Application of Magnetic Bead-based Proteomic Fingerprinting Technology to the Detection of Liver Fibrosis in Patients with Chronic Hepatitis B Infection

WONG, Yee Man Melody

A Thesis Submitted in Partial Fulfillment for the Requirements for the

Degree of Master of Philosophy

In

Medical Sciences

The Chinese University of Hong Kong

August 2009

Abstract

Liver biopsy is a gold standard for detecting liver fibrosis. However, it is invasive and subject to sampling error. Non-invasive predictors are therefore urgently needed. Surface-enhanced laser desorption/ionization (SELDI) Proteinchip technology is a useful tool for biomarker discovery. Nevertheless, the main drawback is its incapability of obtaining protein identity of a biomarker. The aims of the study were to establish an automated magnetic bead-based proteomic profiling technology, and to identify proteomic markers for detecting liver fibrosis in patients with chronic hepatitis B virus (HBV) infection using the established technology.

We successfully established an automated magnetic bead-based technology for high-throughput quantitative serum profiling, which can greatly reduce batch to batch variation and increase reproducibility. The intra-assay and inter-assay coefficients of variation (CV) of the quantitative proteomic profiling assay were evaluated to be less than 30%. It allowed simultaneous analytic profiling and obtaining preparative proteome fractions for subsequent protein identification experiments.

In the 2nd part of the study, the established automated proteomic profiling technology was applied to analyze 214 treatment-naïve and post treatment chronic HBV infected patients with different degrees of liver fibrosis (Ishak score 0-6). Quantitative proteomic profiling was obtained by using C18 hydrophobic magnetic beads and strong anion exchange (SAX) beads and analyzed by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). Linear regression method (forward stepwise) was used to develop a diagnostic model for predicting the degrees of liver fibrosis from the serum proteomic profiles, clinical data, serological data, and biochemical data.

Pre-treatment samples were used to develop diagnostic model. 4 proteomic features, platelet count, HBV DNA, bilirubin and prothrombin time were found to be associated with the degrees of liver fibrosis. By linear regression method (forward stepwise), two proteomic features (*m/z*

9,165 and *m/z* 12,443) and prothrombin time was included in the diagnostic model. Post-treatment samples were used for independent validation. In the independent validation, the AUROC curve was 0.750 (76% sensitivity, 79% specificity). *m/z* 9,165 proteomic feature was identified as apolipoprotein C-III.

In conclusion, serum proteomic fingerprint is a good non-invasive method to supplement liver biopsy for assessment of liver fibrosis, regardless of treatment status.

摘要

肝臟切片檢查是偵查肝臟纖維化的標準方法。然而，它是侵入性的和其準確性受抽樣誤差所影響，因此非侵入性的檢查方法有迫切性的需求。表面增强激光解吸附電離起飛行時間質譜技術 (SELDI-TOF-MS) 能幫助尋找標誌蛋白，其主要缺點是無法直接鑒定標誌蛋白的身份。研究目的是建立自動化磁珠蛋白指紋圖譜技術作血清學分析，並應用此技術尋找標誌蛋白來偵查慢性乙型肝炎患者肝臟纖維化。

我們成功建立自動化磁珠蛋白指紋圖譜技術作大量快速篩選。大量快速篩選的好處是減少逐批質量差異和增加可重複性。同日與異日之變異係數少於30%。這種磁珠技術允許同時分析血清蛋白及獲得蛋白樣本爲蛋白標誌作證明實驗。

我們利用磁珠技術分析214個來自未接受治療及已治療慢性乙型肝炎肝臟纖維化患者的血清樣本，透過使用C18磁珠及基質輔助激光解吸附電離起飛行時間質儀(MALDI-TOF MS) 獲得定量蛋白表達圖譜。首先利用未接受治療的血清樣本尋找標誌蛋白，之後利用線性迴歸(linear regression) 順向逐步方式(forward stepwise) 分析標誌蛋白、臨床、血清的生物化學數據，從而建立診斷模型來預測肝臟纖維化。

研究結果發現，4 個標誌蛋白、血小板計數、HBV脫氧核糖核酸、膽紅素和凝血酶原時間的變化均與肝臟纖維化有關。經過線性迴歸(linear regression) 順向逐步方式(forward stepwise)分析，診斷模型包括 2 個標誌蛋白 (*m/z* 9,165和*m/z* 12,443)和凝血酶原時間。以診斷模型評估已治療的血清樣本，接受器運作指標曲線下面積是0.750 (76%敏感性， 79%特異性)。蛋白質鑑定的結果顯示其中1個標誌蛋白爲載脂蛋白CIII（Apolipoprotein CIII）。

總括而言，建基於磁珠血清蛋白指紋圖譜技術是一項能有效評估肝臟纖維化的非侵入性方法。

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors, Prof. Terence C. W. Poon, for accepting me to pursue these studies, his teaching, help, guidance and encouragement throughout this work. I am also thankful for the financial support, a CERG grant to Dr. Henry L.Y. Chan, offered by University Grants Committee, Hong Kong. Without this financial assistance this work may not have been possible.

I would like to thank Prof. Joseph J.Y. Sung for his continuous encouragement and support; Dr. Alex Y. Hui, former staff in the Department of Medicines and Therapeutics, CUHK, and Dr. Henry L.Y. Chan, for providing the patient samples and clinical data for this study.

I would also like to thank Kiwi Y.W. Chan for working with me to the optimization of BioDot nano-liter dispenser. Special thanks are given to fellows labmates in Medical Proteomics Laboratory especially Ms Eddy W. Y. Ng for assisting in the protein identification work. Moreover, many thanks to Dr. Poon's former students, Ms Irene L. Ang, Ms Karen O.Y. Yu and Ms Man K. Law for all their help and support. I am indepted to my supervisor for teaching me the statistical and bioinformatics analyses.

I would like to thank my family for providing me encouragement and emotional support throughout my studying period.

ABBREVIATIONS

Units

Milligram	mg
Microgram	μg
Milliliter	ml
Microliter	ml
Molar	mol
Millimolar	mmol
Dalton	Da
Kilodalton	kDa
Hour	hr
Minute	min
m/z	mass to charge ratio

Diseases

Hepatitis B virus	HBV
Hepatitis C virus	HCV
Chronic Hepatitis B	CHB
Chronic Hepatitis C	CHC
Hepatocellular carcinoma	HCC
Chronic Liver Disease	CLD

Biochemical Terms and Terminology

Matrix-Assisted Laser Desorption/Ionization	MALDI
Surface Enhanced Laser Desorption /Ionization	SELDI
Electrospray	ESI
Time-of-flight	TOF
Mass spectrometry	MS
High Performance Liquid Chromatography	HPLC
Liquid Chromatography	LC
Linear Ion Trap Quadruple	LTQ
Fourier-Transform Ion Cyclotron Resonance	FTICR
Significance Analysis of Microarray	SAM
Artificial Neural Network	ANN
Hepatic stellate cell	HSC
Extracellular matrix	ECM
Apolipoprotein	Apo

CONTENTS

Review of literature

Overview of liver fibrosis

Liver fibrosis is a scarring process caused by chronic liver injury. This wound-healing response leads to the accumulation of fibrillar extracellular matrix (ECM) components which causes the progressive distortion of the hepatic architecture and eventually loss of hepatic function. [1] Liver fibrosis can be induced by different etiologies such as hepatitis B or C viral infection, alcoholic and nonalcoholic steatohepatitis (NASH). [2] It is a step-by-step process starting from minimal fibrosis which is usually undetectable, followed by liver cirrhosis and ultimately ending in liver cancer. Hepatitis B and C, two major causes of chronic liver diseases (CLD), are endemic in worldwide with 350 million and 170 million people live with chronic infection respectively. [3] Hepatitis C remains the most common cause of CLD in western countries such as the United States and European countries while it is not common in Asia. On the contrary, hepatitis B is common in Asia especially in China. The prevalence of chronic hepatitis B is also high in Asia relative to Europe. 75% of the chronically infected population is in Asia while less than 1% of the chronic population is found in Western Europe and North America. Fung *et al.* reported that the prevalence of severe fibrosis in chronic hepatitis B was 34% in an Asian cohort, they also showed that hepatitis B was a leading cause of liver fibrosis, cirrhosis and liver cancer. [4] Other factors such as antioxidants, hepatic iron stores, age, HIV coinfection and obesity are considered to enhance hepatic injury [5][6] and thus increase the progression of liver fibrosis.

Pathophysiology of liver fibrosis

Fibrosis is a dynamic process and evidences show that it is reversible even in advanced fibrosis. [7] It is governed by the turnover rate of ECM which include glycoproteins, collagens and proteoglycans. ECM can be synthesized by hepatocytes, bile duct epithelial cells and endothelial cells but is mainly made by fibroblastic cell population. [8] Once these cells are injured, the composition of hepatic ECM changes and the liver becomes fibrotic.

Historically hepatic stellate cell (HSC) has been considered as the key contributor for the abnormal deposition of fibrillar ECM. [9] It is also known as lipocyte, ito cell or perisinusoidal cell which stores vitamin A. Upon hepatic injury, "quiescent" stellate cell will become "activated", changing from vitamin A-storing cell to myofibroblast-like cell with morphological and functional changes. [10] Morphological changes include loss of vitamin A and increase in rough endoplasmic reticulum while functional changes consist of increase in collagens, fibronectin, laminin and proteoglycans secretion. [11] Following HSC activation, gene expression pattern also changes; for example, a remarkable increase in types I and III collagens and expression of HSP47, a collagen-binding stress proteins reported as a collagen-specific molecular chaperone during collagen biosynthesis. [12] In addition, HSC activation can be mediated by induction of signaling cascades. Several signaling cascades, transforming growth factor beta (TGF-β), connective tissue growth factor (CTGF) and platelet-derived growth factor (PDGF) can control the induction of gene expression and fibrogenic response of HSC during liver fibrosis. [13-15]

The synthesis and degradation of ECM must be maintained in a critical balance and is tightly regulated by matrix metalloproteinases (MMPs)–mediated turnover of ECM proteins. [16] MMPs can destroy ECM by cleaving native collagen through the helical proportion of the molecule. On the contrary, natural inhibitors, tissue inhibitors of metalloproteinases (TIMPs) block and control the ECM turnover by interacting with the active sites of MMPS in non-covalent manner. Therefore, the occurrence of fibrosis is due to the decreased activity of MMPs made by the overexpression of TIMPs. [17]
In addition, Kupffer cells, specialized leukocytes present in the liver, can also cooperate with cytokines released by infected liver cells to stimulate stellate cell to produce collagen fibers. [18]

Fibrosis in chronic viral hepatitis patients is a dynamic process that occurs throughout one or two decades to reach cirrhosis and even longer for liver cancer. Initially, it is limited to the portal tracts which cause portal hypertension to the patients, followed by extensive fibrosis septa occurring in the liver parenchyma and ended by septum formation and rings of scars surrounding nodules of hepatocytes. [19]

Histological classification of liver fibrosis

As liver fibrosis takes years or decades to develop, it is essential to monitor the status of liver fibrosis so that clinicians can make good use of this important parameter to assess the risk of disease and make therapy decision. It is particularly valuable for patients who are in compensated state since therapy at right time is critical for the reversal of fibrotic state in the liver.

Liver biopsy is the gold standard for staging liver fibrosis. It allows direct histological examination of liver for diagnosis and prognosis. With new insights of liver fibrosis under different etiologies, grading of liver fibrosis aims to confirm clinical diagnosis, describe and quantify necroinflammation and fibrosis, evaluate possible risk of disease and assess the treatment progress. [20] Several semiquantative scoring systems have been used recently, namely histological activity index (HAI), Ishak's score and METAVIR scoring system.

HAI

HAI scoring system was developed in 1981 by Knodell and colleagues. [21] It is a combined score classified into four categories: periportal necrosis (1-10); parenchymal damage (0-4); portal inflammation (0-4) and fibrosis (0-6). The score for each category is discontinuous and with weighting. Periportal necrosis has the greatest weighting among others because of its greatest influence in determining the activity and severe of chronic aggressive (active) hepatitis (CAH). [22]

HAI was the first reproducible histological scoring system in assessing liver biopsy. With a total score of 22, it provides a broad range of scores which can differentiate small differences between biopsies. However, the main frequently cited criticism is the combination of necro-inflammation (grading) and fibrosis (staging) in the existing system should be separately assessed due to their differences in nature. [23]

Ishak score

Due to the inappropriate combination of necrosis and fibrosis, a modified HAI grading and staging was developed. It allows separate assessment in grading and staging. Grading is used to describe the intensity of inflammation while staging refers to the

amount of fibrosis detected by liver biopsy. [24] Another modification upon traditional HAI is that it has continuous range in nature. In modified HAI grading, there are four categories with the maximum possible grading score of 18.

In staging, fibrosis, architectural disturbance and cirrhosis are important features needed to consider. Staging ranges from no fibrosis to definite cirrhosis (Ishak score 0-6). An Ishak score > 2 indicates the presence of significant fibrosis and an Ishak score > 4 indicates the presence of liver cirrhosis. [25]

METAVIR scoring system

METAVIR scoring system was developed specifically for patients with hepatitis C. It also assesses grading and staging separately. The grading score ranges from 0 (no activity) to 3 (severe activity) and the fibrosis score (staging) is a 5-point scale from 0 (no fibrosis, no scarring) to 4 (cirrhosis, advanced scarring). [26]

Gold standard for fibrosis assessment – Liver biopsy

Liver biopsy has long considered as the gold standard to diagnose the presence of fibrosis and stage the disease. It is a direct method for assessing liver injury, providing information on fibrosis, inflammation, necrosis, steatosis and hepatic iron load. [19] In addition, it allows the identification of suspected and unexpected cofactors as well as comorbidities. However, there are several limitations that hinder the use of liver biopsy. Researchers started to focus on the development of non-invasive methods to assess liver fibrosis. Liver biopsy is an invasive method and may cause patients suffering from pain and complications, especially for patients with prolonged prothrombin time and low platelet count as excess bleeding may occur after tissue

collection. Cadranel *et al.* reported that 20% and 0.57% of the patients in the study group suffered from pain and severe complication respectively. [27] The other criticism is sampling error. As only 1/50,000 of the whole liver is examined, the biopsy result may not reflect the full picture of the entire liver. Reports showed that discordance of fibrosis stages occurs in left and right lobes of liver for the same patient. [28] [29] In addition, the size of the tissue greatly affects the diagnostic accuracy of liver biopsy. Colloredo *et al.* reported that smaller sample might give milder stage and thus staging will be significantly underscored. [30] Another well-known drawback is inter-observer variability among pathologists which gives uncertainty on the accuracy of liver biopsy. [29]

On the other hand, its invasive nature also gains reluctance from patients. Because of this, it is not able to monitor treatment progression which is definitely important for antifibrotic therapies. It is costly as patients require hospitalization for 6-18 hours and consumes manpower issues. [19] Because of the sampling error, observer variation and invasive nature in examining liver biopsy, new noninvasive tests for detecting liver fibrosis are likely to have a role in the future.

Biomarker in blood – noninvasive method for assessing diseases

Human blood contains huge amount of proteins and is thought to be a reservoir of biomarkers. Liotta *et al.* pointed out that every cell would leave a record of its physiological state by producing a product that shed to the blood. Their relative cellular abundance, together with cleaved or modified form, can reflect the ongoing physiological and pathological events. [31] More publications showed that the

peptidome (peptides less than 50,000 Da) might contain rich unstudied disease-specific diagnostic biomarkers. [32]

By Food and Drug Administration (FDA) definition, "diagnostic" includes determining the disease risk, screening for diseases and confirming the suspected diseases. It covers determining prognosis or staging the diseases as well as monitoring treatment response. [33] A good biomarker should have properties including good sensitivity, specificity and strong association with the diseases. Furthermore, it should be useful in the majority of the population and have strong diagnostic, prognostic and predictive significance. [34]

With an enormous diversity of proteins in blood which is readily accessible from patients with minimal pain, it is the best source for detecting diseases. However, its complexity gives researchers a great challenge in biomarker discovery. Lack of sensitivity and specificity are still the major obstacles for current biomarkers. Only few biomarker assays have been submitted to FDA for approval though an overwhelming interest in biomarkers discovery rose in the past decades. [33] To get approved from FDA, the new assay should be able to establish adequate analytical (accuracy and precision) and clinical (sensitivity, specificity and some indication of clinical utility) performance. Several successful examples include prostate-specific antigen (PSA) for prostate cancer and alpha-fetoprotein (AFP) for liver cancer. Failure in fulfilling the requirements can be explained by heterogeneity among patients. The severity of diseases and epidemiological heterogeneity including age, sex, ethic and genetic background can reduce the specificity of biomarkers.

Despite of these challenges, there is a growing interest in biomarker discovery field. With more advanced technologies and increasing knowledge of the pathophysiology of diseases, more good quality biomarkers can be discovered and accepted by FDA for clinical use.

Significance of non-invasive markers of liver fibrosis

Due to the invasive nature and limitations of liver biopsy, serum markers offer an attractive alternative to liver biopsy. There has been considerable interest in non-invasive tests for accurate assessment of liver fibrosis. The main merits of serum markers are non-invasive and readily accessible from patients. It is cost effective and less manpower is needed compared with liver biopsy. Liver biopsy needs to be obtained by hepatologists or radiologists [35] while nurses or skilled medical workers can help patients taking blood for assessment and no hospitalization is required. Minimal pain in blood taking makes it more versatile and easily accepted by patients. Also, severe complication is unlikely to occur during blood taking. This enables the evaluation of the efficacy of treatment regimens as patients are pleased to back to clinics for continuous assessment.

Yet, no true serum marker has been validated to date and acted as a surrogate marker of liver fibrosis. [1] Limitations of current non-invasive serum markers include inadequate reproducibility, sensitivity and specificity of the marker, giving at least 20% expected rate of misdiagnosis for non-invasive models. [19] Moreover, most non-invasive markers reported in literatures can classify well for severe cases like cirrhosis [36] but are lack of reliable identification and classification of the early stage cases.

In order for non-invasive markers to be implemented in clinical practices, some requirements should be fulfilled. Sebastiani *et al.* proposed that an ideal non-invasive marker for liver fibrosis should be liver specific, easy to perform and able to provide information on fibrosis stage and fibrogenesis activity. It is even better if the half-life and excretion route is known and independent of comorbidities. More importantly, the biomarker should be sensitive enough to discriminate between different fibrosis stages and reproducible across different centers. [19]

Biomarkers of liver fibrosis

Serum markers of liver fibrosis can be classified into two categories, direct and indirect markers. [1][34] Direct markers are those reflect the ECM turnover which are thought to be closely related with deposition of ECM in liver. They may be the structural elements of fibrogenesis or key mediators involved in degradation of fibrosis. Indirect markers reflect perturbations in liver function. They have potential to directly reflect morphological, functional, and metabolic changes of the liver, monitoring the progression of liver diseases.

Direct markers

With clearer insight into the mechanism of liver fibrosis, researchers start to investigate the components involved in the ECM biosynthesis and degradation, searching for potential markers particularly govern the ECM turnover and correlate with the stages of liver fibrosis. Potential markers include compounds from collagen synthesis or degradation, enzymes in matrix biosynthesis or degradation, ECM

glycoproteins and proteoglycans. The mechanism of liver fibrosis is independent of the etiologies of the liver injury, as no obvious difference is found in the progression of fibrosis. Therefore, it is believed that direct markers can be implemented to all liver fibrosis patients with different etiologies. However, the costs of checking those direct markers are so high that it is not routinely available in most clinical setting. There are some currently available direct biomarkers, namely hyaluronic acid (HA), laminin, YKL-40 (Chondrex), type I and IV collagens and cytokines. [37] HA is a glycosaminoglycan found in virtually all connective tissues and in liver fibrosis. It is a component of ECM synthesized by HSC. HA is increased particularly in cirrhosis patients and gives better sensitivity and specificity to cirrhosis than fibrosis. [38] It can be concluded that HA may be a good marker in excluding cirrhosis but not for fibrosis. It is included in some non-invasive models such as Hepascore and FibroSpect score. [39] [40] Laminin is a major noncollagenous glycoprotein synthesized by HSC. [41] It increases during fibrosis due to alcohol consumption and viral hepatitis and shows reasonable performance with greater than 80% accuracy. [42] It is deposited around the vessels, in the perisinusoidal spaces and portal tracts. [34] YKL-40 is a glycoprotein in chitinase family. It is a relatively new marker of liver fibrosis. The cellular source is supposed to be activated HSC. [43] Although its physiological function is unknown, it is believed that YKL-40 may be a growth factor for fibroblasts, chondrocytes, and synovial cells. It is also a migration factor for endothelial cells.

Indirect markers

Indirect markers can be further classified to routine serum tests markers and imaging test. Based on readily available biochemical parameters from routine blood tests, researchers found a number of them that were correlated with the severity of liver fibrosis and could be served as an indication of liver function. The use of those indirect markers is cheaper compared with direct markers and liver biopsy. Several non-invasive models were constructed for detection of liver fibrosis and their corresponding diagnostic accuracies were summarized in Table 1. Nevertheless, these models were greatly limited by lacking external validation and therefore further extensive investigation was needed. Naveau *et al.* reported that FibroTest was the best models of liver fibrosis among all non-invasive models [49] while other studies show it only gave moderate performance like other non-invasive models. Further validation is needed to confirm its clinical significance. Moreover, some models contain markers (i.e. GGT and apolipoprotein A1) that are not included in routine blood test parameters, extra test needs to be done and it will increase the test cost, making it less versatile to the public.

Despite many new non-invasive models were developed and seemed to be clinically useful, discrepancies between centers made them inappropriate for clinical practice. Diagnostic models using inexpensive biochemical parameters related to liver function indeed have its clinical value, more inter-laboratories validation and systematic reviews together with standardization of laboratory procedures should be done so that these models can be launched for diagnosis.

Different from direct markers, indirect markers are measures of liver function, evaluating the extent of liver injury. Therefore, liver injuries caused by different

diseases may perform differently in the same models. Currently available models are mainly constructed based on HCV patients. It is necessary to check if the models are applicable to other etiologies and evaluate their differences in performance. Some publications concerned about the possibility of applying HCV-based models to CHB patients as their clinical courses and viral pathologies were different. [50] Patients with HBV often have increased ALT level and fluctuated liver enzyme levels while for HCV patients, disease progression is slow and silent without fluctuation of liver enzymes. Some publications even reported that CHB patients have a greater disease progression rate to cirrhosis than CHC patients. [50] It is believed that more evaluation is needed before directly applying HCV-based model to HBV studies.

