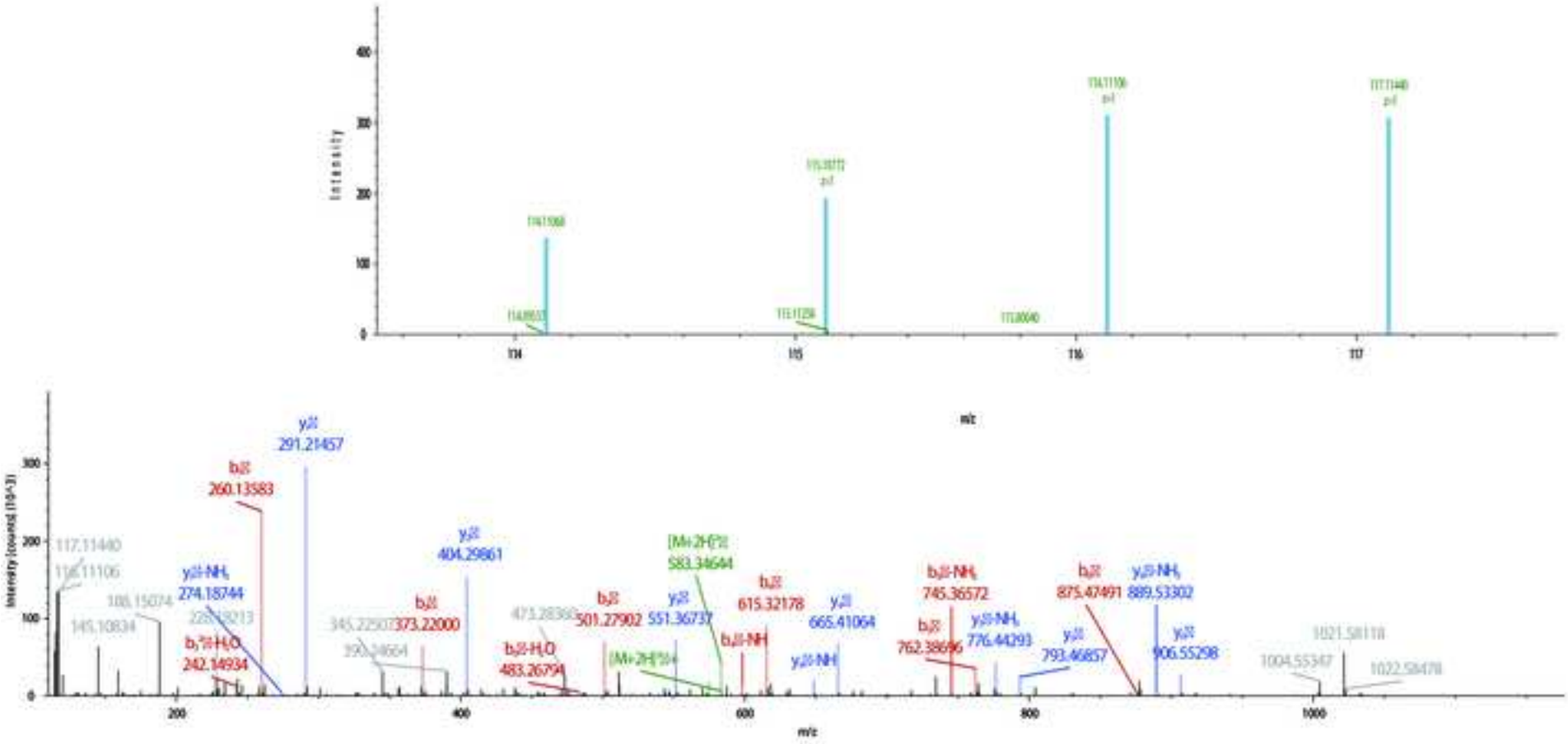

Figure4A-D. Validation results

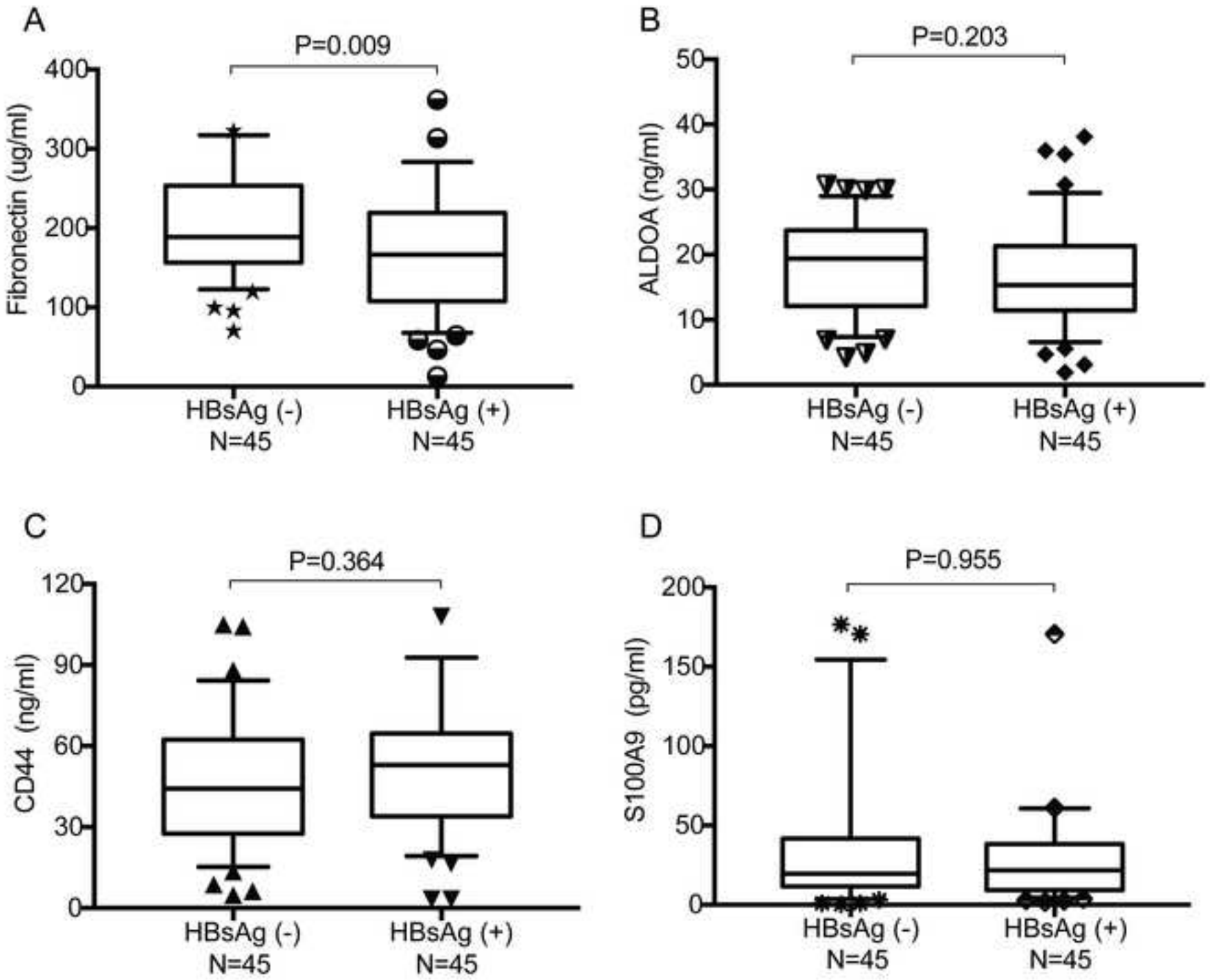

Figure 5. Serial plasma fibronectin levels in patients with spontaneous HBsAg seroclearance and HBsAg positive (controls) at Year -3, Year -1 and Year 0 in Stage III predictive study. The p values were calculated via Wilcoxon matched pairs tests. The horizontal line indicates the median value for each group. The error bar represents 95% confidence interval.