Table 1 Summary of current non-invasive diagnostic models. N.A. = Not applicable

Marker / Model	Markers included in diagnostic model	Accuracy (AUC)		Cohort	Reference
		Fibrosis	Cirrhosis		
Hepascore	Age, sex, total bilirubin, GGT, α2-macroglobulin, and hyaluronic acid	0.81	N.A.	CHC	39
FibroSpect	HA, TIMP-1, and α-2 macroglobulin	0.92	N.A.	CHC	40
APRI	AST and platelet count	0.82 0.76 0.69 0.87	0.92 0.82 0.75 0.84	CHC 22 HCV studies CHB CHC	44 45 46 46
Forns' index	Age, GGT, cholesterol, and platelet count	0.81	N.A.	CHC	47
FibroTest	GGT, bilirubin, haptoglobin, apolipoprotein A1, α2-macroglobin	0.84 0.83	 0.94	CHC Alcoholic liver diseases	48 49
ASPRI	Age, spleen size and platelet count	N.A.	0.91	CHB	50
Mohamadnejad *et al.* model	HBV DNA levels, alkaline phosphatase, albumin, and platelet counts	0.85	N.A.	CHB	51
Hui *et al.* model	Body mass index (BMI), platelet count, serum albumin, and total bilirubin	0.77	N.A.	CHB	25
Zeng *et al.* model	α2-macroglobulin, age, GGT and hyaluronic acid	0.77	N.A.	CHB	52

Other non-invasive method is imaging. Imaging methods include transient elastography (Fibroscan), magnetic resonance imaging (MRI) and contrast ultrasonography. [53] Recently, Fibroscan has been evaluated and implemented for the assessment of liver fibrosis in many countries. [54] It assesses the presence of liver fibrosis by measuring the stiffness of the liver. An ultrasound transducer probe is mounted on the axis of a vibrator. The mild amplitudes with low frequency vibrations are transmitted by the transducer, inducing elastic shear waves that propagate through the tissue. Pulse-echo ultrasound acquisitions can follow the propagation of the shear wave and measure its velocity. The velocity is directly related to tissue stiffness; fibrotic liver generates higher FibroScan measurement signals measured in kilopascals (kPa). Measurements are taken in the right lobe of the liver through the intercoastal spaces. This painless and rapid (less than 5 minutes assessing time) method has other advantages including immediate result with good reproducibility and more representative as sample volume is greater than liver biopsy by at least 100 times. Many studies have evaluated the clinical performance of FibroScan with different etiologies of CLDs and found that Fibroscan can give fairly stable performance among different etiologies. [55] A meta-analysis of Fibroscan showed that it has excellent utility for the identification of cirrhosis but large variation of AUROC was found in different liver diseases. [56] In France, FibroScan is recommended to be the initial assessment for treatment naïve CHC patients or patients with no concomitant disorders. [57] Yet, Sandrin *et al.* reported that FibroScan was limited to patients with ascites, obese and narrow intercoastal space. [58] Liver stiffness measurements are also influenced by ALT flares. Chan *et al.* reported that patients with the same fibrosis staging but higher ALT

levels are more likely to have higher liver stiffness measurement (LSM). The performance of FibroScan is also seriously affected for those patients with low fibrosis stage and elevated ALT. Different cut-offs and algorithms were used for definite patients groups. [59] Careful evaluation should be done to define a standardized cut-offs and algorithms for clinical use.

Why Proteomics?

With the completion of human genome project in 2003, surprisingly, it was discovered that human had barely more genes than much simpler organisms such as roundworm and fly. Human complexity is therefore thought to be contributed by gene functions and the way gene products interact, putting a new insight on protein study. [60] Considering the biological flow, from DNA to RNA to protein, genome only represents the first step of understanding biological functions. It has already an idea that one gene to one protein is not hold; there are about 500,000 proteins in human proteome while only 40,000 genes are present in human genome, showing that there would be more complicated in protein level. [61] In genomics, a moderate correlation was found ($r=0.61$) between protein abundance and mRNA level, indicating that there is a gap between the genome sequence and cellular behaviour. It is obvious that genomics cannot provide a full answer for biological questions. A new level of gene regulation, microRNA (miRNA), was first introduced in 2001, [62] and proteomics may become supplement of current work on genomics. More effort should be passed from genomics to miRNA and proteomic studies.

"Proteomics" is the study of all proteins in a biological system. It provides a more realistic view of biological status and therefore is more useful than genomics for evaluating diseases. Poor correlation between gene and protein can be explained by the differences in the synthesis rate and half-life of mRNA as well as the protein encoded. [60] In addition, there are some distinguish advantages on proteomics. Proteomics describes dynamic cellular process. It provides information about post-translational modifications (PTM) such as glycosylation and phosphorylation which may be essential for protein function and activity. The contribution of proteomics is fundamental in understanding factors which can alter gene expression but cannot be determined by DNA sequence. [63]

In conclusion, proteins are the core of cellular functions. Study of proteins offers researchers the possibility of identifying disease-associated biomarkers for diagnosis, prognosis or drug development.

Clinical values of proteomics in biomarker discovery

Clinical proteomics aims to discover diagnostic, prognostic and predictive biological markers (clinical application) as well as to detect and validate novel drug targets (pharmaceutical application). In this regard, proteomics can be classified into expression proteomics, functional proteomics and structural proteomics. [64]

Expression proteomics is the large-scale study of variations in protein expression. [60] It is the most popular and expensive proteomics strategy which is analogous to differential gene expression. [60] It involves in cataloging proteins expressed in serum,

cell or tissue to identify alternations between control and disease samples. Furthermore, characterization of the biomarkers has potential to improve patient outcomes.

Functional proteomics, a specialized form of proteomics, aims to characterize protein activities, protein-protein or protein-DNA/RNA interactions and signaling pathways. Human genes produce huge amount of proteins via PTMs and mutations which can be the sources for examining diseases. Also, a detailed description of the cellular signaling pathways can help the elucidation of protein-protein interaction in vivo. [65] Unlike expression proteomics, specialized proteins or subproteome are isolated for characterization by using affinity chromatography rather than obtaining a protein profile to search for differential proteins. More biological functions of proteins can be known by understanding the rapid and transient association within large proteins complex.

Structural proteomics aims to elucidate structure-function relationships of uncharacterized gene products based on 3D protein structure. It is the deduction of biological function from the predicted protein structure. [66] Protein structure can be identified by using X-ray crystallography, NMR spectrometry and computational method such as comparative and de novo structure prediction. The structural information derived from structural proteomics is important in drug development in the pharmaceutical industry.

Challenges in proteomics

Thousands of proteins per study can be identified in typical mammalian cell but it cannot apply to human plasma because of complexity. Much less proteins can be

identified in human plasma or serum due to the presence of high abundance proteins. In human plasma, the 20 most abundance proteins contribute 99% of the total protein and mask the remaining 1% of low abundance proteins which are usually potential biomarkers. Large dynamic range of at least 9-10 orders of magnitude in serum proteome is the major challenge in clinical proteomics, from the most abundance protein, albumin with concentration of 45mg/mL, to low abundance proteins with 1-10pg/mL. [67] Currently there is no publication proposing a proteomics method that can withstand a dynamic range of 10^9. Most of the proteomics technologies have the dynamic range limit of 10^3. In combination with other advanced separation techniques, the limit can be increased to 10^4 or 10^5. Even with the most advanced LC-FTICR technology, the limit is still at $<10^6$ which is far from 10^{10}, showing that current technologies cannot fully penetrate into the deep proteome. Many researchers keen to find ways to reduce the dynamic range of proteins and the most common way is to deplete those high abundance proteins. However, depletion of high abundance proteins may cause reduction of protein signals and loss of potential proteins of interest. Liotta *et al.* pointed out that low molecular weight (LMW) proteins could stay in blood circulation by binding to large carrier proteins. If it is true, existing depletion methods will discard the high abundance proteins together with potential biomarkers and fail to capture this valuable resource. [31]

Many studies used plasma or serum as the starting material. Systematic evaluation done by HUPO Plasma Proteome Project (HUPO PPP) showed that about 40% of all detected signals were unique in serum and cannot be found in plasma. These proteins include intracellular, coagulation dependent and enzymatic activity-derived peptides. [68] Moreover, it is well known that proteomics is easily subject to sample

bias. Proteomic patterns can be changed upon sample collection, processing, storage and number of freeze-thaw cycles. HUPO PPP reported that multiple peak differences were found in samples stored at freezing, refrigerated (4°C) and room temperature conditions for 2 months. [68] Baumann *et al.* also pointed out that unfrozen serum showed great variations in protein profiling, giving poor reproducibility. In addition, freeze-thawing markedly decreased protein intensities especially in low mass range (m/z 1500-3300). [69] Another important issue is the presence of protease in blood samples. No review concerning about the origin of proteins or peptides present in blood, whether they are generated in vivo or ex vivo. Marshall *et al.* illustrated that there were differences in proteomic pattern between plasma samples with and without serine-centered protease inhibitor, phenylmethylsulphonyl fluoride (PMSF). Serum sample treated with 1mM PMSF produced an obvious reduction in the high mass range and more peaks were observed in the low mass range. Marshall *et al.* concluded that the action of serine centered protease contributed to the proteomic profiling pattern ex vivo. [70] However, Ayache *et al.* illustrated that the dramatic changes in proteomic pattern is independent of protease inhibitor. They demonstrated that protein distribution was altered by the processing temperature instead. Remarkable variation of proteins was observed in samples left at room temperature for 2 hours and was due to cytokine production and/or released by leukocytes and platelets. [71]

Comparative proteomics is the common strategy used in biomarker discovery. It was done by comparing proteomic pattern between disease and control groups. However, it is difficult to find a normal applicable to patients with different ages, genders, ethnic and menstrual cycles. The "normal" sample is very important in comparative study; it

directly influences the interpretation of the proteomics patterns between control and disease groups.

Other challenge is patient heterogeneity, including gender, age, menstrual cycles and medications. It is well known that they all contribute changes in proteome patterns but no extensive and systematic evaluation was carried out to illustrate how much they really affect the proteome content in different patients. This uncertainty hinders the development of biomarker discovery.

Current proteomics technologies in biomarker discovery

Current proteomics technologies used in biomarker discovery can be classified into two categories, gel based and gel-free based approaches.

Gel based

Though many chromatographic and electrophoresis-based technologies are available for protein separation, two-dimensional polyacrylamide gel electrophoresis (2-D PAGE), first introduced in 1975, is still the benchmark and routinely used in proteomics. [72] The principle of 2-D PAGE is to separate protein mixtures on a gel in two dimensions, charges and masses. Proteins are first separated according to their isoelectric points (pI) using isoelectric focusing (IEF) and then by their molecular weights using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Separated protein spots are visualized by staining the gels with coomassie blue, sliver or fluorescent dye such as SYPRO Ruby and Flamingo, depending on the abundance of proteins on the gels. [73] The detection limit of different dyes ranges from 0.5 ng to 25

ng. SYPRO Ruby and Flamingo show a linear dynamic range over three orders of magnitude and have higher sensitivity than others but they need expensive laser-based excitation scanner systems for gel visualization whereas coomassie blue has the least sensitivity and poor dynamic range. [74] The differential protein spots are then excised, destained and performed in-gel digestion with trypsin. After extraction, the peptide mixtures are spotted on the target plate and analyzed by Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF-MS). With the use of peptide mass fingerprinting (PMF) and tandem MS technology, the protein identity of the protein spot can be obtained with high accuracy. Wang *et al.* used 2-DE to investigate the global proteome responses of liver-derived cells to HBV infection and treatment response. By comparing two cell lines before and after interferon alpha treatment, 6 down-regulated and 11 up-regulated protein spots were identified. Identified proteins include vinculin, calumenin and prohibitin. [75] Spano *et al.* reported that 46 differentially expressed proteins were identified in HBV transgenic and nontransgenic livers at the early stage of liver fibrosis by 2-D DIGE. [76]

The major merit of 2-D PAGE is that it is capable of separating thousands of proteins simultaneously which cannot be achieved by other methods effectively. In the first dimension, proteins can be separated in different ranges of pH (i.e. pH 3-10, 3-10NL, 4-12, 5-8). To increase the resolution in particular pH, narrow pH range or zoom IEF can be used. Narrow pH range in IEF is important for avoiding or minimizing spot overlapping on the gels, a well-known problem in 2-D PAGE. Large gels using in 2nd dimension can also increase the resolution dramatically as reported by Wildgruber *et al.* [77] However, it is a time-consuming procedure and requires high technical expertise to

carry out the separation. In addition, this low-throughput method also greatly suffers from gel-to-gel variation, making it unsuitable for comparative proteomics analysis. A modified technology, 2-D differential gel electrophoresis (2-D DIGE), was developed to increase reproducibility, speed and sensitivity of 2-D PAGE. 2-D DIGE is based on measuring three samples on one gel, labeling with different cyanine dyes (Cy2, Cy3 and Cy5). By measuring specific wavelength using fluorescent scanner, each labeled proteins can be visualized and compared with the other two. It greatly increases reproducibility; yet, 2-D DIGE and 2-D PAGE also share limitations that they are not applicable for hydrophobic proteins which are insoluble for protein separation. Furthermore, proteins with molecular weight less than 5,000 Da cannot be analyzed and identified. Due to its low throughput and manually operated nature, gel-free approaches have been developed to examine the whole proteome thoroughly and effectively. [78] [79]

Gel free approach – MS based

Upon 2-D PAGE, mass spectrometry (MS) has gained the popularity and become an indispensable tool for proteomics. With the combination of chromatography columns and MS, MS based gel free approach can detect many low abundance proteins which are usually masked by numbers of high abundant proteins. The primary applications of MS to proteomics are cataloging protein expression, defining protein interactions and identifying protein interaction sites. [80] Mass spectrometer consists of an ion source to ionize the analyte molecules, mass analyzer to separate ionized analytes on the basis of mass to charge ratio (m/z) and a mass detector to count the ions at

particular m/z value. Currently, two soft ionization techniques, MALDI and electrospray ionization (ESI), are widely used for analyzing peptides and proteins. With the use of these two ionization techniques, affinity based technologies such as Surface Enhanced Laser Desorption/ Ionization (SELDI) and liquid chromatography (LC) become major tools in clinical proteomics.

SELDI-TOF MS

Surface Enhanced Laser Desorption/ Ionization Time-of-Flight Mass Spectrometry (SELDI-TOF-MS) technology was developed by Hutchens and Yip. [81] The principle of SELDI technology is similar to classical adsorption-desorption mechanism in column chromatography. It is an affinity-based MS technology, selectively capturing subset of proteins based on surface chemistries (i.e. hydrophobic, ion exchange, metal chelate) or antibodies. [82] This technology enables the analysis of small proteins and peptides with mass between 1,000 Da and 20 kDa. It allows direct mass spectrometric analysis of the retained proteins/peptides on the array [83] by MALDI-TOF MS. Besides capturing low mass of proteins, its non-invasiveness and high-throughput manner make it possible for fast screening for novel biomarkers. [84-86] It is capable of analyzing wide range of samples such as serum, plasma, cell lysate, cerebrospinal fluid [87] and urine. [88] It is an easy, effective and sensitive approach of protein profiling for identifying biomarkers especially from crude samples. Its sensitivity can be up to femtomolar range. [83] On the other hand, SELDI technology is capable of detecting protein variants. A study of renal cancer patients found multiple variants of serum amyloid alpha in patient serum but not in healthy control. [89]

Jayanthi *et al.* reported that 3 apolipoprotein C-III isoforms were found by SELDI ProteinChip and was upregulated in heavy marijuana users. [90]

As each protein has a unique m/z value and physicochemical properties, they can be detected precisely and quantified in patients' fluids using specific ProteinChip arrays without knowing their protein identities. However, further characterization of cancer- or disease-associated proteomic features not only enables us to understand the disease pathology, but also allows us to develop a simple immunoassay to measure potential biomarkers. [91] However, the main drawback of ProteinChip SELDI technology is the interested protein or potential biomarkers cannot be recovered from the ProteinChip. Hence, independent experiment is needed to carry out for protein identification which makes the procedures tedious and time-consuming. Protein identification is usually the bottle-neck in most of the biomarker studies employing the ProteinChip SELDI technology.

Another drawback, reproducibility, has been controversial in ProteinChip SELDI technology. [85] Because sample binding steps in this technology is usually operated manually and the serum profiling is sensitive to small changes in the operating procedure, large variation between batches or different centers makes this technology less reliable. Diamandis pointed out that different patterns and biomarkers were identified by various research groups using the same types of biological specimens and the same analytical platform. [92] Moreover, hindrance effect occurs during the binding reaction as small, information rich proteins needs to compete with large, high abundance proteins such as albumin on a small binding surface. Yet this hindrance problem cannot be solved by depletion of high abundance proteins. The overall peak intensities were

decreased after albumin depletion as reported by Cheng *et al.* [84] It is believed that some albumin-binding proteins were also depleted out with albumin. [93]

Ward *et al.* showed that two SELDI peaks, identified as κ and λ immunoglopbulin light chains, were up-regulated in patients with hepatocellular carcinoma in HCV related cirrhosis. [94] Gogel *et al.* identified 5 protein markers (m/z 2,873, 6,646, 7,775, 10,525 and 67,867 Da) to be used to differentiate liver cirrhosis and hepatocellular carcinoma in patients with CHC. One protein peak, m/z 6,646, was identified as apolipoprotein C-I. [95]

LC-MS

An alternate protein separation technology is 1-D or 2-D liquid chromatography (1-D or 2-D LC). Its high sensitivity and specificity allows the analysis of biological molecules in a complex mixture. LC-MS has become an instrument routinely used for protein separation and identification in proteomics laboratory. By using different chromatographic columns, whole proteome in biological sample can be fractionated into less complex fraction, making it possible for the detection of low abundant protein and the deeper proteome. [64] LC-MS is a bottom-up strategy which involves separation and digestion of proteins followed by PMF. LC can be linked to different ion sources and ion analyzers, providing different platforms for variety uses. Examples of ion sources found for proteomics include MALDI, electrospray ionization (ESI), nanospray ionization (NSI) and atmospheric pressure chemical ionization (APCI). Ion analyzers can be TOF, quadurpole IT (QIT), fourier transform ICR (FT-ICR) and quadruple mass filter (QMF). Due to the high complexity of sample, many proteomics studies are limited to identify high abundant proteins, those low abundant proteins, potential

biomarkers for human diseases, are still the challenge in biomarker discovery. Tucholska *et al.* demonstrated the feasibility of using LC-ESI-MS/MS to identify low abundance proteins in complex blood sample. By using multiple partition chromatography resins, many polypeptides can be purified to nearly homogeneous from serum or plasma samples. Enrichment of low abundance proteins can be achieved by depletion of high abundance proteins using custom-designed or commercially available depletion columns. [96] Yang *et al.* demonstrated that after the depletion of HSA and IgG, low abundance protein like vitamin-K-dependent clotting factors and inhibitors were identified by LC-ESI-MS/MS. [97] Gao *et al.* also showed that after the depletion of 58 high abundance proteins, the number of identified proteins was 2.7 times as that of the nondepletion method under LC-MALDI-TOF/TOF MS platform. [98]

For shotgun proteomics, the multidimensional protein identification technology (MudPIT), developed by Yates and colleagues, is now widely implemented to analyze complete cell lysates, tissue extracts and other subproteomes. [99] The protein mixtures are directly digested into peptides which are then separated by on-line multidimensional chromatography followed by tandem mass spectrometric analysis. The digested peptide mixtures are separated in strong cation exchange (SCX) column followed by reversed phase (RP) column. The eluted fractions are then directly passed into the mass spectrometer for protein identification. This system can separate and identify 1,484 proteins for S. cerevisiae proteome, comprising proteins with extreme pI, MW, abundance and hydrophobicity. [100] It is a fast, sensitive method for analyzing complex peptide mixtures prior to 2-D PAGE as it is an online process. The main drawback is that only one sample can be analyzed in single run, the throughput is low.

Furthermore, as digestion is performed before separation of protein mixture, it generates a more complex peptide mixture than the original one and makes the separation more challenging. In serum, there are at least 200,000 peptides coexist after digestion, effective separation on SAX-RP system may not be achieved and more sensitive and effective separation technique is needed. [101] Some information may lose upon the digestion of proteins to peptides which may result in false identification, especially those with diverse or unexpected modifications. In addition, the suppression effect by abundance difference in mass spectrometer may restrict the probability of identifying medium and low abundance proteins especially when high abundance proteins are co-eluted with medium and low abundance proteins. More advanced and highly sensitive mass spectrometers such as triple quadrupole MS, LTQ FT and LTQ-orbitrap can be used to allow more accurate mass measurement and peptide fragmentation schemes.

Apart from shotgun proteomics approach, a newer strategy known as top-down approach was developed for studying intact proteins. This top-down approach does not need the purification and digestion of proteins for analysis. It requires high-resolution mass measurement of intact proteins and their direct fragmentations with the use of very expensive instruments such as Orbitrap or FTICR only. A key to this approach is the ability to fragment the intact proteins, it can be performed by using electron transfer dissociation (ETD) or electron capture dissociation (ECD) techniques. Using this dissociation techniques, small to medium-sized proteins can be fragmented while it is incapable of analyzing large proteins (>50kDa) because of increasing complexity between the protein ions' tertiary structure and many nonconvalent interactions. It is believed that top-down approach can be a tool for sequencing peptides and determining

the post-translational modification sites which cannot be done by bottom-up approach. [102] [103] However, since it is performed with a single protein or simple protein mixture fractionated off-line, the analytical throughput and efficiency for large-scale proteome analysis is still a challenge. Recently, Bryan *et al.* demonstrated proof-of-concept for high throughput top-down proteomics made from the measurement of high-resolution MS/MS fragment ions in an online fashion. [104] With the integration of offline weak anion exchange (WCX) column and online LC-LTQ FT, 22 proteins were identified on yeast whole cell lysate.

1.3.4.3 Quantitative proteomics

The major progress in proteome study is quantitative proteomics. It is important to obtain quantitative information from healthy and disease group for biomarker discovery. Stable isotope and molecular labeling methods are currently used for quantitative proteomics. Because the labels are chemically identical, the peptide pair (light and heavy peptides) will behave identically in terms of chromatographic retention and ionization efficiency, allowing samples to be analyzed and compared simultaneously.

Stable isotope labeling by amino acids in cell culture (SILAC) is a common stable isotope labeling technique used in cell culture. [105] Two populations of cells are cultivated in different culture medium. One of the populations is fed by stable isotope-labeled essential amino acids while the other is fed by normal one. The stable isotope will then be absorbed and secreted by cells in the synthesis of proteins in vitro. ^{13}C is usually chosen for labeling, $^{12}C_6$ and $^{13}C_6$ lysine is labeled in HCC cell lines with low

and high metastatic potentials for quantitative analysis respectively. [106] Peptide pair was separated by 6 Da corresponding to the mass difference between $^{12}C_6$ and $^{13}C_6$ isotopes.By measuring their intensities ratio, differential proteins can be found and identified. Nearly all peptides can be isotopic labeled by SILAC, greatly improving the sequence coverage of proteins. However, it is not practical for clinical protein samples in vivo.

Isotope-coded affinity tag (ICAT) was introduced in 1999 by Aebersold and co-workers. [107] The ICAT reagent consists of a reactive group specific for free thiol functionality of cysteine residues, a polyether linker region with deuteriums or ^{13}C with their light forms (i.e. ^{1}H and ^{12}C) and a biotin tag that allows recovery of labeled peptides. In ICAT experiment, the light and heavy isotopic tags bind covalently to cysteine moieties of amino acids within proteins. The proteins are then eluted from an avidin column and quantified with MS. The relative amount of the identified peptides can be calculated by the ratio between heavy and light forms. Because only cysteine-containing peptides are isolated, complexity of the sample will not affect quantitative analysis and it allows detection of low abundance peptides. This technology has been applied to study proteome in mammalian cells and human liver cells. Yet, many important proteins which do not have cysteine cannot be detected. [108] About 300-400 proteins per sample can be analyzed by ICAT, far less than that with 2-D PAGE technology. Also, deuterium affects peptide retention time in LC and intense fragment ions from biotin moiety makes the complicated MS/MS spectra not suitable for database searching.

Isobaric tag for relative and absolute quantization (iTRAQ) is increasingly accepted in secretome analysis. iTRAQ labels consist of reactive groups label with all free amines, at the N-terminus of all peptides and also the side chain of internal lysine residues, a balance group and a reporter group . [109] Unlike ICAT, they target all tryptic peptides and therefore enhance the depth of the coverage. There are currently two types of reagents: 4-plex and 8-plex. Each label has an identical mass but produces different fragments during collision-induced fragmentation in MS/MS analysis. Mixed differentially labeled peptide mixtures at equimolar concentration, separated by 2-D chromatography and subjected to tandem MS/MS analysis, a special mass signal at m/z 114.1, 115.1, 116.1 and 117.1 can be obtained from the fragmentation between the reporter and balance groups. Each signal correlates directly to the amount of peptides present in the sample, thus protein concentration can be deduced. [110] The advantage of iTRAQ technology is that different tags will not affect the separation performance which makes it possible for analyzing at least four samples in single run. However, it suffers from some limitations. Fragmentation efficiency depends on charge and sequence, small amount of co-eluted labeled fractions may have similar m/z fragment ions that can lead to significant error in reporter ion distribution.

The MS-based technique of multiple reaction monitoring (MRM) is now being explored for validation of protein biomarkers in clinical samples. It is not a new technique for small molecules which has been used for quantification of metabolites in drug metabolism studies over 30 years. Due to the emergence of new MS instrumentations, MRM is now capable of analyzing large molecules such as proteins and peptides. [111] For low molecular weight analysis, triple quadrupole mass

spectrometer enables the precursor ion (parent ion) to be selected in the first quadrupole, followed by fragmentation by collision with inert gas atom in the second quadrupole and finally passed to the third quadrupole for analyzing selective fragment ion (daughter ion). In principle, this technique is able to confirm the identity of the precursor ion by fragment ion in a specific transition. Moreover, in combination of stable isotope-labeled internal standards (SISs), absolute quantitation can also be obtained. [112] For protein and peptide analysis, a triple quadrupole-linear ion trap (QqQ-LIT) instrument like Q TRAP® has advantages. Q TRAP® has a high sensitivity that is able to detect low abundance analytes in complex matrices. Several recent proof of principle studies showed that MRM can be identified and quantified proteins at ng/mL level. Keshishian *et al.* showed that 6 spiked proteins in immunodepleted human plasma can be quantitated in the 1-10ng/ml range with coefficient of variation ranges from 3 to 15%. [113] Kay *et al.* also demonstrated that in combination with ACN precipitation, insulin-like growth factor-I (IGF-I) can be detected in MRM experiment which has the concentration of 100ng/mL in human serum. [114] Besides, Q TRAP® is able to switch rapidly between triple quadrupole and linear ion trap mode, making it possible to carry out MRM-initiated detection and sequencing (MIDAS). MIDAS, first developed for the identification of phosphorylation site on the yeast cell cycle proteins, consists of standard MRM experiment and a full-scan product ion to confirm the identity of the parent ion, especially for peptide post-translational modifications. [115] Positive MRMs trigger an MS/MS experiment to confirm the nature and modification site. It is a more sensitive approach due to efficient duty cycle and reduced background signal. Although MRM technique shows promising result in the quantitation limit, it still faces lots of

limitations that hinder the development of this technique. Detection of low abundance proteins which is always masked by high abundance proteins remains the main challenge for MRM. It remains unclear the potential of MRM to detect and quantify trace amount of protein in a "sea" of high abundance proteins which cover about 99% of the total protein. Though there are several depletion methods for enrichment of low abundance proteins, it showed that depletion couldn't increase the number of peptide transition detected. [116] In addition, designing and validating individual peptide transition is a major bottleneck of MRM. In *silico* methods are available to derive the possible MRM transition based on the theoretical rules and empirical observation of the target peptide, however, Anderson *et al.* showed that by using in *silico* method to derive peptide transition, only 11 of 30 proteins were identified, indicating that in *silico* method is not reliable enough for transition determination. [116]

1.3.5 Applications of proteomics to discovery of biomarkers for diagnosis of liver fibrosis

Only few studies used SELDI ProteinChip technology to find biomarkers for the detection of HBV-associated liver fibrosis or cirrhosis. Zhu *et al.* found two markers, 7,772 and 3,933 *m/z*, for detection of HBV-induced liver cirrhosis using SELDI technology. Two markers were found to be down-regulated in liver cirrhosis group. The decision tree model achieved good accuracy (80% sensitivity and 81.8% specificity) with 75% positive predictive value. [117] Another study conducted by Cui *et al.*, two differential protein peaks (2,050 and 3,166 *m/z*) were found for detection of CHB with 100% sensitivity and 86.5 % specificity. [118] Poon *et al.* reported 30 protein features

together with several routine laboratory parameters were included in artificial neural network (ANN) model. With a cut-off of 3.0, the ANN model attained good accuracy (96% sensitivity, 89% specificity) for detecting liver fibrosis. [119] Though these studies reported models with good diagnostic accuracy in their study group, the identities of these protein markers were still unknown and the calculation of these models were so complicated for clinical use.

Rationale and Objectives of the Project

There is a lack of sensitive and specific serum markers for diagnosis of liver fibrosis. Our recent pilot study showed that serum proteomic fingerprint identified by the ProteinChip SELDI technology might allow diagnosis of liver fibrosis in CHB patients with high accuracy. Unfortunately, the non-recovery nature of the ProteinChip SELDI prohibited the identification of the proteomic features forming the diagnostic fingerprint. Hence it was important to develop a new technology comparable to the ProteinChip SELDI, but it allows purification of the proteins corresponding to the proteomic features in parallel.

The objectives of this research project are:

1 To develop a magnetic beads-based proteomic profiling method for quantitative proteomic profiling and micro-purification in parallel;

2 To develop a proteome-based fingerprinting model for detecting liver fibrosis in patients with chronic hepatitis B infection

3 To uncover the protein identities of proteomic features forming the diagnostic fingerprint

Section 1 Method development of magnetic beads-based proteomic profiling for quantitative proteomic profiling and micro-purification in parallel

1.1 Introduction

High throughput quantitative serum profiling method is a useful tool for biomarker discovery and is a potential blood test for disease diagnosis/prognosis. By comparative proteomic approach, new disease- associated biomarkers can be discovered for clinical use. 2-D PAGE was a traditional technique used in biomarker discovery. Though it can effectively separate hundreds to thousands of proteins on the gel, low throughput and lack of reproducibility are still the main limitations on 2-D PAGE which makes it unsuitable for high throughput screening.

In the past 10 years, there was an overwhelming interest in using SELDI technology in biomarker discovery. It is a variant of MALDI-TOF MS used to retain a subproteome on the functionalized affinity surface. Proteins retained on the SELDI chip are then directly analyzed by MALDI-TOF MS to generate a proteomic profile. By using statistics methods, differential proteomic features can be found for the underlying diseases. Though biomarker can be reproducibly detected and quantified by the SELDI technology even not knowing the protein identity, for clinical use, knowing the protein identity is indispensable so that the protein-protein interaction or disease pathologies can be understood. The main drawback of the SELDI technology is the retained proteins cannot be recovered for further protein identification and analysis. Independent time-consuming small-scale chromatography needs to be performed to purify the corresponding proteins for subsequent protein identification work. Protein identification

is usually the bottle-neck in most of the biomarker studies employing the ProteinChip SELDI technology. In addition, there are several limitations of the SELDI technology, for example, high cost, high susceptibility to identify false-significant biomarkers due to systemic bias and sensitive to small changes of experimental procedures and analytical variables. [120] Also, competition between large and small proteins occurs during sample binding on a small chromatographic surface.

On the other hand, instrument performance plays an indispensable role in quantitative serum profiling. The instrument sensitivity greatly affects the possibility of detecting low concentration of proteins on the spot. In addition, stable instrument performance can give reproducible MS spectra, reducing variations during MS analysis.

In this section, we aimed to develop a magnetic bead-based proteomic fingerprinting method which was comparable to ProteinChip SELDI technology, but allowed quantitative proteomic profiling and microscale purification of the profiled proteins in parallel. The performance of the instrument, reproducibility of the proteomic profile and the feasibility of protein identification of the MS peak were examined.

1.2 Materials and methods

1.2.1 Biological specimen

Blood was collected from a healthy volunteer with consent. The blood was clotted at room temperature for 30 minutes, and centrifuged at 3000xg for 15 minutes. The serum was stored in 10 μL aliquots at -80 °C before analysis.

1.2.2 Automated platform for protein capture with magnetic beads

To improve reproducibility and achieve a high throughput, the capture procedures were automated by using the KingFisher 96 magnetic particle processor (ThermoFisher), which could process 96 samples per run. Figure 1 showed the schematic diagram of the whole magnetic bead-based serum profiling procedure.

1.2.3 Capture of hydrophobic serum proteins

2μL serum was diluted and inactivated with 198μL of binding buffer (0.9% sodium chloride (NaCl) containing 0.1% trifluoroacetic acid (TFA)). 0.5 mg of C18 magnetic beads in 10 μL (Dynabeads RPC18, Invitrogen) were pre-washed sequentially with 90μL of pure acetonitrile (ACN) and 100μL of binding buffer. 80μL of diluted serum samples were mixed with the washed C18 beads for 10 minutes. The C18 beads were then washed 3 times with 100μL of 0.2% TFA. Finally, the captured proteins were eluted with 60μL of universal elution solution (50% ACN containing 0.2% TFA).

1.2.4 Capture of anionic serum proteins

2μL serum was first denatured by adding 8μL of UC buffer (9M Urea, 2% CHAPS, 2.5mM Tris-base, 10mM Tris-HCl). After incubation at room temperature for 30 minutes, the denatured serum samples were diluted with 190μL of pH 8 phosphate

binding buffer (20 mM sodium phosphate, 30% ethanol, pH 8.0). 0.5 mg of SAX magnetic beads in 10 μL (Dynabeads SAX, Invitrogen) were mixed with 60μL of 1.5 M NaCl and 30 μL of ethanol for prewashing, and then washed with 100 μL of pH 8 phosphate binding buffer. 80μL of diluted serum samples were mixed with the washed SAX beads for 10 minutes. The SAX beads were then washed with pH 8 phosphate binding buffer, followed by 3 washes with washing buffer (2.3μM ethanolamine, 30% ethanol, pH 9.0). Finally, the anionic proteins were eluted with 60μL of universal elution solution.

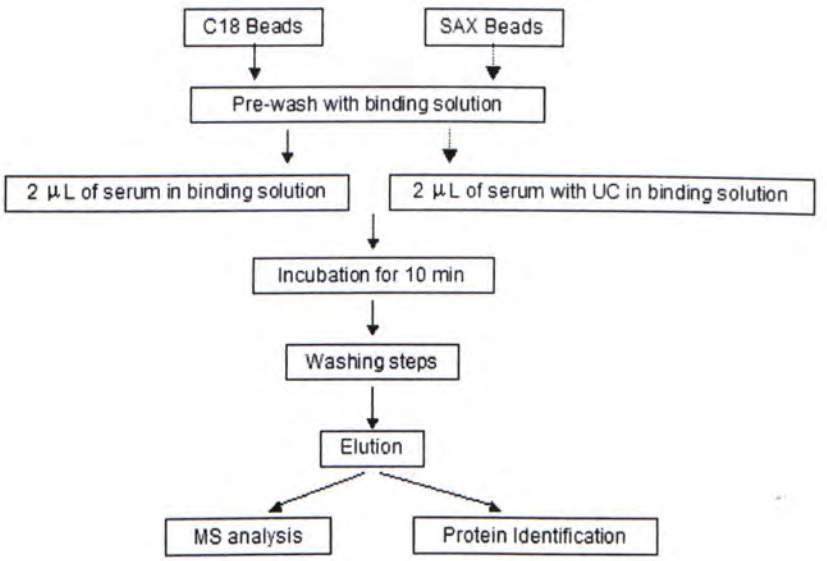


Figure 1. Schematic diagram of the magnetic bead-based serum proteomic profiling method for parallel analytical analysis and micropreparative purification. In the C18 assay, serum was diluted and inactivated with 0.9% NaCl containing 0.1% TFA; in the SAX assay, the serum was denatured with urea-CHAPS (UC) buffer. All the binding and washing procedures were performed automatically with a 96-sample magnetic particle processor.

1.2.5 Evaluation of the performance of linear MALDI-TOF MS

The performance of PCS 4000 ProteinChip reader (Bio-Rad Laboratories) was evaluated by using ProteinChip OQ Kit (Bio-Rad Laboratories). Unless specified, all testing peaks were selected on the spectra and peak information was exported and pasted on ProteinChip SELDI OQ form for calculation. ProteinChip detector qualification array was used to test the detector sensitivity of PCS 4000 ProteinChip reader. Two different concentrations (10fmol and 140fmol) of immunoglobulin (IgG) were analyzed and the corresponding signal-to-noise ratio (S/N) was measured. 1 of 1 partition was used for MS analysis. For high concentration, it passed if S/N was greater than 1,000. For low concentration, it passed if S/N was greater than 5. ProteinChip peptide standard array was used for mass drift, resolution and mass accuracy tests. 1 of 4 partitions was used for MS analysis for these three tests. Mass drift and resolution was measured on insulin peak (5.96 kDa). It passed if the mass drift was less than 7 Da and average resolution was greater than 750. Resolution was further assessed on arg-8-vasopressin peak (1.08 kDa). It passed if the resolution was greater than 1,000. Mass accuracy of the system's internal and external calibration was tested on arg-8-vasopressin (1,084.247 Da), somatostatin (1,637.903 Da), dynorphin A (2147.5 Da), ACTH (2,933.5 Da), beta endorphin (3,465 Da), arg-insulin (5,963.8 Da) and cytochrome C (12,230.92 Da). For external calibration, it passed if the average mass within 0.1% of calibrant mass and pooled CV of <0.05. For internal calibration, it passed if the average mass with 0.01% of calibrant mass and pooled CV of <0.01.

1.2.6 Quantitative serum proteomic profiling by linear MALDI-TOF MS

1μL of eluate was spotted onto an 8-spot gold ProteinChip array (Ciphergen Biosystems) in duplicate. 0.5μL of sinapinic acid matrix (13mg/mL sinapinic acid dissolving in 50% ACN/ 0.5% TFA) (Sigma-aldrich) was added by using of nano-liter non-contact dispenser (BioDot AD3050) and air-dried in a chamber with humidity control at 80%. Another 0.5μL of sinapinic acid matrix was added and air-dried. The gold ProteinChip array was subjected to the PCS4000 ProteinChip reader (Bio-Rad Laboratories) to determine the masses and intensities of all the peaks over the m/z range from 1,000 to 250,000 m/z. Two acquisition protocols were used for low and high mass ranges. Intensities of peaks between 1,000 and 20,000 m/z were obtained at a laser power of 6,000nJ and the focus mass was 8,000 m/z; intensities of peaks between 10,000 to 250,000 m/z were obtained at a laser power of 10,000nJ and the focus mass was 80,000 m/z. The mass spectra were externally calibrated with a mixture of peptide/protein standards (angiotensin, 1,296.51 m/z; ACTH (clip 1-17), 2,093.46 m/z; ACTH (clip 18-39), 2465.72 m/z; double charged horse apomyoglobin 8475.8 m/z; E. coli Thioredoxin, 11673.5 m/z, horse apomyoglobin, 16951.6 m/z; bovine serum albumin, 66430 m/z; bovine serum albumin dimer, 132861 m/z) (Applied Biosystems). The common peaks among the mass spectra were identified and quantified using the Biomarker Wizard software (Bio-Rad Laboratories). The peak intensities were normalized with the total ion current, and subsequently with the total peak intensities.

1.2.7 Reproducibility analysis of magnetic bead-based protein spectra

The peaks with normalized intensities > 0.15% of total peak intensities were used to evaluate the reproducibility of magnetic bead-based mass spectra. Same serum sample was subjected to quantitative proteomic profiling in at least 20 replicates in a single experiment for assessing the intra-assay (i.e. within assay) error. For assessing interassay (between assays) error, same serum sample was subjected to quantitative proteomic profiling for at least 15 times on different days. For each peak, the intra-assay and interassay coefficients of variation (CVs) of the normalized intensities were calculated according to the standard formula (CV = standard deviation / mean value).

1.2.8 Non-reducing 2-D gel electrophoresis

Eluted protein sample was dried at 45 °C using speedvac concentrator (Enppendorf) and reconstituted with 185μL of rehydration buffer (8M Urea, 2% CHAPS, 0.2% Biolyte 3-10 ampholyte, 0.001% bromophenol blue, 1mM EDTA). An immobilized pH gradient (IPG) strip (11cm 3-10NL, Bio-Rad Laboratories) was rehydrated with the sample overnight. For the first dimension IEF separation, the running condition was as follows: 100V for 10min, 250V for 65min, 500V for 25min, 1000V for 40min, and finally 8000V for 140min. Second dimension SDS-PAGE was performed on 4-12% Bris-Tris polyacrylamide gels (Bio-Rad Laboratories) and the proteins were separated at 200 V for 40min in ice bath. The 2D gel was then stained with silver nitrate using Amersham PlusOne silver staining kit (GE Healthcare) with some modifications to reduce the loss of proteins with MW < 10 kDa. The gel was fixed in 40% methanol/10% acetic acid for 30 min and then sensitized by thiosulfate solution

with 25% w/v Glutardialdehyde. Washed with 30% ethanol for 15 min, the gel was immersed with silver solution in 30% ethanol and 37% w/v formaldehyde for 1 hr. The gel was washed with Milli-Q water for 1min for 3 times and then developed in sodium carbonate solution with 37%w/v formaldehye. The development was stopped by EDTA solution and the gel was rinsed with Milli-Q water for 3 times before performing image analysis with GS-800 calibrated densitometer (Bio-Rad Laboratories).

1.2.9 Protein identification of eluted proteins

Protein spots were excised from silver stained gels. The gel pieces were destained, reduced with 1.75% DTT, alkylated with 350 mM IAA, and digested with modified porcine trypsin overnight (Sequencing grade modified typsin from Promega, Madison, WI). The tryptic digest was harvested and cleaned up with C18 ZipTips (Millipore). The cleaned tryptic peptides were subjected to MALDI-TOF/TOF MS (Applied Biosystems 4700 Proteomic Analyzer, Applied Biosystems) with α-cyano 4-hydroxy cinnamic acid as matrix. The MS and MS/MS spectra were then processed with Data Explorer software (Ver. 4.4; Applied Biosystems). The spectra were subjected to gaussian smoothing with a filter width of 5 points, and the baselines were corrected with default settings. Peaks were detected based on a S/N > 15. The MS spectrum data were searched via the online ProFound search engine to obtain the protein identity by undertaking the peptide mass fingerprinting (PMF) approach. Tandem MS data were subjected to MS/MS ion search via the Mascot search engine to obtain the protein sequence of a particular peptide. For the search parameters, the 1 missed cleavage in trypsin digestion was allowed; partial oxidation of methionine, phosphorylation of

serine/threonine/tyrosine, and iodoacetamide modification of cysteine residues were selected. The error tolerance values of the parent peptides and the MS/MS ion masses were 200ppm and 0.5 Da, respectively. A protein identification result was considered valid when both PMF and MS/MS ion search identified the same protein as the statistically significant hit from the NCBInr database, and/or when MS/MS ion search identified at least tryptic peptides with sequences from the same protein as the statistically significant hits.

1.3 Results

1.3.1 Serum proteome profiles obtained with different types of chromatographic magnetic beads

Two different assays for profiling hydrophobic proteins and anionic proteins were successfully developed by using C18 and SAX magnetic beads respectively. Different chromatographic magnetic beads resulted in different proteomic profiles (Figure 2 and 3). C18 hydrophobic beads captured more proteins in m/z 2000-8000 range than SAX beads. Peak at m/z 2750 was unique in C18 assay. However, there was an exception in this range. Proteins at m/z 6,400 and 6,600 had greater affinities in SAX, giving 1.5 to 2 times of the signal intensities compared with the hydrophobic condition. There was also a unique peak found in the SAX profiling assay at *m/z* 13,850 which cannot be found in C18 profiling assays. However, in high mass range, there was no obvious peak difference between the two profiles. The only difference was SAX beads captured the proteins in greater amount than C18 beads; the overall intensities from SAX profile were about 5 times greater than that of C18 beads.

For the C18 profiling assay, 88 peaks (74 peaks with normalized intensities > 0.15% of total intensities) between 1,600 m/z and 20,000 m/z and 143 peaks (93 peaks with normalized intensities > 0.15%) between 20,000 m/z and 250,000 m/z were observed. For the SAX profiling assay, 65 peaks (54 peaks with normalized intensities > 0.15%) between 1,600 m/z and 20,000 m/z and 74 peaks (46 peaks with normalized intensities > 0.15%) between 20,000 m/z and 250,000 m/z were observed.

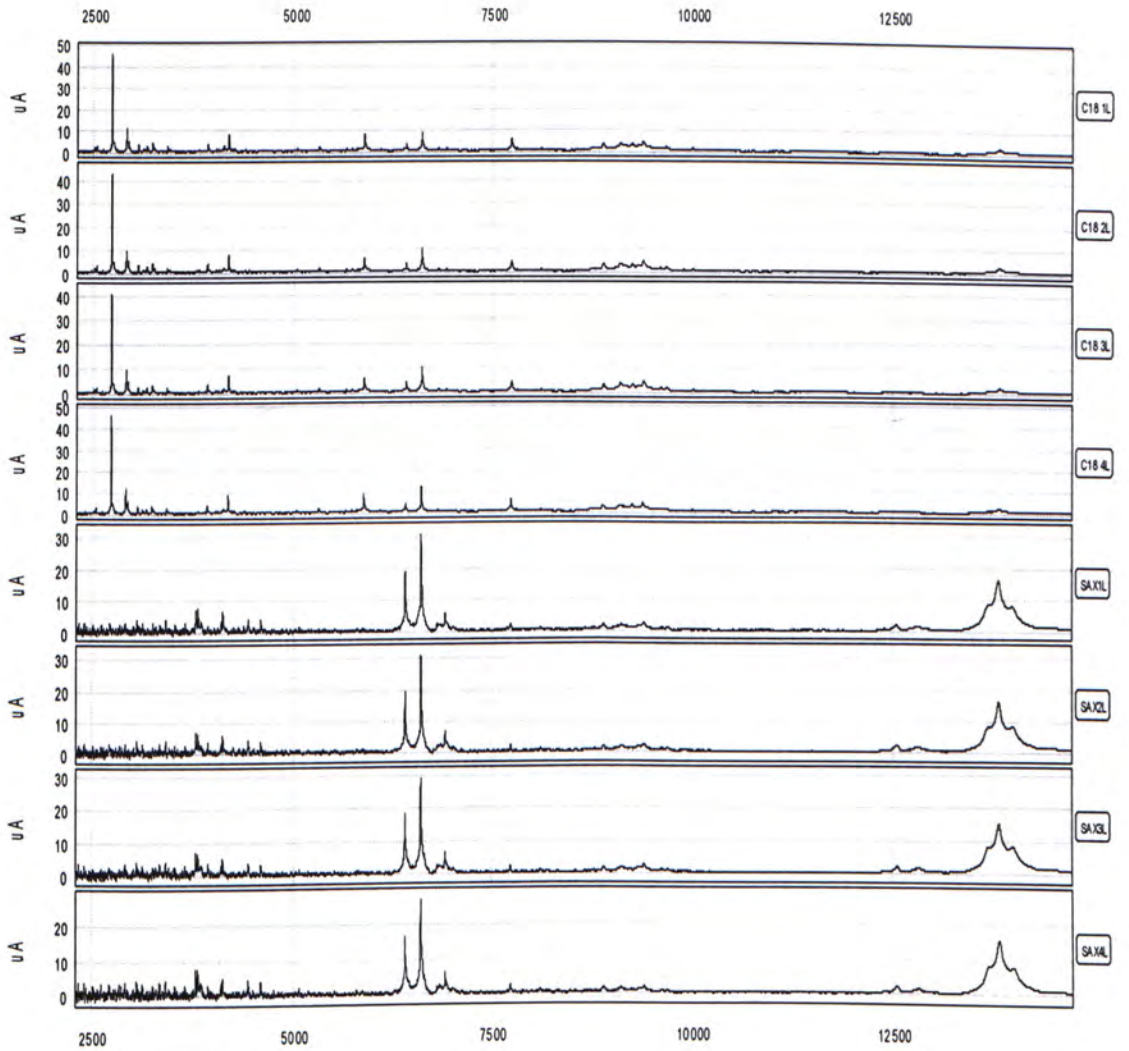


Figure 2. Representative proteomic profiles obtained with different bead types at low mass range (1,000-20,000 m/z). Top: C18 beads; Bottom: SAX beads.

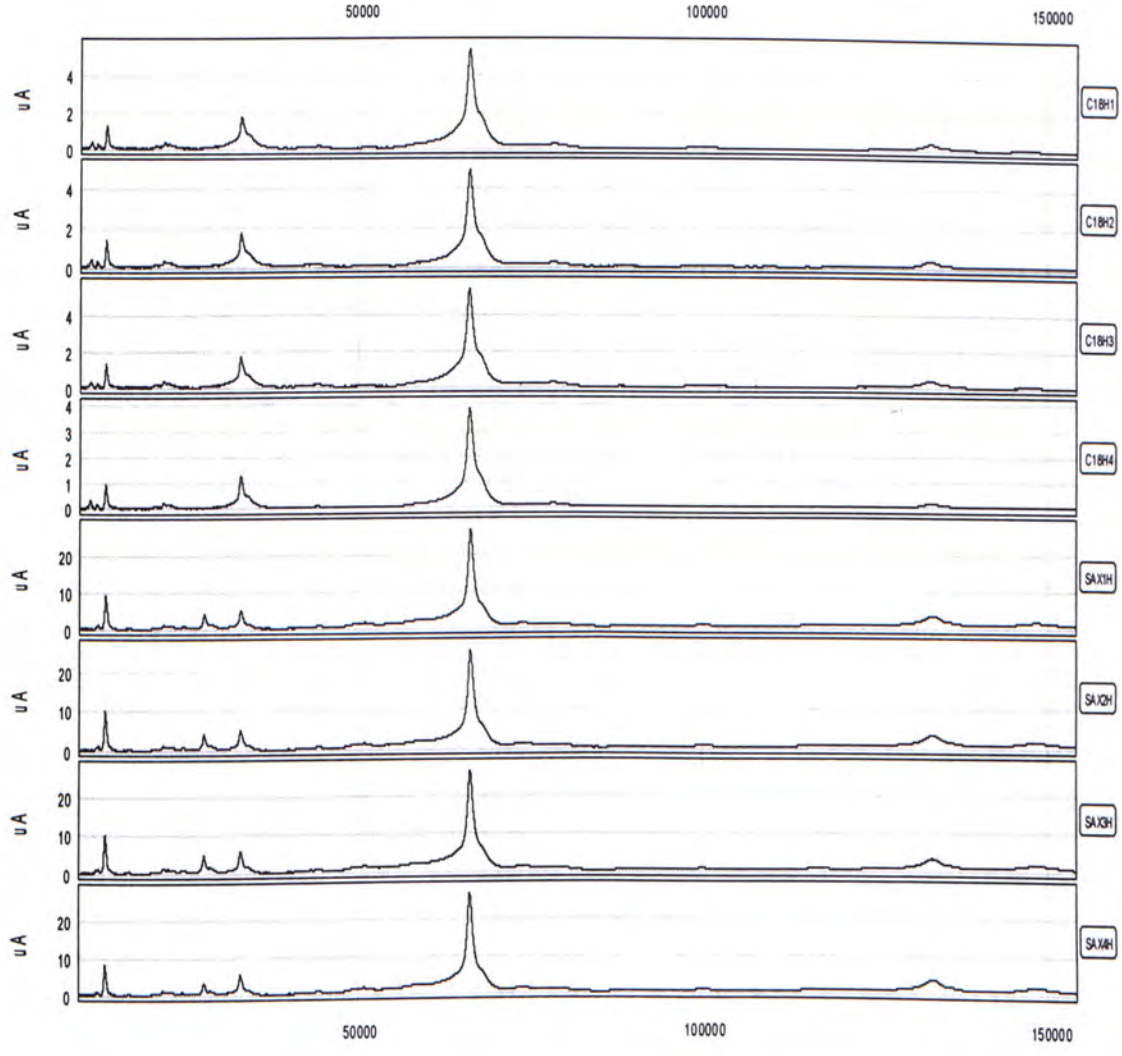


Figure 3. Representative proteomic profiles obtained with different bead types at low mass range (10,000-250,000 m/z). Top: C18 beads; Bottom: SAX beads.

1.3.2 Performance of PCS 4000 ProteinChip reader

All performance tests were passed. For detector sensitivity test, the average S/N for 140 fmol and 10 fmol were 1147.37 and 51.26 respectively. For mass drift and resolution test at 5.96 kDa, the mass drift and resolution were found to be 2.94 and 817.67 respectively. For resolution at 1 kDa, the resolution was 1681.91. For external and internal calibration, the average mass of the calibrants was within 0.03% and 0.01% and the pooled CV was 0.023 and 0.005 respectively. Figures 4 – 7 showed the resulting spectra of each test.

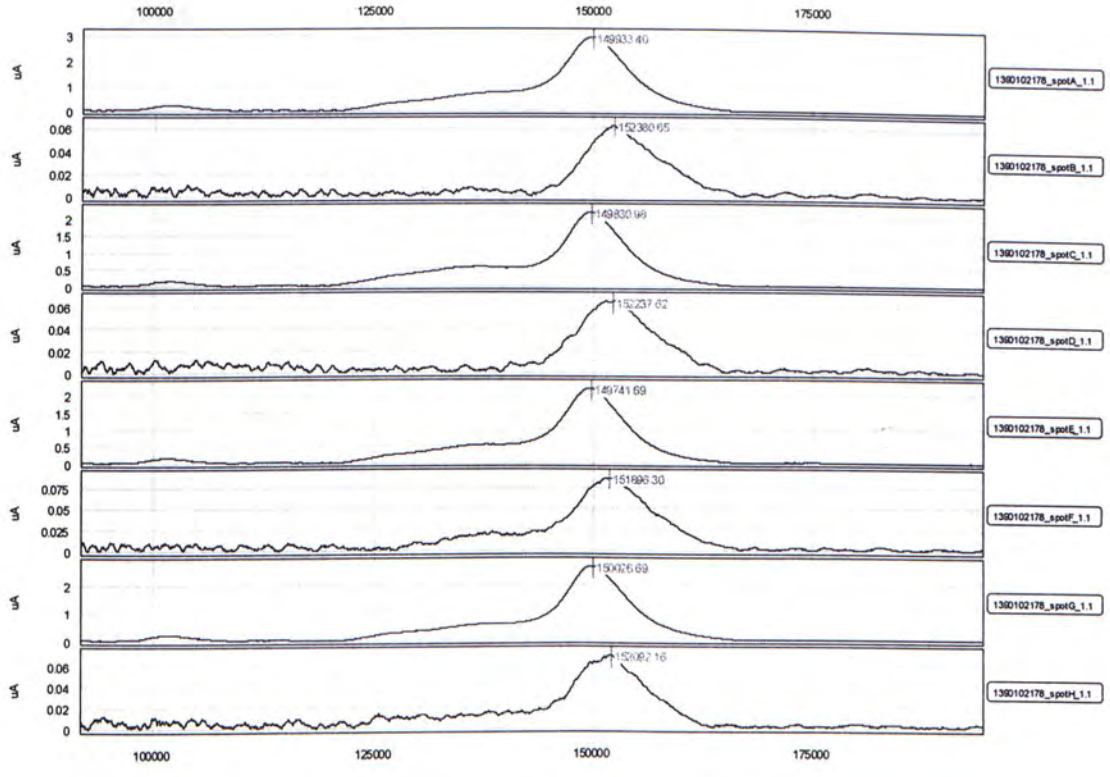


Figure 4. Spectra showing different concentrations of IgG for evaluating detector sensitivity. Spots with high concentrations (140fmol) (spots A, C, E and G) alternate with those with low concentrations (10fmol) (spots B, D, F and H).

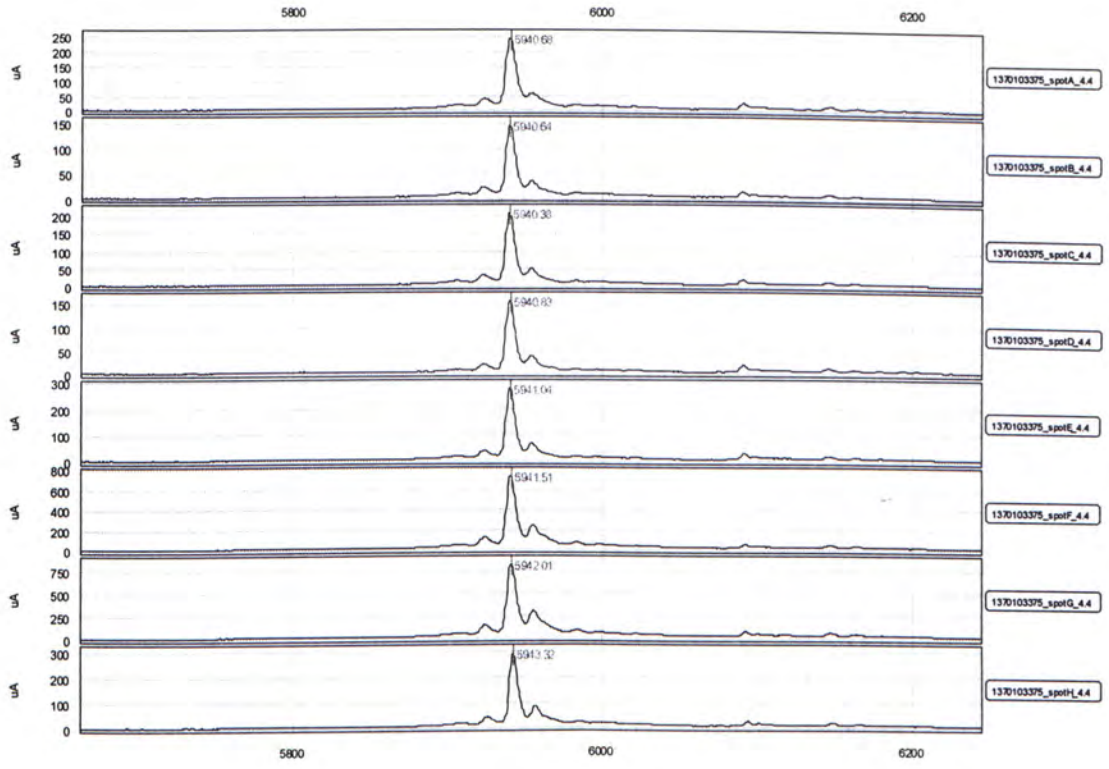


Figure 5. Spectra showing insulin peak (m/z 5,963 Da) for evaluating mass drift and resolution

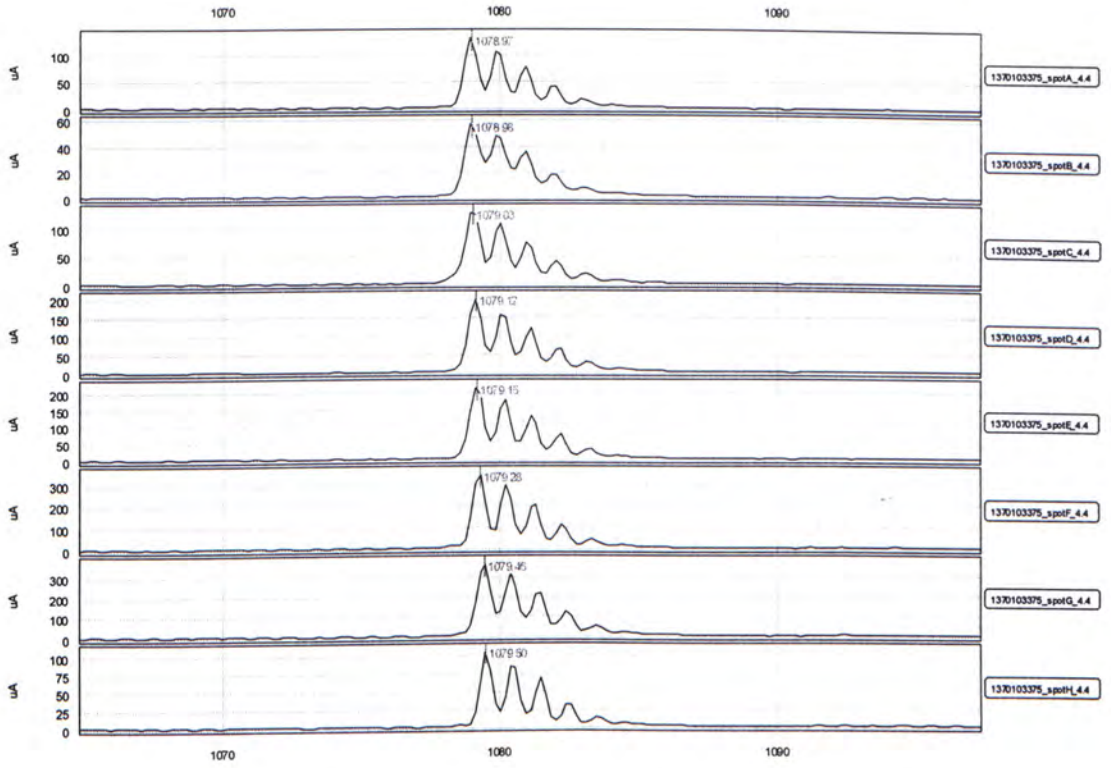


Figure 6. Spectra showing arg-8-vasopressin peak (m/z 1,084 Da) for evaluating resolution

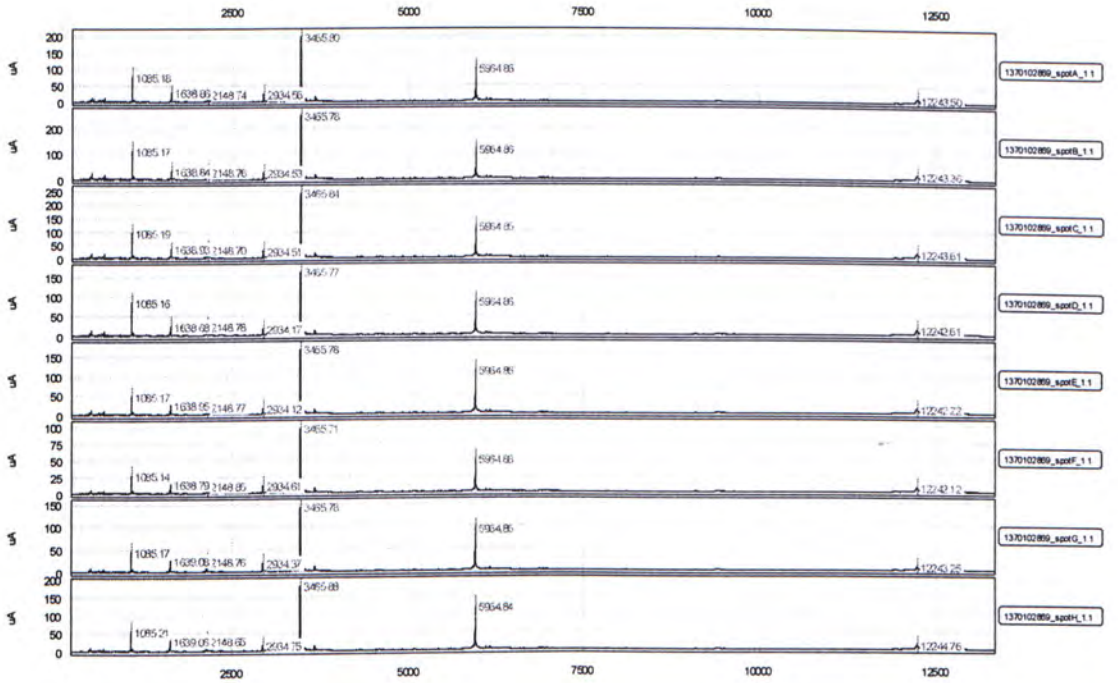


Figure 7. Spectra showing arg-8-vasopressin (1,084.247 Da), somatostatin (1,637.903 Da), dynorphin A (2147.5 Da), ACTH (2,933.5 Da), beta endorphin (3,465 Da), arg-insulin (5,963.8 Da) and cytochrome C (12,230.92 Da) for evaluating mass accuracy

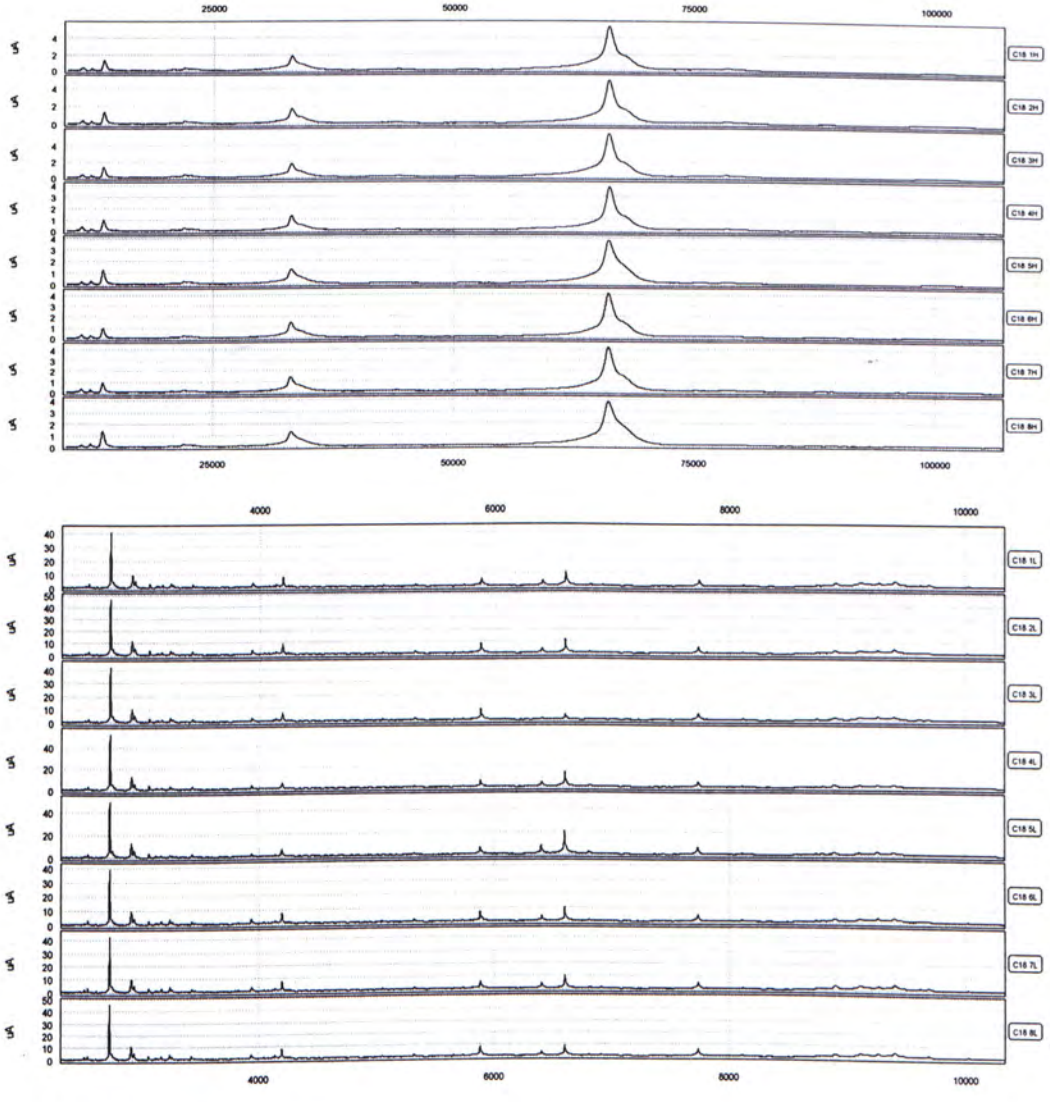
1.3.3 Reproducibility of magnetic beads-based serum proteomic profiling

The reproducibility of each profiling assay was assessed in terms of coefficients of variation of the normalized intensities of the protein peaks. The intra- and inter-assay CVs for the 2 assays were summarized in Table 2. The reproducibility of the 2 assays was similar, but it appeared that the SAX assay had the best performance. The normalized intensities of peaks at higher m/z range appeared to be more precise. The intra-assay and interassay CVs of different peaks within a proteomic profile were similar, and both in the range of 4 to 30%. Figures 8 and 9 showed representative spectra for the reproducibility of this magnetic bead-based approach.

Table 2. Reproducibility of quantitative proteomic profiles by using C18 and SAX beads. Intra-assay and interassay coefficients of variation (CVs) of the normalized intensity of each MS peak were calculated, and mean values of the CVs from multiple peaks were presented.

[a]n – number of replicates
[b]The peaks were ranked from the highest normalized intensities to the lowest.

Bead type	Mass range	Intra-assay CV, % mean (min. – max.)			Interassay CV, % mean (min. – max.)		
		n[a]	Top 10[b]	11th-20th	n[a]	Top 10	11th-20th
C18	1600 – 20,000 m/z	20	24 (18-29)	24 (15-30)	16	25 (21-30)	26 (13-30)
	20,000 – 250,000 m/z		14 (8-21)	22 (14-30)		18 (15-22)	20 (14-29)
SAX	1600 – 20,000 m/z	24	12 (6-21)	12 (6-28)	24	20 (16-29)	24 (18-30)
	20,000 – 250,000 m/z		8 (4-17)	8 (6-13)		14 (7-19)	16 (10-22)



HMW

LMW

Figure 8. Representative spectra obtained with C18 beads. Top: High molecular range (*m/z* 20,000-250,000) Bottom: Low molecular range (*m/z* 1000-20,000)

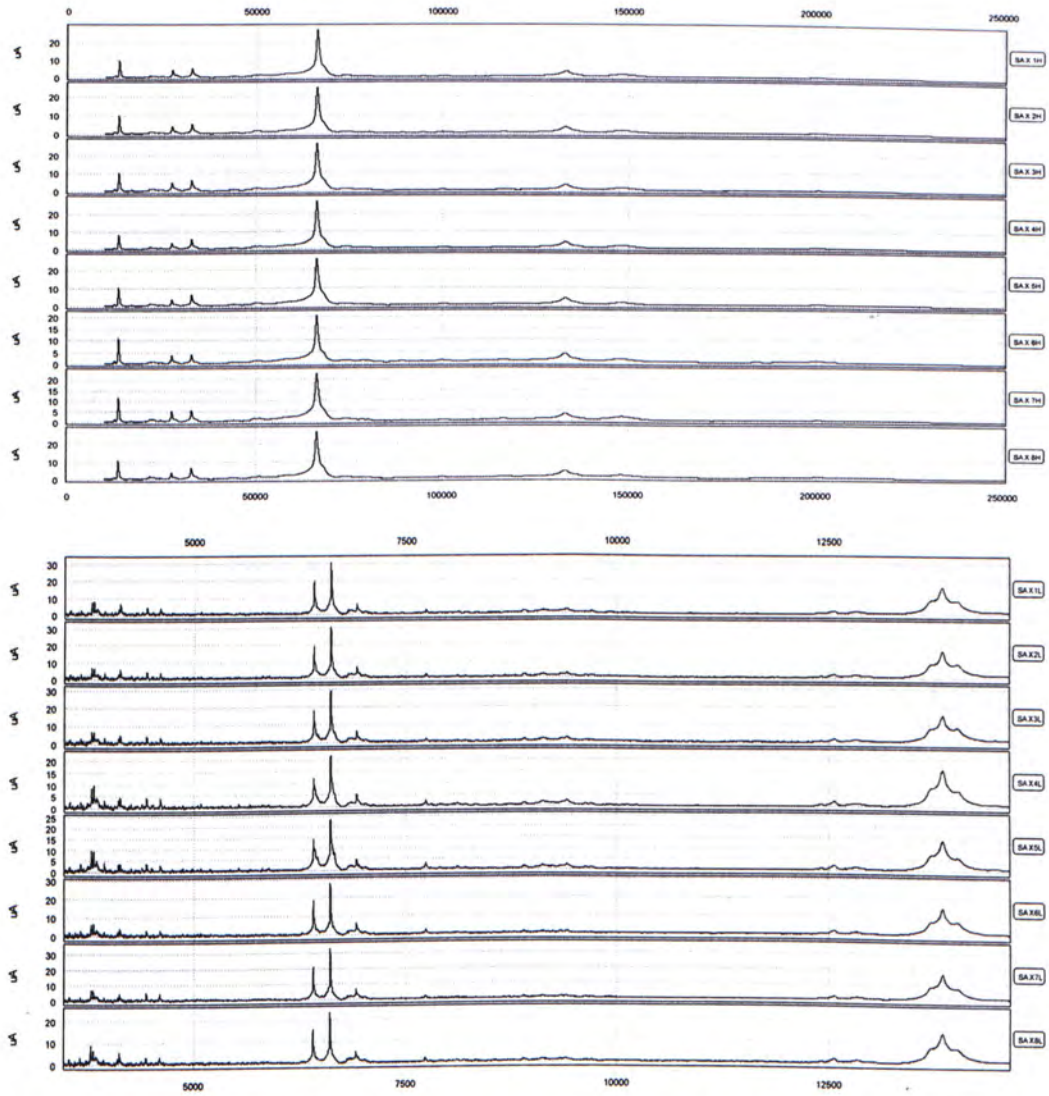


Figure 9. Representative spectra obtained with SAX beads. Top: High molecular range (*m/z* 20,000-250,000) Bottom: Low molecular range (*m/z* 1000-20,000)

1.3.4 Gel electrophoresis of the eluted proteins

50% ACN solution containing 0.2% TFA was successfully developed as the universal elution solution regardless of the types of functionalized magnetic beads used. Because it is a volatile solvent, it can be easily removed by speed vac. Then the dried proteins could be redissolved in any appropriate solvents for subsequently gel electrophoresis. The eluted proteins could be separated by non-reducing 2D PAGE for the ease of subsequent protein identification (Figure 10A).

1.3.5 Identification of the protein peaks

The protein identities of the protein spots in a 2D PAGE can be obtained by undertaking standard methods employing MALDI-TOF/TOF MS. The protein peaks in the mass spectrum could be easily matched with the protein spots with similar molecular weight to obtain their identities. Figure 10C is a typical example illustrating the proteins identified in the proteomic profile from a normal healthy subject, which were consistent to results previously reported by our team and by other teams using the SELDI ProteinChip technology for biomarker discovery. For example, the m/z 11685 peak were found to be serum amyloid A, which was consistent to what we found in a previous study in which SELDI ProteinChip technology was used. [121]

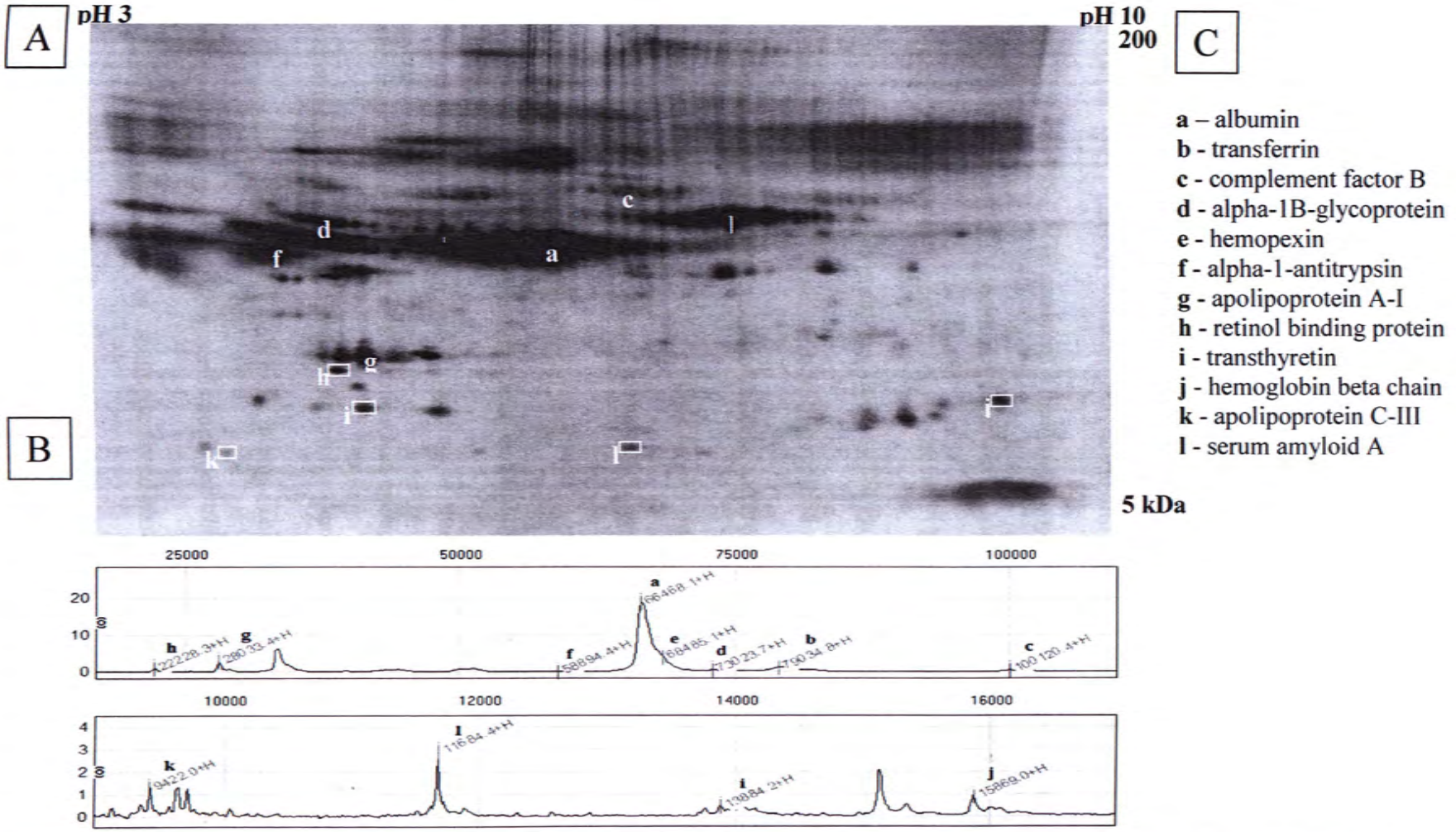


Figure 10. The protein identities of the MS peaks in the serum proteomic profile from a normal healthy subject. (A) Non-reducing 2D gel image of proteins eluted from C18 beads; (B) MS spectrum (i.e. proteomic profile) of the same preparation of eluted proteins; 12 proteins (C) were identified by MS approaches, and the corresponding protein spots and peaks were marked on (A) and (B), respectively.

1.4 Discussion

It is believed that low abundance and small proteins (<15kDa) of the serum proteome contain important diagnostic information and should be a rich source of undiscovered biomarkers. However, tradition techniques like 2D PAGE, only allows the separation and quantification of proteins with molecular weight in a range of 10 to 300kDa. [122] Lots of valuable diagnostic information may be lost if the sample is subjected to 2D PAGE analysis. This limitation has been overcome by the invention of the ProteinChip SELDI technology. However, low reproducibility and not applicable for protein identification hinders its development in biomarker discovery.

To overcome the problems faced by the ProteinChip SELDI technology, magnetic bead-based quantitative serum profiling method was developed in the present study. An automatic platform was used to process different types of chromatographic magnetic beads, such as hydrophobic C18 beads and strong cation exchange beads, to capture and release serum/plasma proteins with specific physicochemical properties. One microliter of the eluted proteins (60 uL) was directly spotted on a gold ProteinChip array in duplicate, overlaid with sinapinic acid, and finally subjected to the ProteinChip reader of SELDI-TOF MS system for acquiring the quantitative proteomic profile (i.e. mass spectrum), while the rest of the eluate was enough for subsequent protein identification experiments. The resulted proteomic profiles can be analyzed with existing informatic softwares for SELDI-TOF MS experiments. This method is cost effective, and can be easily adopted by those laboratories equipped with a SELDI-TOF MS system.

The performance of magnetic bead-based method was comparable to the ProteinChip SELDI technology. Different types of ProteinChip arrays could produce

different proteomic profiles. Similarly, the use of different types of magnetic beads produced different proteomic profiles. Different profiles from C18 and SAX beads illustrate that combined use of different magnetic bead-based profiling assays can obtain a more comprehensive proteomic profiles for biomarker discovery. C18 beads were functionalized with C18 alkyl groups which were hydrophobic in nature. It was used for reversed phase fractionation of complex protein/peptide mixtures. It was believed that the major proteins captured from C18 beads were mainly hydrophobic proteins. As most proteins are hydrophobic due to long carbon chain, C18 beads would than give a global profile for protein analysis. SAX beads were functionalized with strong anion exchanger which was used to adsorb anionic proteins. By using optimized pH (i.e. pH 8) for binding, proteins with their pI values below 8 would become negatively charged and captured on the magnetic beads. Washing steps were introduced to wash away any contaminants and non-specific binding proteins. The pH of the washing solution was prepared to have one pH unit above the binding solution (i.e. pH 9) to ensure only negatively charged proteins were adsorbed. This greatly enhances the specificity of the protein profile. This can explain why more peaks were found in C18 protein profile and some unique peaks were found in either protein profile. With the combination of C18 and SAX profiling assays, proteins with different properties were captured and more information can be obtained for protein analysis.

Instrument performance was important for acquiring good quality of mass spectra for further analysis. The detector sensitivity, resolution, mass drift and mass accuracy were the main factors affecting the quality of the mass spectrum. Failure in any one of these may cause wrong peak clustering or failure in detecting low abundance protein

peaks for data analysis. The performance of the PCS 4000 ProteinChip reader was evaluated using OQ kit (Bio-Rad Laboratories). All tests including detector sensitivity, mass drift, resolution and mass accuracy were passed, indicating the MALDI-TOF MS was in good condition for acquiring MS spectra for proteomic profiling. The detector sensitivity test showed that protein can be detected at 10fmol level with S/N greater than 5. It showed the MALDI-TOF MS was sensitive enough for detection of low abundance proteins which were believed to be potential biomarkers. A severe mass drift may cause wrong peak clustering for data analysis which would led to make erroneous conclusion on the results. Mass drift of the spectra was therefore important to be evaluated. The mass drift between spectra was found to be negligible (2.94 Da) and wrong peak clustering was not likely occurred. Moreover, if the resolution of the MALDI-TOF MS was not good enough, proteins with similar molecular weight would merge to form one single peak, reducing protein peaks detected in the whole profile. The resolution at 1 kDa was high even 3 other peaks were very close to the testing protein, arg-8-vasopressin. The result showed that PCS 4000 ProteinChip reader was capable of detecting peak proteins with similar masses with high resolution. Mass accuracy was important in identifying the interested proteins in the spectra. External and internal calibration was carried out to ensure the experimental mass did not deviate too much with the theoretical mass. The result showed that the experimental mass was close enough with the theoretical mass.

The reproducibility of peak intensities of magnetic bead-based profiling assays (both intra-assay and interassay CVs in range of 4 to 30%) was comparable to the values reported for the ProteinChip SELDI technology. [123][124] Alteroviz *et al.* reported an

intra-assay CV of 28% for automatic procedure and an intra-assay CV of 45% when done manually. [123] Using a robotic sample-processing station, Semmes *et al.* reported an intra-laboratory CV from 9 to 43%. [124] In the present study, the use of the automated magnetic particle processor for processing the magnetic beads and serum samples was one of the keys to achieve good reproducibility. Another automated machine used was the nano-liter non-contact dispenser for adding matrix chemical on the gold ProteinChip arrays, and the matrix drying step of the added matrix were well controlled in a humidity control chamber. It was found that amount of matrix added and the drying time could greatly affect the reproducibility. The use of automated machines helped to reduce assay-to-assay and person-to-person variations. This also explained why the intra-assay CVs and the interassay CVs of assays were very close.

In addition, reproducibility is greatly governed by binding step which depends on the quality of the chromatographic surface. Compared with ProteinChip arrays, the functionalized magnetic beads can be manufactured in large single batch and distributed to different laboratories while the ProteinChip arrays have to be made one by one. Therefore, the quality of magnetic beads can be more easily controlled.

Protein identification was done to demonstrate the feasibility of the magnetic bead-based approach to obtain enough eluted proteins for both MS analysis and protein identification in one capture step. In contrast, the ProteinChip SELDI technology required additional samples and additional preparative purification steps to obtain protein fractions that were similar to those captured on the ProteinChip arrays.

Another advantage of this profiling method is that magnetic bead-based assays are more sensitive than the ProteinChip-based SELDI assays. The binding capacity of

magnetic beads is much higher than that of ProteinChip arrays. Furthermore, the capability of a magnetic bead-based assay can be easily increased by using more magnetic beads. In this method, the magnetic beads were added in excess to avoid competition among the serum proteins for the binding sites, which happened in the case of ProteinChip SELDI technology. Study showed that the ProteinChip SELDI technology yielded fewer mass peaks than a profiling assay based on magnetic beads and MALDI-TOF MS using urine samples. [125]

1.5 Conclusion

A magnetic bead-based proteomic profiling method was successfully developed. It was comparable to ProteinChip SELDI technology, but allowed quantitative proteomic profiling and microscale purification of the profiled proteins in parallel. The resulted proteomic profiles can be analyzed with existing informatic softwares for SELDI-TOF MS experiments. It is a good alternative to ProteinChip SELDI technology for biomedical research.

Section 2 Development of a proteome-based fingerprinting model for detecting liver fibrosis in patients with chronic hepatitis B infection

2.1 Introduction

Liver fibrosis is the wound-healing scar due to liver injury such as chronic liver diseases. It can progress from liver fibrosis, cirrhosis and ultimately liver cancer. The prevalence of hepatitis B infection varies worldwide and the highest rates are seen in Asia. Each year, an estimated 500,000 people die of cirrhosis and liver cancer caused by chronic infection. CHB adult patients develop liver cancer at a rate of 5 % per decade which is approximately 100-fold higher than the rate among non-infected people. [126]

Liver fibrosis is a reversible and appropriate treatment may reverse the liver fibrosis and prevent further complications. Clinical guidelines published by the American Association for the Study of Liver Diseases in 2007 recommended that patients with HBV DNA > 20,000 IU/mL and persistently or intermittently elevated ALT with age greater than 40 should be evaluated by liver biopsy. [127] Liver biospsy is the gold standard to ascertain the degree of liver injury. The main role of liver biopsy is to assist in deciding the need for antiviral therapy. Although antiviral treatment suppresses the viral replications effectively, complete eradication of HBV infection is rarely achieved with current available therapies. [128] Therefore, treatment monitoring is important to evaluate the treatment efficacy and consider the endpoint of therapy. Feld *et al.* further pointed out that liver fibrosis was perhaps a stable marker for disease progression. However, liver biopsy is subject to sampling error and its invasive nature makes it impossible for repeated measurements.

Blood tests based on the biochemical and virological markers maybe more appropriate for treatment monitoring due to its non-invasiveness. Current blood tests such as FibroTest and hepascore have good accuracy for diagnosis of liver fibrosis and cirrhosis with an area under the ROC curve ~0.8. [129][130]

The formation of liver fibrosis causes disruption of liver architecture and loss of liver function. As proteins are mainly synthesized by liver, alternation of protein contents between disease and control groups may reflect the progression of liver disease. Blood is a good reservoir of proteins. It is believed that differential serum proteins can be found between disease and control groups by using comparative approach. Serum proteomic profiling is a useful tool in comparative proteomics and has been widely applied to biomarker discovery in different diseases such as breast cancer [131], liver cancer [132] and esophagus cancer. [133]

Serum proteomic profiling can be obtained using 2-D PAGE and gel–free mass spectrometry-based technologies, such as ProteinChip SELDI technology. By comparing the protein spot intensities in the 2D gel images between disease and control groups, differential spots were excised for protein identification. White *et al.* identified 7 differential proteins using 2-D PAGE/LC MSMS platform in a CHC-associated liver fibrosis study. [134] However, it is a time consuming procedure and subject to large gel-to-gel variation which limits the sample size of the study. Gel–free mass spectrometry-based technology was then served as an alternative for biomarker discovery.

Certain protein markers were found to be associated with HBV-related liver fibrosis and cirrhosis using SELDI technology. [117-119] Serum protein markers may reflect not only liver function and inflammation but also other physiologic conditions of

liver fibrosis. Diagnostic models were constructed based on the differential protein markers. Though these reported models achieved good correlation with liver fibrosis (ROC curve area >0.8), the unknown protein identities and complex diagnostic models made them less applicable for clinical use.

In addition, most studies used treatment naïve samples to construct diagnostic models, but seldom assess the accuracy of the diagnostic model using post-treatment samples. The accuracy of the model in detecting liver fibrosis after antiviral treatment remains unclear. Only treatment-independent diagnostic model can be used in treatment monitoring. For example, acute phase response proteins markers identified by SELDI-TOF technology may not be useful in treatment monitoring as they can be easily altered by antiviral treatment. [135] In addition, hematologic abnormalities may occur after antiviral treatment. [126] As some of the current non-invasive diagnostic models consist of biochemical and hematological parameters, the accuracy of them may be affected. Further examination on these models using serum/plasma samples that were collected after the antiviral treatment should be done to assess the possible use in treatment monitoring.

In this section, we aimed to identify the serum proteomic fingerprint, and develop a diagnostic model for the detection of liver fibrosis in patients with CHB infection using the magnetic bead-based proteomic fingerprinting approach developed in the first part of this M.Phil. study. The diagnostic model was validated in a second independent group of serum samples that were collected after antiviral treatment to investigate the its possible use in treatment monitoring.

2.2 Materials and methods

2.2.1 Patient materials

The current study included patients with CHB who were recruited or screened for other therapeutic trials as well as patients who were suspected of having active liver disease based on laboratory or radiologic investigations between 1998 to 2005 in the Prince of Wales Hospital, Hong Kong. All patients had compensated liver disease at the time of recruitment. Patients had received combined treatment of peginterferon and lamivudine, lamivudine or placebo treatment during the study period. Patients had received antiviral treatments for 1 year, and liver biopsy were collected 6 months after the termination of treatment. After obtaining informed consent, fasting blood samples were collected by venipuncture and liver function tests were carried out within 4 weeks prior to liver biopsy. All patients had been fasting for at least 6h before the blood sampling. Serum was stored at -80°C before proteomic profiling analysis. Clinical, biochemical and hematological data were recorded from each patient. Hematological and biochemical parameters (complete blood screens, coagulation tests, bilirubin, total protein, albumin, AST, ALT, alkaline phosphatase (ALP)) were measured by a Modular Analytic system (Roche Diagnostics). All patients gave written consent for use of these data for research purposes and use of these clinical samples for biomarker discovery was approved by the university ethics committee.

Liver biopsies were obtained with 16-gauge Temno needles (Bauer Medical). The specimens were fixed with formalin, embedded in paraffin, and stained with hematoxylin-eosin. Liver tissues were at least 1.5cm in length with a minimum of five

portal tracts for diagnosis. Histological staging of liver fibrosis was assessed using Ishak fibrosis score by a single pathologist blinded for the clinical and proteomic data.

A total of 214 randomly selected serum samples (104 treatment naïve, 110 post-treatment) were used in this study. 102 patients had both pre- and post-treatment samples; 2 patients had pre-treatment samples only and 8 patients had post-treatment samples only.

Pre-treatment serum samples were used for construction of diagnostic models using magnetic-based proteomic profiling technology and post-treatment samples were used for independent validation, to examine the independence of the model upon anti-viral therapy.

2.2.2 Serum proteomic profiling

Serum samples were analyzed by two types of magnetic beads, C18 and SAX, to capture hydrophobic and anion proteins respectively. An automated machine, Kingfisher 96, was used for protein extraction in 96-well format for single run. The experimental conditions were briefly described. For C18 condition, 2μL serum was diluted and inactivated with 198μL of binding buffer (0.9% sodium chloride (NaCl) containing 0.1% trifluoroacetic acid (TFA)). 0.5 mg of C18 magnetic beads in 10 μL (Dynabeads RPC18, Invitrogen) were pre-washed sequentially with 90μL of pure acetonitrile (ACN) and 100μL of binding buffer. 80μL of diluted serum samples were mixed with the washed C18 beads for 10 minutes. The C18 beads were then washed 3 times with 100μL of 0.2% TFA. Finally, the captured proteins were eluted with 60μL of universal elution solution (50% ACN containing 0.2% TFA). For SAX condition, 2μL serum was first

denatured by adding 8μL of UC buffer (9M Urea, 2% CHAPS, 2.5mM Tris-base, 10mM Tris-HCl). After incubation at room temperature for 30 minutes, the denatured serum samples were diluted with 190μL of pH 8 phosphate binding buffer (20 mM sodium phosphate, 30% ethanol, pH 8.0). 0.5 mg of SAX magnetic beads in 10 μL (Dynabeads SAX, Invitrogen) were mixed with 60μL of 1.5 M NaCl and 30 μL of ethanol for prewashing, and then washed with 100 μL of pH 8 phosphate binding buffer. 80μL of diluted serum samples were mixed with the washed SAX beads for 10 minutes. The SAX beads were then washed with pH 8 phosphate binding buffer, followed by 3 washes with washing buffer (2.3μM ethanolamine, 30% ethanol, pH 9.0). Finally, the anionic proteins were eluted with 60μL of universal elution solution.

2.2.3 Quantitative serum proteomic profiling by linear MALDI-TOF MS

1μL of eluate was spotted onto an 8-spot gold ProteinChip array (Ciphergen Biosystems) in duplicate. 0.5μL of sinapinic acid matrix (13mg/mL sinapinic acid dissolving in 50% ACN/ 0.5% TFA) (Sigma-aldrich) was added by using of nano-liter non-contact dispenser (BioDot AD3050) and air-dried in a chamber with humidity control at 80%. Another 0.5μL of sinapinic acid matrix was added and air-dried. The gold ProteinChip array was subjected to the PCS4000 ProteinChip reader (Bio-Rad Laboratories) to determine the masses and intensities of all the peaks over the m/z range from 1,000 to 250,000 m/z. Two acquisition protocols were used for low and high mass ranges. Intensities of peaks between 1,000 and 20,000 m/z were obtained at a laser power of 6,000nJ and the focus mass was 8,000 m/z; intensities of peaks between 10,000 to 250,000 m/z were obtained at a laser power of 10,000nJ and the focus mass was 80,000 m/z. The mass spectra were externally calibrated with a mixture of

peptide/protein standards (angiotensin, 1,296.51 m/z; ACTH (clip 1-17), 2,093.46 m/z; ACTH (clip 18-39), 2465.72 m/z; double charged horse apomyoglobin 8475.8 m/z; E. coli Thioredoxin, 11673.5 m/z, horse apomyoglobin, 16951.6 m/z; bovine serum albumin, 66430 m/z; bovine serum albumin dimer, 132861 m/z) (Applied Biosystems). The common peaks among the mass spectra were identified and quantified using the Biomarker Wizard software (Bio-Rad Laboratories). The peak intensities were normalized with the total ion current, and subsequently with the total peak intensities. The duplicate normalized peak intensity measurements were averaged for further statistical analysis.

2.2.4 Statistical analysis

Serum samples were divided into two groups. Treatment naïve samples were used as training group for both biomarker discovery and diagnostic model construction and post-treatment samples were used as validation group. The correlations among Ishak score, protein peak intensities obtained from magnetic bead protein profiles, age (calculated to the date of liver biopsy), circulating HBV DNA, haemoglobin level (HB), white blood cell count (WBC), platelet count (PLT), INR, total protein concentration, albumin concentration, bilirubin concentration, ALP, AST and ALT were analyzed by Spearman's rank correlation test. Mann-Whitney U and bivariate Spearman's rank correlation tests were conducted with SPSS16.0 (SPSS, Inc, Chicago, IL).

For the discovery dataset, 4 criteria should be fulfilled for a differential proteomic features. Firstly, the normalized protein peaks should be significantly higher or lower in patients with significant fibrosis than those patients without significant

fibrosis. 2-independent sample comparison tests were performed by using Mann-Whitney U test. 2 groups were classified based on Ishak scores, Minimal fibrosis (Ishak score = 0, 1 and 2) and typical fibrosis (Ishak score = 4, 5 and 6) groups. Secondly, the normalized protein peaks should be associated with the progression of liver fibrosis. Spearman's rank correlation test was used to examine the correlations between proteomic features/serological parameters and the degrees of liver fibrosis. Thirdly, the proteomic features should be identified to be significantly associated with Ishak score at a false discovery rate (FDR) less than 5%. To assess the FDR, significance analysis of microarray (SAM) analysis was used (Stanford University, 161, 162). In the SAM analysis, "Quantitative" was selected in response type list, "ranks" was chosen in regression method, and 5,000 permutations were performed. Fourthly, the differential proteomic features should be significantly correlated with at least 1 of the serological/biochemical/clinical parameters to illustrate its biological relevance. Proteomic features showed statistically significant in all 4 tests were considered as potential biomarkers.

2.2.5 Development of diagnostic model

The potential proteomic, biochemical and serological markers were $\log_2$ transformed before constructing diagnostic model. The $\log_2$ values were subjected to multiple linear regression (forward stepwise) to formulate predictive models for detecting significant liver fibrosis. During model training, the potential proteomic, biochemical and serological markers were used as independent variables and Ishak score as dependent output variable. For the biomarker discovery group, the performance of the

predictive model was evaluated by 10 fold cross-validation. Briefly, all treatment naïve cases were divided into 10 groups, the model was trained on 9 groups and the trained model was then used to test the case that had been left out. The process repeated until every case in the dataset had been used once as an unseen test case. The results were averaged across all cases to evaluate the prediction performance. Receiver Operating Characteristic (ROC) curves was used to assess the sensitivity and specificity of the diagnostic models in detecting significant fibrosis (Ishak score >2) and cirrhosis (Ishark score >4). The positive (PPV) and negative predictive values (NPV) were then calculated to investigate the diagnostic value of the predictive model.

2.2.6 Independent validation of diagnostic model

Independent validation was carried out using post-treatment data to evaluate the accuracy of the predictive model and examine whether the model was affected by antiviral therapy. The diagnostic value of the model was assessed by ROC curve analysis, and sensitivity, specificity, PPV and NPV in detecting significant fibrosis were calculated according to standard equations.

2.2.7 Comparison of the constructed diagnostic model with other non-invasive models

Two non-invasive models constructed for detecting CHB associated liver fibrosis, Aspartate aminotransferase to platelet ratio index (APRI) [Hepatology 2003, 38, 518-26] and age-platelet index (API) [J Viral Hepatol 1997;4:199-208], were investigated, and their diagnostic performances were compared with our predictive model. APRI and API were calculated according to the following equation and procedure.

Equation for calculating APRI:

$$\text{APRI} = \frac{\text{AST level/ULN}}{\text{Platelet counts } (10^9/\text{L})} \times 100$$

Procedure for calculating AP index:

1) Age (years): <30=0; 30-39 =1; 40-49 =2; 50-59 =3; 60-69 = 4; ≥70 =5
2) Platelet count (10^9/L): ≥225 =0; 200-224 =1; 175-199 =2; 150-174 =3; 125-149 =4; <125 =5
3) AP index is the sum of the age-platelet count index (possible value 0-10)

2.3 Results

2.3.1 Patient characteristics

For pre-treatment group, the mean (SD) of Ishak fibrosis score was 1.81 (1.55). 10 patients had no fibrosis (Ishak score =0); 52 had minimal fibrosis (Ishak score =1); 23 had mild fibrosis (Ishak score =2); 8 had moderate fibrosis (Ishak score =3 or 4); 6 had severe fibrosis (Ishak score =5), and 5 had probable/ definite cirrhosis (Ishak score =6). The prevalence of significant fibrosis in discovery group was 18.27%.

For post-treatment group, the mean (SD) of Ishak fibrosis score was 1.84 (1.76). 14 patients had no fibrosis (Ishak score =0); 61 had minimal fibrosis (Ishak score =1); 9 had mild fibrosis (Ishak score =2); 11 had moderate fibrosis (Ishak score =3 or 4); 7 had severe fibrosis (Ishak score =5), and 8 had probable/ definite cirrhosis (Ishak score =6). The prevalence of significant fibrosis in validation group was 23.64%. Minimal fibrosis was defined as Ishak score lower than 3 while significant fibrosis was defined as Ishak score greater than 3, indicating the presence of bridging fibrosis and/or cirrhosis. Table 3 showed the distribution of fibrosis stages in both pre- and post-treatment groups. Results showed that the prevalence of significant fibrosis was similar in both training and validation groups.

The demographic data of pre- and post-treatment groups were compared to find any differences in biochemical/ hematological data after antiviral treatment using two-tailed Student-t test. To avoid multiple comparisons, p-value was adjusted using Bonferroni method and $p < 0.003$ was considered as statistically significant. There were no significant difference in age, Ishak score, HB, WBC, PLT, INR, total protein and bilirubin concentration (p-values > 0.003). On the other hand, statistically significant

differences were found in PT, ALB, ALP, AST, ALT and HBV DNA. All these parameters except ALB were found to have lower level in post-treatment groups.

Further evaluation was done on pre- and post-treatment dataset. The serum samples from patients with and without significant liver fibrosis were separated and investigated to evaluate the effects of the treatment on the serological/biochemical parameters. For patients without significant fibrosis, PT, INR, ALB, ALP, AST, ALT and HBV DNA were found to be statistically significant between pre- and post-treatment groups (p-values < 0.003). For patients with significant fibrosis, ALB, AST and ALT were significantly different between pre- and post-treatment groups (p-values <0.003). All the comparisons in different groups were summarized in Tables 4-6.

Table 3 Distribution table of patients with different Ishak scores in pre- and post-treatment samples. N.A.= Not applicable

Ishak score	Pre-treatment samples (n = 104)	Percentage (%)	Post-treatment samples (n = 110)	Percentage (%)
0	10	9.62	14	12.73
1	52	50.00	61	55.45
2	23	22.12	9	8.18
3	2	1.92	4	3.64
4	6	5.77	7	6.36
5	6	5.77	7	6.36
6	5	4.80	8	7.27
Mean	1.81	N.A.	1.84	N.A.
SD	1.55	N.A.	1.76	N.A.
Prevalence of significant fibrosis	19	18.27	26	23.64

Table 4 A summary of demographic data of all patients in pre- and post-treatment samples, mean (SD) (*p <0.003) N.A.= Not applicable

	Pre-treatment	Post-treatment	p-value
	All	All	
Sex (Male/Female)	3.00	3.07	N.A.
Age (yr)	35.8 (10.0)	37.90 (10.5)	0.136
Ishak score	1.81 (1.55)	1.84 (1.76)	0.900
HB (g/L)	14.83 (1.49)	14.86 (1.55)	0.867
WBC (10^9/L)	6.45 (5.53)	6.18 (1.65)	0.633
PLT (10^9/L)	183.51 (54.28)	203.79 (60.27)	0.011
PT (sec)	11.07 (0.83)	10.45 (0.64)	<0.0005*
INR	1.07 (0.08)	1.04 (0.07)	0.004
Total protein (g/L)	76.83 (4.53)	77.75 (5.03)	0.163
ALB (g/L)	39.93 (3.20)	43.95 (3.21)	<0.0005*
Bilirubin (mg/dl)	10.42 (4.33)	11.35 (5.37)	0.165
ALP (IU/L)	87.90 (24.84)	74.00 (22.79)	<0.0005*
AST (IU/L)	80.82 (51.27)	31.42 (21.17)	<0.0005*
ALT (IU/L)	151.73 (103.48)	51.94 (59.04)	<0.0005*
HBVDNA	5.71E+08 (1.01E+09)	4.84E+06 (1.91E+07)	<0.0005*

Table 5 A summary of demographic data of patients with minimal fibrosis in pre- and post-treatment samples, mean (SD). (*p <0.003) N.A.= Not applicable

	Pre-treatment	Post-treatment	p-value
	Minimal fibrosis	Minimal fibrosis	
Sex (Male/Female)	2.70	2.74	N.A.
Age (yr)	34.69 (9.99)	35.92 (9.84)	0.424
Ishak score	1.15 (0.61)	0.94 (0.52)	0.016
HB (g/L)	14.76 (1.48)	14.66 (1.63)	0.694
WBC (10^9/L)	6.52 (6.05)	6.40 (1.68)	0.863
PLT (10^9/L)	191.96 (52.78)	215.38 (58.09)	0.007
PT (sec)	10.93 (0.79)	10.28 (0.51)	<0.0001*
INR	1.07 (0.08)	1.03 (0.05)	<0.0001*
Total protein (g/L)	76.79 (4.82)	77.94 (4.62)	0.114
ALB (g/L)	40.22 (3.31)	44.08 (3.16)	<0.0001*
Bilirubin (mg/dl)	9.98 (4.38)	10.69 (5.55)	0.354
ALP (IU/L)	86.07 (24.19)	71.08 (21.08)	<0.0001*
AST (IU/L)	75.08 (48.44)	30.54 (20.93)	<0.0001*
ALT (IU/L)	149.93 (105.09)	48.70 (57.92)	<0.0001*
HBVDNA	6.68E+08 (1.08E+09)	3.68E+06 (1.71E+07)	<0.0001*

Table 6 A summary of demographic data of patients with significant fibrosis in pre- and post-treatment samples, mean (SD). (*p <0.003) N.A.= Not applicable

	Pre-treatment	Post-treatment	p-value
	Significant fibrosis	Significant fibrosis	
Sex (Male/Female)	5.33	5.00	N.A.
Age (yr)	40.74 (8.48)	44.31 (10.36)	0.225
Ishak score	4.74 (0.99)	4.73 (1.08)	0.985
HB (g/L)	15.14 (1.55)	15.50 (1.04)	0.347
WBC (10^9/L)	6.09 (1.99)	5.46 (1.36)	0.212
PLT (10^9/L)	145.68 (44.88)	166.35 (52.22)	0.172
PT (sec)	11.68 (0.77)	11.01 (0.70)	0.004
INR	1.12 (0.08)	1.11 (0.07)	0.560
Total protein (g/L)	77.00 (3.04)	77.11 (6.22)	0.941
ALB (g/L)	38.63 (2.31)	43.54 (3.42)	<0.0001*
Bilirubin (mg/dl)	12.42 (3.52)	13.50 (4.12)	0.362
ALP (IU/L)	96.11 (26.72)	83.42 (25.84)	0.116
AST (IU/L)	110.00 (53.24)	36.31 (22.72)	0.001*
ALT (IU/L)	159.79 (98.24)	62.38 (62.52)	0.001*
HBVDNA	1.49E+08 (3.14E+08)	1.27E+07 (2.94E+07)	0.144

2.3.2 Correlation between biochemical/serological markers and the degrees of liver fibrosis

4 significant serological markers (PLT, PT, bilirubin and HBV DNA level) were correlated with the degrees of liver fibrosis. PLT and HBV DNA were negatively correlated with the degrees of liver fibrosis while PT and bilirubin were positively correlated. List of differential serological/biochemical markers were summarized in Tables 7 and 8.

2.3.3 Serum proteomic profiling by linear MALDI-TOF MS

Protein profiles were obtained in two mass ranges, low (1,000 and 20,000 m/z) and high mass range (10,000 to 250,000 m/z). In C18 profiles, 129 and 74 peaks were found in low and high mass range respectively. In SAX profiles, 69 and 81 peaks were found in low and high mass range respectively.

2.3.4 Correlation of proteomic features with Ishak score

5 differential proteomic features were associated with Ishak scores and significantly different in minimal and significant fibrosis groups. At a FDR < 5%, 4 of the differential proteomic features (9,165, 9,452, 10,045 and 12,243 m/z) were found from C18 profiles and one (17,286 m/z) was from SAX profiles. Among 5 potential proteomic markers, 3 (9,165, 9,452 and 12,243 m/z) were down-regulated and 2 (m/z 10,045 and 17,286) were up-regulated. List of differential proteomic features were summarized in Table 9. Representative spectra of individual peaks were shown in Figures 11-14.

Table 7 List of significant differential proteomic features (p <0.05, FDR<5%), mean (SD)

	Correlation coefficient	p-value		Minimal fibrosis	Significant fibrosis
		Spearman correlation	Mann-Whitney U		
PLT	-0.399	<0.0005	<0.0005	191.96 (52.78)	145.68 (44.88)
PT	0.349	<0.0005	0.001	10.93 (0.79)	11.68 (0.77)
Bilirubin	0.255	0.009	0.005	9.98 (4.38)	12.42 (3.52)
HBV DNA	-0.369	0.002	0.001	6.67E+08 (1.08E+09)	1.49E+08 (3.14E+09)

	Serological markers (r, p-value)
PLT	WBC (0.255, 0.009), PT (-0.288, 0.003), INR (-0.311, 0.003), Bilirubin (-0.212, 0.03),
PT	PLT (-0.288, 0.003), INR (0.882, <0.0005), Bilirubin (0.325, 0.001), ALP (0.320,0.001), AST (0.339, 0.002), HBVDNA (-0.257, 0.033)
Bilirubin	HB (0.236, 0.016), PLT (-0.212, 0.03), PT (0.325, 0.001), INR (0.327, 0.002), ALT (0.200, 0.042), AST (0.297, 0.008)
HBV DNA	PT (-0.257, 0.033)

Table 8 A summary of correlation of biochemical / serological markers ($p < 0.05$)

Table 9 List of significant differential proteomic features ($p < 0.05$, FDR <5%), mean (SD)

	Correlation coefficient	p-value		Minimal fibrosis	Significant fibrosis
		Spearman correlation	Mann-Whitney U		
m/z 9165	-0.279	0.004	0.045	0.0116 (0.0047)	0.0096(0.0036)
m/z 9452	-0.324	0.001	0.025	0.0141 (0.0069)	0.0106 (0.0049)
m/z 10045	0.248	0.011	0.041	0.0235 (0.0105)	0.0282 (0.0123)
m/z 12243	-0.317	0.001	0.02	0.0044 (0.0027)	0.0032 (0.0017)
m/z 17286	0.271	0.006	0.019	0.3290 (0.1337)	0.3762 (0.1322)

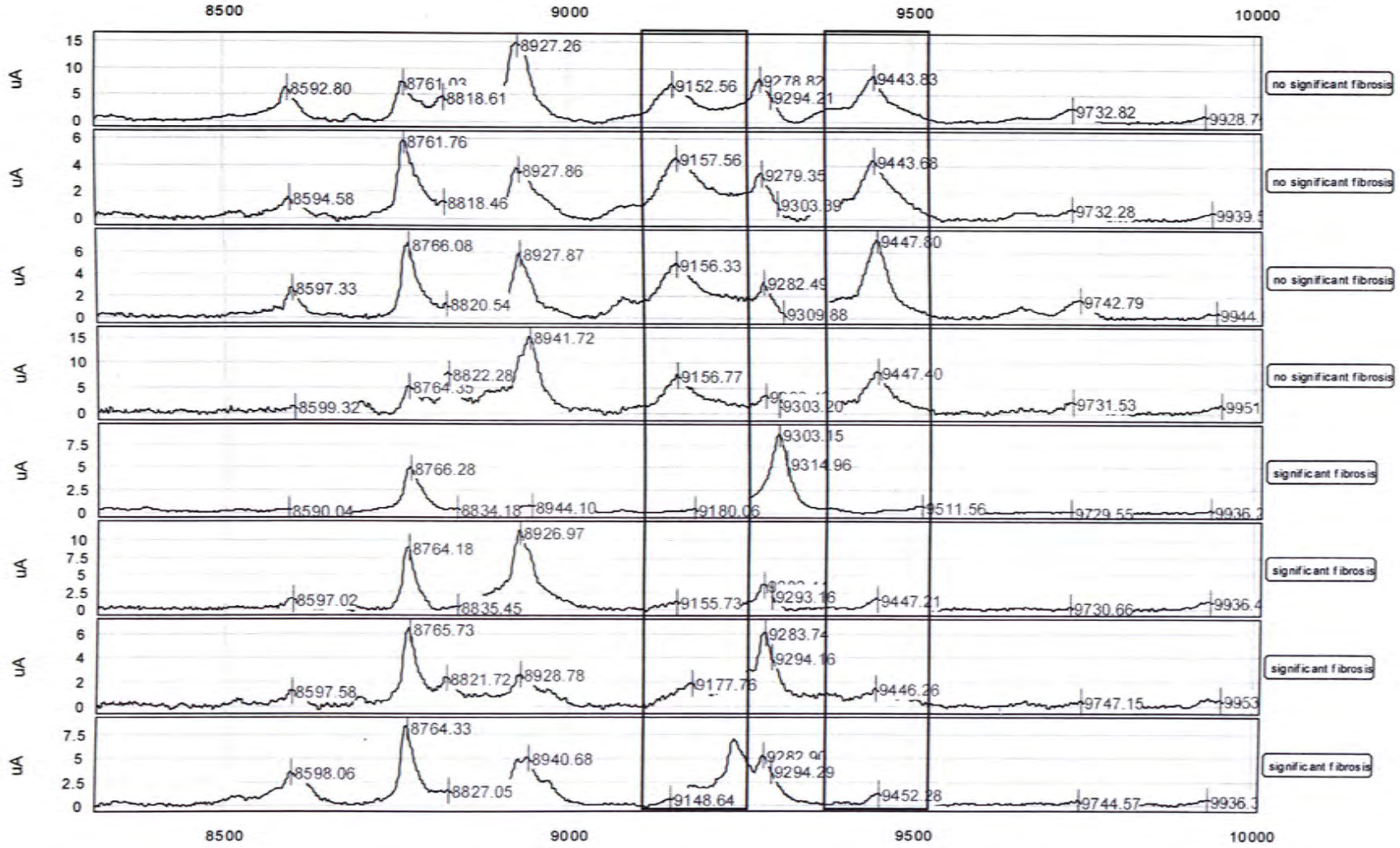


Figure 11 Representative spectra of 9.2 and 9.5 kDa differential proteomic features. Top 4: Minimal fibrosis; Bottom 4: Significant fibrosis.

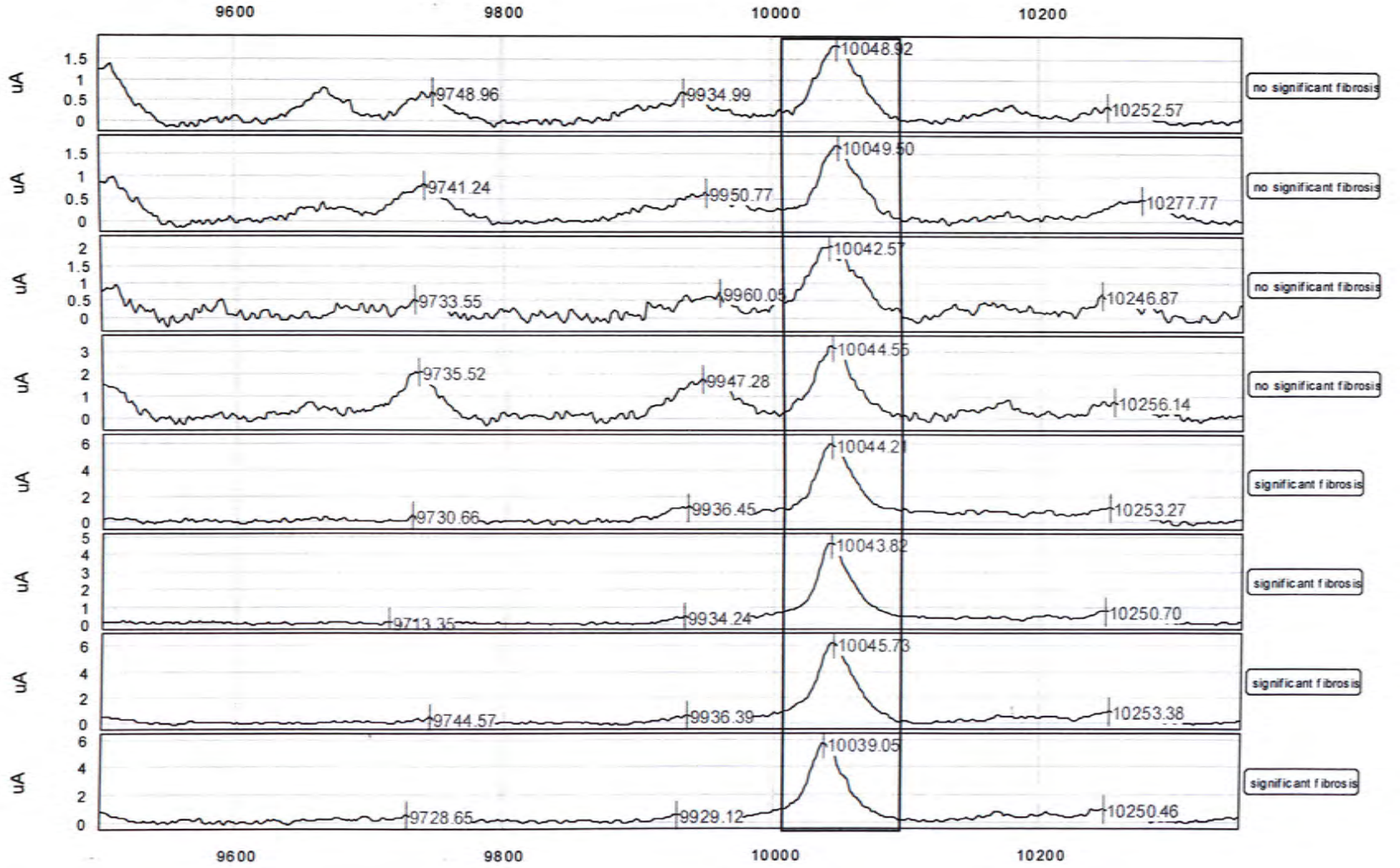


Figure 12 Representative spectra of 10 kDa differential proteomic features. Top 4: Minimal fibrosis; Bottom 4: Significant fibrosis

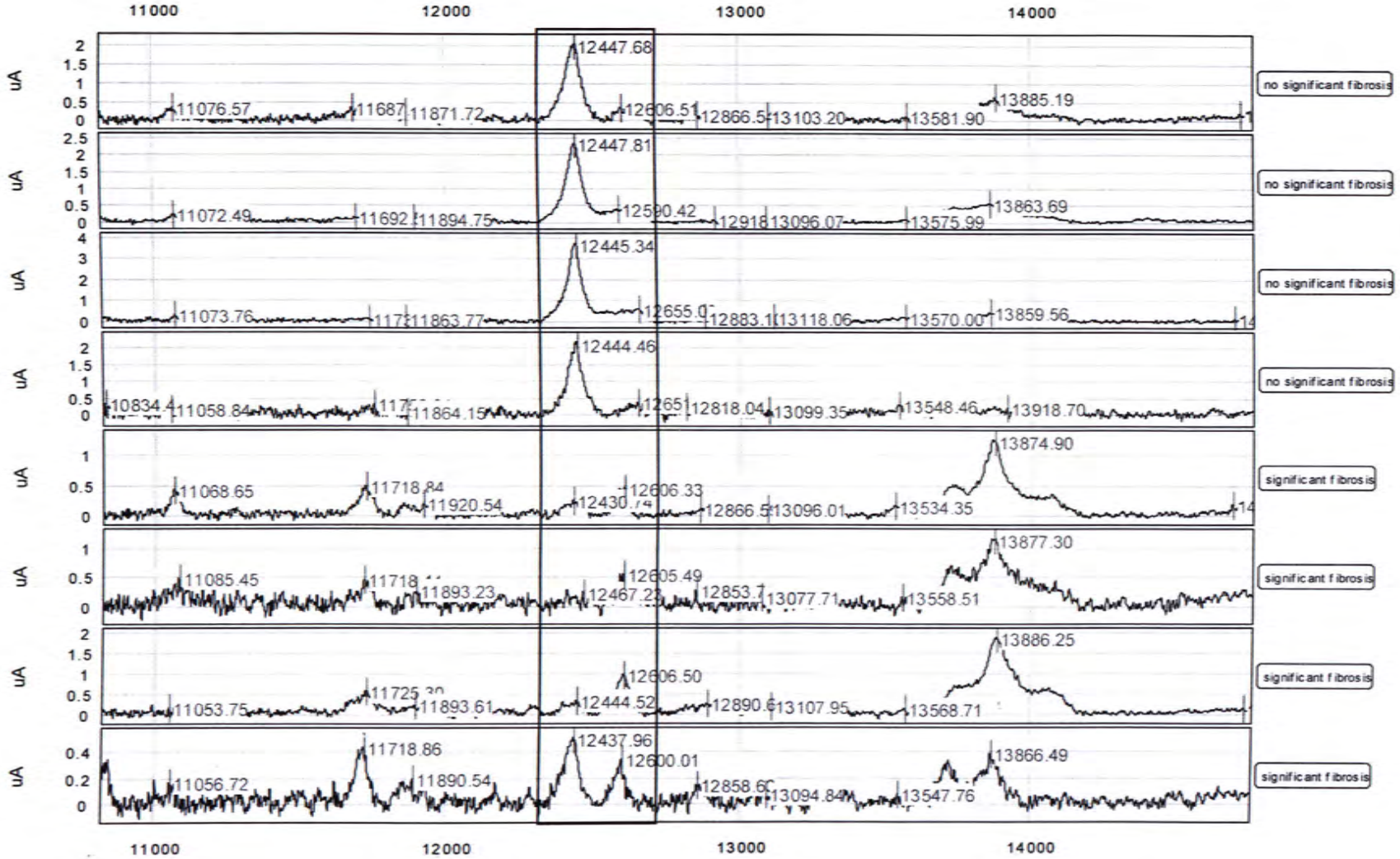


Figure 13 Representative spectra of 12.4 kDa differential proteomic feature. Top 4: Minimal fibrosis; Bottom 4: Significant fibrosis

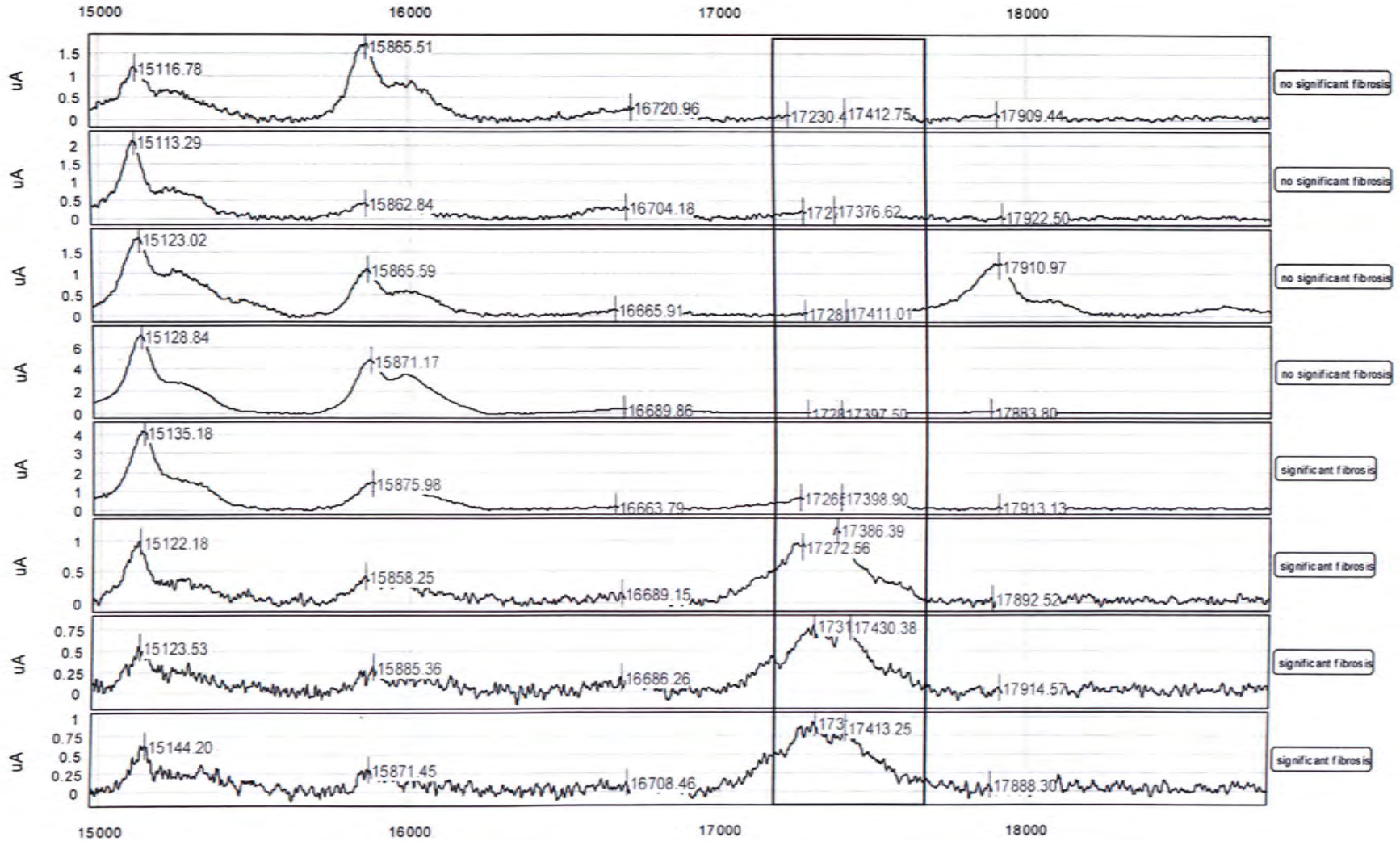


Figure 14 Representative spectra of 17.3 kDa differential proteomic feature. Top 4: Minimal fibrosis; Bottom 4: Significant fibrosis

2.3.5 Correlation of significant proteomic features with serological markers

4 significant proteomic features were correlated with serological markers. 9,165 m/z was positively correlated with WBC ($r = 0.200$, $p = 0.042$). Two proteomic features, 10,045 and 12,243 m/z were associated with clotting factors. Proteomic feature 10,045 m/z, which was up-regulated with the degree of liver fibrosis, was negatively correlated with PLT ($r = 0.228$, $p = 0.02$) and positively correlated with PT ($r = 0.240$, $p = 0.015$). 12,243 m/z, a down-regulated marker, was negatively correlated with PT ($r = -0.208$, $p = 0.035$) and INR ($r = -0.223$, $p = 0.034$) and positively correlated with PLT ($r = 0.228$, $p = 0.02$). The correlation between proteomic features and serological markers were summarized in Table 10. These 4 proteomic features were then regarded as potential diagnostic markers.

Table 10 A summary of correlation of significant proteomic fearures and biochemical / serological markers ($p < 0.05$)

Proteomic feature	Serological markers (r, p-value)
m/z 9165	WBC (0.200, 0.042)
m/z 10045	PLT (-0.227, 0.021), PT (0.240, 0.015),
m/z 12243	PLT (0.228, 0.02), PT (-0.208, 0.035), INR (-0.223, 0.034),
m/z 17286	PLT (-0.183, 0.013)

2.3.6 Construction of diagnostic model in detecting liver fibrosis and cirrhosis

A total of 8 $\log_2$ of normalized potential protein markers and serological markers were selected to construct linear regression model (forward stepwise). Two protein markers (9,165 and 12,443 m/z) and PT were selected as the independent variables in the diagnostic model. The linear regression model was found to be Proteomic fibrosis index = -0.739*$\log_2$ (peak intensity of m/z 9165) - 0.398* $\log_2$ (peak intensity of m/z 12443) + 3.689* $\log_2$ (PT) -19.09. For predicting liver fibrosis (Ishak score >2), the diagnostic model was useful in identifying cases with significant liver fibrosis in the CHB patients with AUROC = 0.758 (95% CI 0.640-0.875, p = 0.0005). At a high cut-off of 2.4693, the specificity and sensitivity were 88% and 42% respectively. At a low cut-off of 1.5294, the specificity and sensitivity were 48% and 90% respectively. For predicting liver cirrhosis (Ishak score > 4), the diagnostic model was useful in identifying cases with significant liver cirrhosis in the CHB patients with AUROC = 0.851 (95% CI 0.743-0.958, p < 0.0005). At a high cut-off of 2.4693, the specificity and sensitivity were 88% and 64% respectively. At a low cut-off of 1.9762, the specificity and sensitivity were 66% and 91% respectively.

2.3.7 Cross-validation of the diagnostic model using pre-treatment samples in detecting liver fibrosis and cirrhosis

After 10 fold cross-validation, the diagnostic model showed similar performance. Our predictive index was consistent with Ishak score with a good correlation (r = 0.452, p < 0.0005) and a positive trend was shown in the boxplot. [Figure 15]

For predicting liver fibrosis, the diagnostic model was useful in identifying cases with significant liver fibrosis in the CHB patients with AUROC = 0.726 (95% CI 0.605-0.846, $p < 0.005$), indicating overfitting did not significantly exist in the predictive model. With the use of high (3.0844) and low cutoffs (1.3068), the PPV and NPV were found to be 80% and 96% respectively.

For predicting liver cirrhosis, the model was significantly useful with AUROC = 0.825 (95% CI 0.709-0.941, $p<0.0005$). With the use of high (3.0844) and low cutoffs (1.8981), the PPV and NPV of detection of liver fibrosis were found to be 80% and 98% respectively. Figure 15 showed a summary of ROC curves for both discovery and cross-validation group in detecting significant liver fibrosis and cirrhosis.

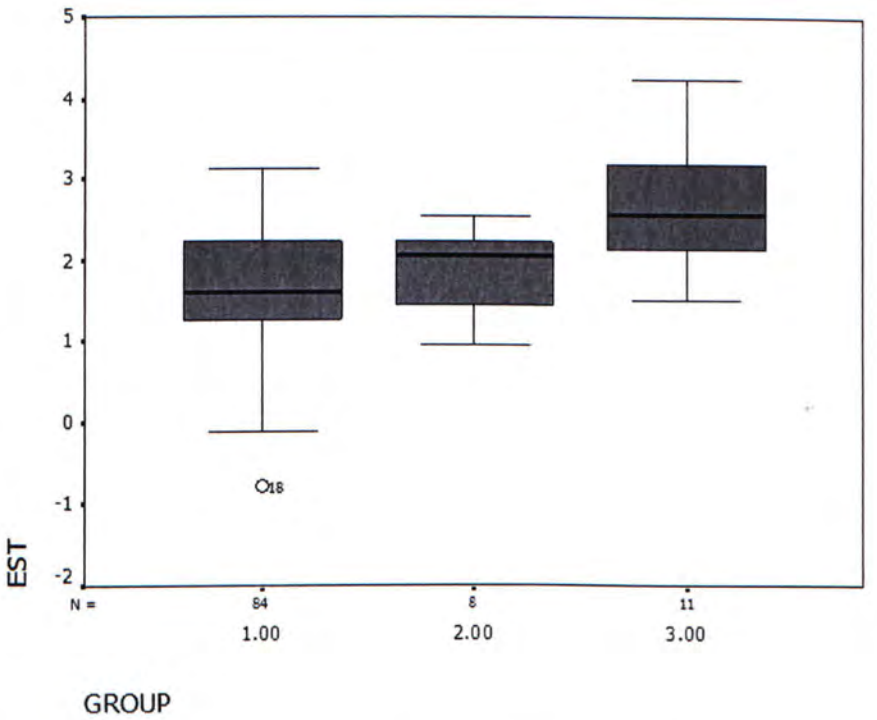


Figure 15 Boxplot of the predictive index in pre-treatment group (Group 1: minimal fibrosis (Ishak score = 0, 1, 2); Group 2: moderate fibrosis (Ishak score = 3, 4); Group3: cirrhosis (Ishak score = 5,6))

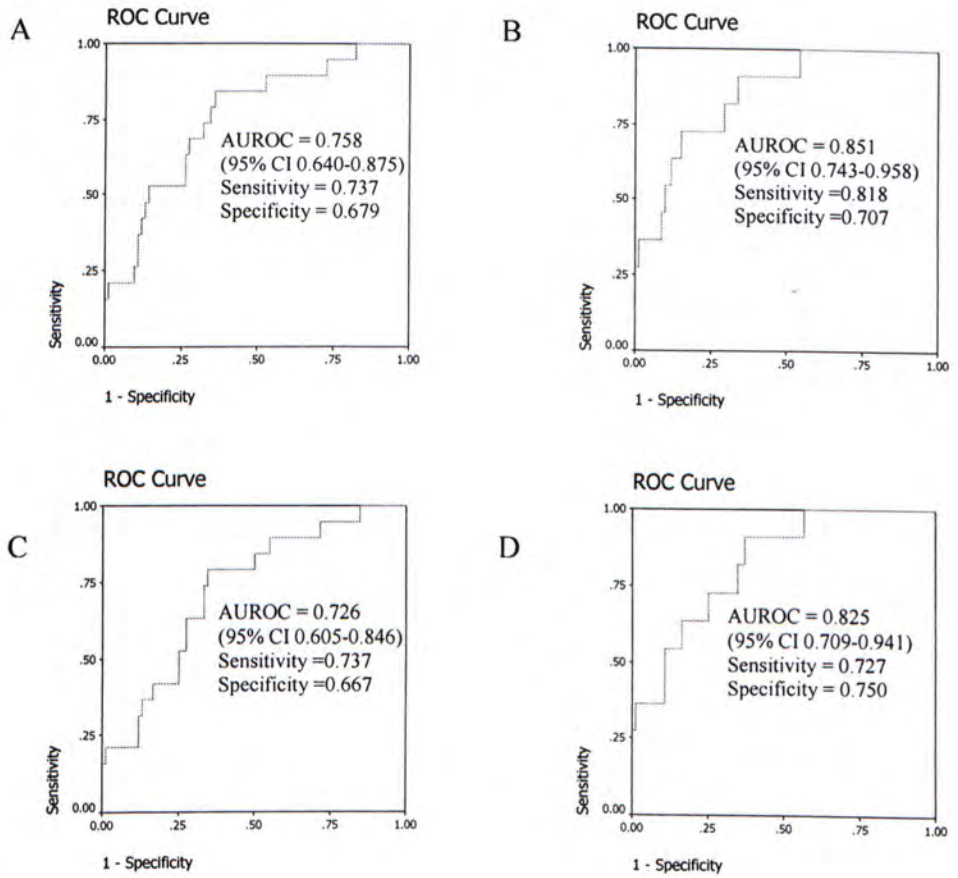


Figure 16 ROC curves of the diagnostic models in detecting significant fibrosis and cirrhosis in discovery group (pre-treatment). A) Liver fibrosis B) Liver cirrhosis C) 10 fold cross-validation for liver fibrosis D) 10 fold cross-validation for liver cirrhosis

2.3.8 Independent validation of the diagnostic model using post-treatment samples in detecting liver fibrosis and cirrhosis

Post-treatment samples protein data was used to evaluate the independence of the diagnostic model to the anti-viral therapy. Our predictive index was consistent with Ishak score with a good correlation ($r = 0.211$, $p < 0.05$) and a positive trend was shown in the boxplot. [Figure 17]

For liver fibrosis, the diagnostic model was useful in identifying cases with significant liver fibrosis in the CHB patients with AUROC = 0.750 (95% CI 0.640-0.860, $p < 0.0005$), similar to the performance in the pre-treatment discovery group. For predicting liver cirrhosis, the diagnostic model was useful in identifying cases with significant liver cirrhosis in the CHB patients with AUROC = 0.783 (95% CI 0.672-0.894, $p < 0.0005$). Figure 18 showed the ROC curves of the diagnostic model in detecting liver fibrosis and cirrhosis. Same cutoffs were obtained in predicting fibrosis and cirrhosis with the same accuracy. With the use of high (2.7303) and low cutoffs (0.3427), the PPV and NPV for detection of liver fibrosis were found to be 75% and 92% respectively.

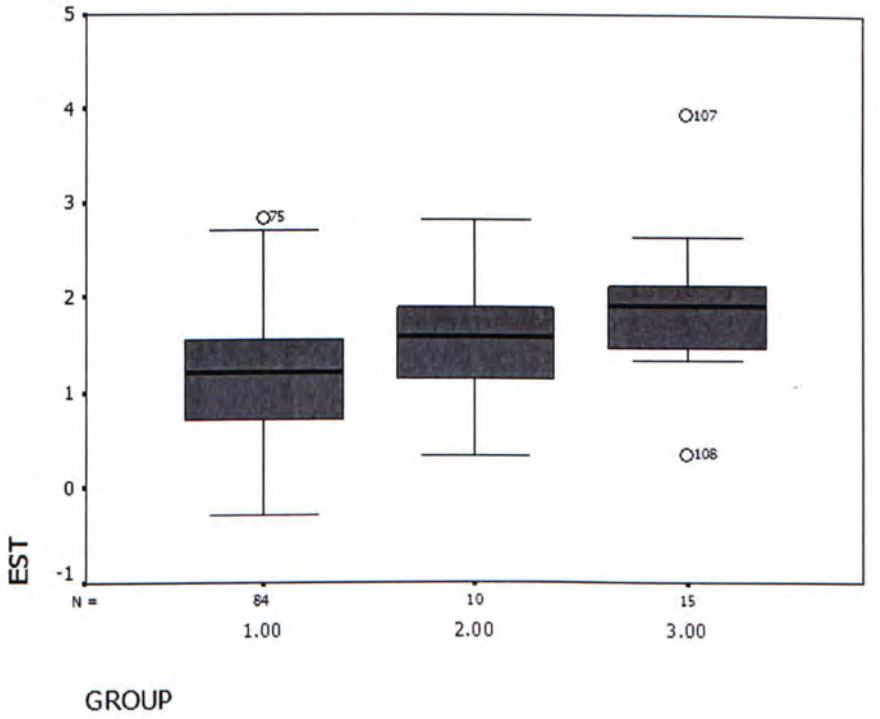


Figure 17 Boxplot of the predictive index in post-treatment group (Group 1: minimal fibrosis (Ishak score = 0, 1, 2); Group 2: moderate fibrosis (Ishak score = 3, 4); Group3: cirrhosis (Ishak score = 5,6))

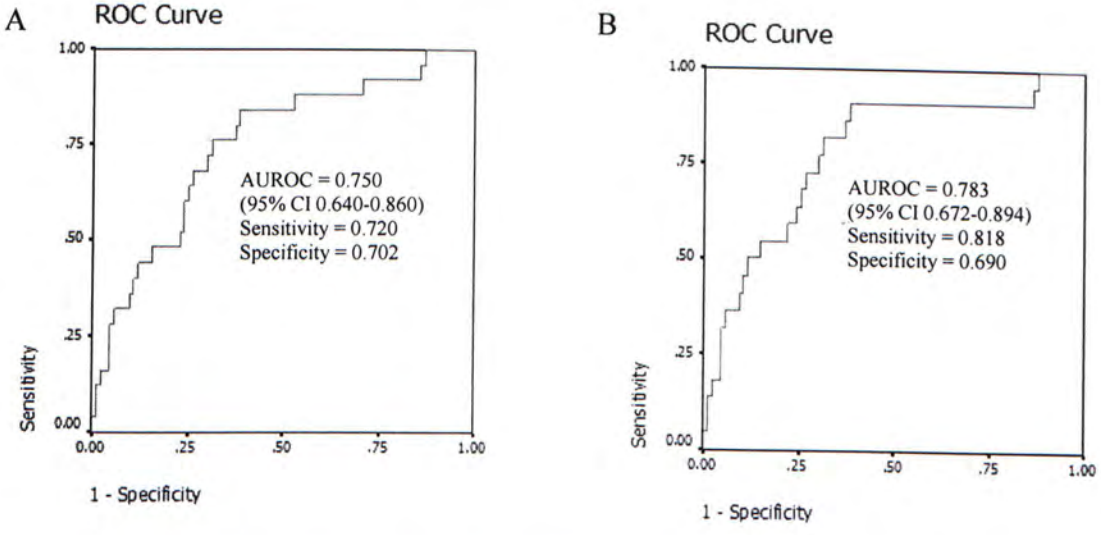


Figure 18 ROC curves of the diagnostic models in detecting significant fibrosis and cirrhosis in validation group (post-treatment). A) Liver fibrosis B) Liver cirrhosis

2.3.9 Comparison against other non-invasive models in detecting liver fibrosis and cirrhosis

The diagnostic performances of two non-invasive models, APRI and API, and our diagnostic model were compared. Our predictive model showed comparable results with the APRI and API models. For pre-treatment samples, both APRI and API models were useful in identifying cases with significant liver fibrosis in the CHB patients with AUROC = 0.791 and 0.781, respectively (p values=0.001). However, for post-treatment samples, the APRI model was not significantly useful in predicting post-treatment samples while API was still significantly useful (p=0.045). However, the AUROC of the API was dropped from 0.781 to 0.676. Our predictive model showed better and more stable performance in detecting post-treatment samples with an AUROC of 0.750. Proteomic model (m/z 9,165 and 12443) was constructed and compared with our reported diagnostic model (m/z 9,165, 12443 and PT). The linear regression model was found to be Proteomic index = $-0.827*\log_2$ (peak intensity of m/z 9,165) - $0.513*\log_2$ (peak intensity of m/z 12,443) –7.833. For predicting liver fibrosis (Ishak score >2), the diagnostic model was useful in identifying cases with significant liver fibrosis in the CHB patients with AUROC = 0.676 (95% CI 0.546-0.807, p = 0.017). However, the proteomic model was not significantly useful in predicting post-treatment samples (p-value >0.05). Figure 19 and Table 11 showed the ROC curves and comparison of the diagnostic accuracy between our diagnostic model, proteomic model, APRI and API using pre-treatment data respectively. Figure 20 and Table 12 showed the ROC curves and comparison of the diagnostic accuracy between our diagnostic model, proteomic model, APRI and API using post-treatment data respectively.

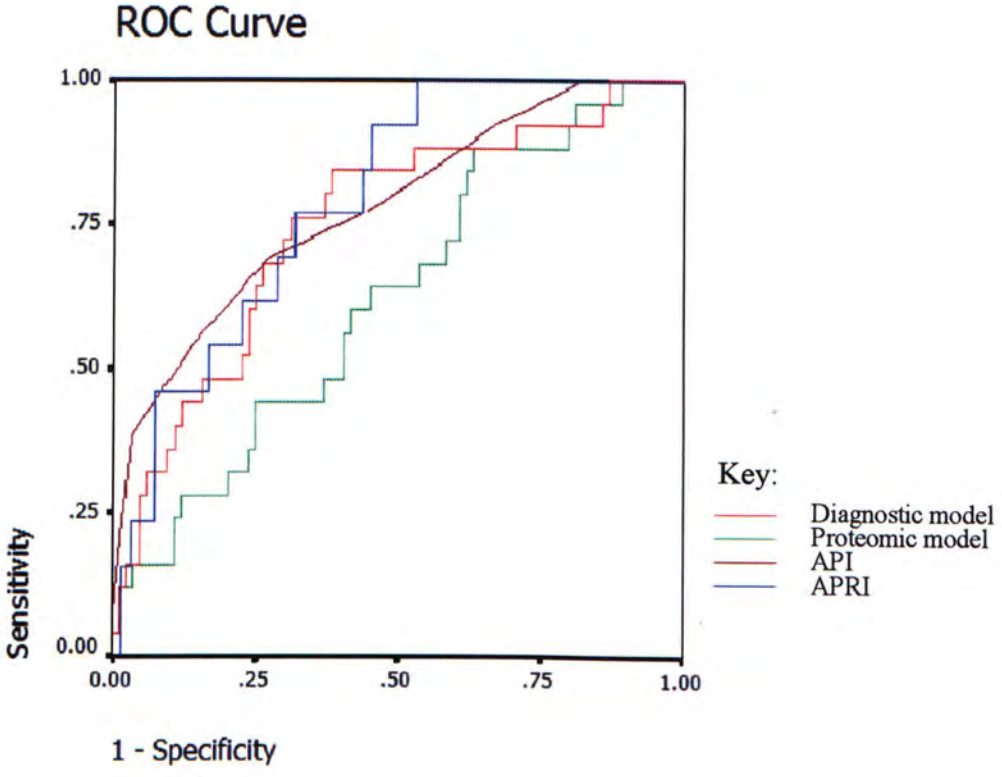


Figure 19 ROC curves of constructed diagnostic model, proteomic model, APRI and API models using pre-treatment samples.

Table 11 A comparison of AUROC of different non-invasive models using pre-treatment samples. (*p <0.05)

Non-invasive model	AUROC	p-value	95 % confidence interval	
			Lower bound	Upper bound
Our diagnostic model	0.726	0.002*	0.605	0.846
Proteomic model	0.675	0.017*	0.545	0.806
APRI	0.791	0.001*	0.676	0.907
API	0.781	0.001*	0.638	0.925

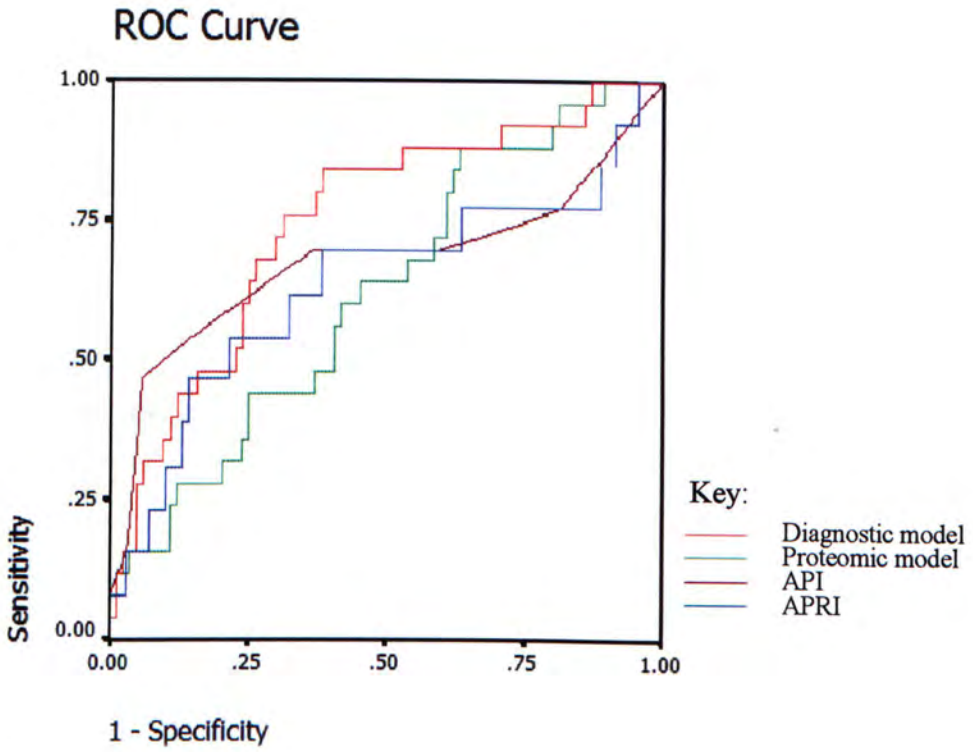


Figure 20 ROC curves of constructed diagnostic model, APRI and API models using post-treatment samples.

Table 12 A comparison of AUROC of different non-invasive models using post-treatment samples. (*p < 0.05)

Non-invasive model	AUROC	p-value	95 % confidence interval	
			Lower bound	Upper bound
Our diagnostic model	0.750	<0.0005*	0.640	0.860
Proteomic model	0.621	0.066	0.500	0.743
APRI	0.633	0.130	0.439	0.827
API	0.676	0.045*	0.471	0.880

2.4 Discussion

Currently there is no drug that can cure liver fibrosis directly though it is thought to be a reversible process. Antiviral therapy is mainly taken by patients with chronic hepatitis diseases to suppress or clear the virus load level so that the stress on the liver can be relieved and allow recovery on the fibrosis area. Several drugs were approved by the U.S. FDA and taken by patients with different treatment responses. In our study, patients were taken combined treatment of peginterferon and lamivudine, lamivudine or placebo treatment. Similar clinical trials studies [136-138] showed that decrease of PLT occurred during the anti-viral treatment, indicating the host responses might be affected by the medication. In our study approach, post-treatment samples were included and therefore a systematic evaluation on the serological parameters was done to investigate the effect of the antiviral treatment which might affect the diagnostic accuracy of the non-invasive models.

Recently, there were 3 studies using SELDI technology to find biomarkers in detecting HBV-associated liver fibrosis. [117-119] Though the recruited patients had the same etiology, different differential proteomic features were reported. There are two reasons for this discrepancy. One explanation is that there are thousands of potential differential features in serum and the chance of two groups finding a common biomarker is low. However, it is unlikely to occur. Another more reasonable explanation is that those markers are subject to systemic bias. It is well known that the SELDI biomarkers are sensitive to experimental details or to serum storage condition even same condition was used. [139]

Zhu *et al.* used peaks at m/z 7,772 and 3933 and Cui *et al.* selected peaks at m/z

2,055 and 3,166 for differentiating liver cirrhosis from healthy controls. Poon *et al.* reported 30 differential proteomic features and 7 of these together with serological markers were used to differentiate significant fibrosis. Compared our result with these 3 studies, our results were only consistent with Poon *et al.* Two proteomic features (m/z 9,498 and 9,714) were also found in our protein peaks list though these were not included in their diagnostic model. M/z 9,498 was believed to be m/z 9,452 in our study. This 9.5 kDa proteomic features showed negatively correlation with the Ishak scores in both studies. [119]

The consistency with Poon *et al.* study was due to similar sample processing method and statistical design used. In our study and Poon *et al.*, all samples were analyzed blindly without knowledge of the fibrosis stages. Pre-treatment samples were used for constructing diagnostic models. Spearman correlation test and SAM analysis were used to find differential proteomic features. SAM is a conservative multivariate bioinformatics test which allows the adjustment of median FDR and therefore a more reliable result could be achieved. Large proteomic data was usually encountered in proteomic profiling and more attention should be paid on multiple comparisons to avoid finding false differential markers by chance. Among 3 studies, only Poon *et al.* adopted the FDR approach and used correlation tests to search for differential peaks. Moreover, several SELDI studies also adopted this FDR concept in finding potential features from proteomic data. [140-142] Moreover, another criterion was set in our study. The differential proteomic features should be not only correlated with the degrees of liver fibrosis but also significantly higher or lower in patients with significant fibrosis. Mann-whitney U test was used to find proteomic features which could differentiate between

with and without significant fibrosis. With these stringent criteria, the resulting differential proteomic features were therefore not likely identified by chance or selection bias.

On the contrary, the discrepancy with these studies was due to several reasons. First, although all liver fibrosis/cirrhosis cases were HBV-associated, different controls samples were used. In Zhu *et al.* and Cui *et al.* studies, healthy patients were used as control while all recruited cases in Poon *et al.* were fibrosis cases from minimal fibrosis to cirrhosis which was the same as our study. It is obvious that different results can be obtained from different normal control. Second, different statistical designs can also give different results even using the same proteomic data. [143] Different statistical tools were used in model construction. Zhu *et al.* and Cui *et al.* used decision tree to construct the diagnostic models while Poon *et al.* developed artificial neural network (ANN) models for prediction of liver fibrosis. Linear regression model was used in our study instead. This could explain that why the common differential peaks found in our study was not included in Poon *et al.* ANN model.

In addition, previous studies of our team showed that case-control experiment design could lead to a discovery of a set of statistically valid differential proteomic features. However, only about 20% of them were genuinely associated with the diseases studied. [142][144] None of the previously published SELDI studies on discovering biomarkers for liver fibrosis had attempted to assess the biological relevance of the differential proteomic features. In the current study, in order to be considered as a potential diagnostic marker, the differential proteomic features needed to have association with at least 1 serological/biochemical biomarker reflecting liver function or

inflammation. This served as indirect evidence that the proteomic features had clinical meanings related to the liver fibrosis. Under these stringent criteria, only 4 proteomic features were found as potential markers in detecting liver fibrosis.

In addition, the diagnostic model was validated with the post-treatment samples. To the best of our knowledge, none of the previously published SELDI studies provided this information. However, due to limited samples provided, most of pre- and post treatment samples were from the same patient group which might introduce sampling bias and confounding factors into our study. Sample heterogeneity between the pre- and post-treatment groups were studied and no difference was found in the mean fibrosis score and prevalence of the significant liver fibrosis between groups, showing there was no sample bias between discovery and validation groups. By comparing the biochemical and serological parameters between discovery and validation groups, result showed that significant differences were found in PT, ALB, ALP, AST, ALT and HBV DNA levels. This implied that the antiviral drugs effectively suppressed the viral load and some liver functions were recovered.

Several preventative measures were done to avoid biased results caused by the nature of clinical samples used, sample storage conditions, experimental details, the mass spectrometric instrument, and/or bioinformatics analyses.[143][145] All recruited patients, including non-liver fibrosis cases, were suffered from CHB infections. All serum samples were collected and processed in the same clinical center under the same laboratory settings. To ensure the quality of the serum samples, they were stored at -80°C before analysis.

For the diagnostic model construction, PT was the only serological/biochemical

markers included. PT is a blood test that measures how long it takes blood to clot. As all the clotting factors except factor VIII are produced by liver, the contents of these clotting factors will decrease if liver injury occurs. Therefore, measurement of PT is an indicator of liver function. People suffered from liver injury will have longer PT, showing that the liver lose its function to synthesize enough clotting factors for blood clotting.

For pre-treatment group, the diagnostic model was useful in identifying cases with significant liver fibrosis in the CHB patients with AUROC = 0.726 (95% CI 0.605-0.846, $p < 0.005$). Comparison between our diagnostic model and other non-invasive models in detection of liver fibrosis was performed. APRI, API and FibroTest (or Fibrosure in USA) were the non-invasive models developed for liver fibrosis and showed good accuracy in detecting liver fibrosis in some studies. Hongbo *et al.* reported that the AUROC of APRI in detection of liver fibrosis was 0.70 in a Chinese CHB cohort. [146] API, developed by Poynard *et al.* mainly for CHC, [147] was evaluated by a Korean group in CHB cohort with AUROC = 0.751. [148] FibroTest (or FibroSure), constructed mainly for CHC-associated liver fibrosis, has been widely evaluated in the detection of liver fibrosis associated with different etiologies. A meta-analysis of FibroTest showed that the mean standardized AUROC was 0.84 (95% CI 0.82-0.87) and 0.80 (95% CI 0.77–0.84) for patients with chronic hepatitis C and CHB, respectively. [149] As the sample size of this study was still small, it is difficult to conclude that the clinical performance of FibroTest was better than of our model in diagnosis of liver fibrosis in CHB patients. With available data, the diagnostic accuracy of APRI and API were calculated and compared with our diagnostic model. As two parameters included in

FibroTest equation such as GGT and α-2 macroglobulin were not available in our cohort, comparison between our diagnostic model and FibroTest could not performed. The AUROC of APRI and API were found to be 0.791 and 0.781 respectively which was consistent with recent studies. Our diagnostic model showed comparable result with these non-invasive models.

However, from our result, APRI and API did not perform well in post-treatment samples with AUROC values of 0.633 ($p = 0.130$) and 0.676 ($p = 0.045$) respectively. It showed that these models were treatment-dependent. For APRI model, it performed well for pre-treatment samples but could not accurately classify significant fibrosis from minimal liver fibrosis patients. AST and PLT were involved in APRI model. As AST only reflected the liver injury, it did not represent the severe of liver fibrosis and the reduction of AST did not imply the reverse of fibrosis. In addition, AST was not significantly correlated with the degrees of liver fibrosis in our study. No significant difference was found between significant and minimal liver fibrosis group decreased its accuracy in APRI model. On the other hand, FibroTest was also found to be treatment-independent. Poynard *et al.* reported that the AUROC for the diagnosis of advanced fibrosis was 0.76 both at baseline and after treatment. [150] Another longitudinal study indicated that the AUROC of FibroTest was 0.77 [151], showing stable performance upon antiviral therapy. It is worth noting the AUROC reported in these studies was similar to the value of our diagnostic model. Hence, we could conclude our diagnostic model and FibroTest both were useful in monitoring changes of liver fibrosis during antiviral treatment with similar accuracy.

On the other hand, our diagnostic model consisted of PT as one of the parameters.

The accuracy of the model without PT was constructed and compared with our diagnostic model. It was found that the accuracy was dropped in model without PT parameter when classifying pre-treatment group. In addition, this model was unable to classify post-treatment group, indicating that PT was an important parameter in detecting liver fibrosis.

For pre-treatment samples, the formed diagnostic model had a good accuracy in detecting liver fibrosis and cirrhosis. This allowed us to classify the liver fibrosis patients into 3 different risk categories with the use of high and low cut-offs, 3.0844 and 1.3068. For those predictive score > 3.0844, they will be classified as high risk of liver fibrosis. Suspended patients with predictive score within 1.3068 and 3.0844 will be classified as medium risk. Those patients with predictive score < 1.3068 will be classified as low risk group. The positive predictive value for liver fibrosis in high risk group is 80%. This means 20 % of the cases with a positive blood test are false positive. Serious follow up, such as histological examination of liver biopsy, should be given to this group of high risk patients. For patients in medium risk group, they should be subjected to transient elastography for further examination. On the other hand, the negative predictive value in the low risk group is 96%, meaning that only 4 % of patients will be wrongly classified as low risk group. Low risk group will be excluded from both liver biopsy and antiviral treatment. For this group of patients, less aggressive follow-up, such as monthly blood test, can be provided.

Though the model was treatment-independent, different cut-offs were set for pre- and post-treatment. Another interesting finding was that same cut-offs were obtained for detecting fibrosis and cirrhosis. With the use of high and low cut-offs, 2.7303 and

0.3427, 3 different risk categories were made. For those predictive score > 2.7303, they will be classified as high risk of liver fibrosis. Suspended patients with predictive score within 0.3427 and 2.7303 will be classified as medium risk group. Those patients with predictive score < 0.3427 will be classified as low risk group. The negative predictive value in the low risk group is 92%, meaning that only 8 % of patients will be wrongly classified as low risk group. Low risk group will be excluded from both liver biopsy and antiviral treatment. For patients in medium risk group, they need to take liver biopsy for histological examination.

For both treatment naïve and post-treatment cases, it is worth nothing that similar high cut-off was obtained for both fibrosis and cirrhosis diagnosis. It may be due to the fact that our model was more applicable to detect liver cirrhosis than liver fibrosis. It might imply that our model was suitable in detecting liver cirrhosis instead of liver fibrosis. Another possible reason was due to the similar sample characteristics, the number of cases in fibrosis and cirrhosis groups were so similar that it was possible to get identical cut-offs with the same accuracy in these two groups. The model should be further investigated with larger sample group to confirm the reason for this phenomenon. However, the significant positive correlation between our predictive index and Ishak score strongly suggested that our model was useful for detecting both liver fibrosis and cirrhosis.

Different cut-offs between pre- and post-treatment could be explained by the different values of PT and the peak intensities in pre- and post-treatment groups even at the same fibrosis stage. This was also due to small sample size issue and the cut-off values carried certain degrees of error, making different cut-off values in two groups. On

the contrary, as the AUROC values were similar, the diagnostic performance of the model was not greatly affected. Further studies with larger sample size are needed before we could decide whether different cut-offs should be set for pre- and post-treatment groups.

2.5 Conclusion

Using the in-house magnetic bead-based proteomic profiling assays, 4 potential proteomic markers were found for the diagnosis of liver fibrosis in CHB patients. A diagnostic model composed of 2 proteomic markers (m/z values of 9,165 and 12,443) and PT was constructed. Independent validation showed the diagnostic performance of this model was comparable to those non-invasive diagnostic models previously published. However, it appeared that our model was more superior because our model maintained similar diagnostic performance for the serum samples collected after antiviral treatment.

Section 3 Identification of proteomic features to form diagnostic fingerprint for the detection of liver fibrosis in patients with chronic hepatitis B infection

3.1 Introduction

As mentioned, SELDI has become a useful tool in proteomic profiling for different diseases in biomarker discovery. However, only knowing the mass values of the differential protein peaks limits its usefulness and significance in clinical field. Wu *et al.* discovered 45 differential protein peaks between HBV-related hepatocelluar carcinoma (HCC) and liver cirrhosis groups. [152] Among them, two protein features (m/z 9,297 and 29,941) and AFP were selected to form a diagnostic model with 79.3% sensitivity and 90.0% specificity. As the protein identities were unknown, the validity and significance of protein biomarkers could not be evaluated. Furthermore, without differential peak identities, it was impossible to study why such abnormalities would occur and how the disease enhances such effects. This was well illustrated by Liu *et al.* [153] Factor analysis was performed from SELDI MS data to characterize serum low-molecular weight proteins/peptides in liver injury. A group of proteins/peptides was all down-regulated in hepatitis serum samples and factor analysis showed that they were highly correlated and influenced by the same factor. However, since the protein identities of these protein group as well as their chemical properties were unknown, it was not possible to draw definite conclusion regarding the nature of the peaks and factors.

Also, protein identity could provide evidence that the differential protein peaks were not artifacts which were common in biomarker discovery studies. In addition,

doubly or triply charged protein signals were usually occurred in SELDI-TOFMS. Sometimes the same protein with different charges may be selected in the same diagnostic model which could affect the diagnostic accuracy. [154] On the other hand, wrong classification of a protein as a multiple charged species can lead to a loss of useful information. The only solution was protein identification of the interested peaks. Kanmura *et al.* found a numbers of peaks that might represent doubly charged peaks. They suspected that the differential peaks at 4,067 m/z might be the doubly charged form of the 8,138 m/z peak. Surprisingly, after protein identification, they concluded that the peak at 4,067 m/z did not appear to be the doubly charged of 8,138 m/z peak. This demonstrated that protein identification was indispensable for biomarker discovery before concrete conclusion could be made.

The major limitation of SELDI technology was that the retained proteins on the proteinchip could not be recovered for protein identification. Other purification techniques were needed to enrich the corresponding proteins. This made the whole procedures complicated and tedious. Several approaches had been reported to enrich the interested proteins. Zinkin *et al.* enriched the interested SELDI peak by using a spin column containing Q Ceramic HyperD F beads. The column was washed thrice with 50mmol/L Tris (pH9) and incubated with diluted serum (40μL serum + 500μL 50mmol/L Tris (pH9) for 90 minutes. The column was then washed with 150μL of 50mmol/L Tris (pH9) twice and the flow through pooled and dried in a Speedvac centrifuge for 1-D SDS PAGE. Wu *et al.* proposed that the biomarkers could be purified by using affinity capture, anion exchange, size exclusion and reversed-phase chromatography depending on their individual biochemical properties. [132] 5 mL of

serum was first subjected to blue sepharose to deplete albumin, the remaining proteins were then separated by C18 reverse-phase chromatography. The purified biomarkers were further digested and subjected to ESI-MS/MS for sequence identification. Alternately, the fractions obtained from liquid chromatography could be further separated by SDS PAGE. The interested band was then excised, digested and identified by MALDI-TOF-MSMS. Another approach was ACN precipitation reported by He *et al.* [155] Pure ACN was added to 200μL of serum to remove high molecular weight proteins. The supernant was then concentrated for 2D-PAGE analysis. All methods required large amount of serum which might not be affordable by many research groups as patient samples were limited.

Western blotting and immundepletion/immunoprecipitation was usually followed to confirm the protein identity of the interested peak. [156] Furthermore, immunoassays were also done to confirm the identities of the biomarkers and evaluate the feasibility of using these biomarkers for clinical use. [135]

In this section, we aimed to identify the protein identities of the protein features selected in the diagnostic model by using the developed methodology mentioned in section 1. The validity of the protein identity to the targeted peak was then confirmed by immunodepletion. The concentrations of the biomarkers were then further quantified by immunoassay.

3.2 Materials and methods

3.2.1 Non-reducing 2-D gel electrophoresis

Proteins eluted from the magnetic beads was dried at 45 °C using speedvac concentrator (Enppendorf) and reconstituted with 185μL of rehydration buffer (8M Urea,

2% CHAPS, 0.2% Biolyte 3-10 ampholyte, 0.001% bromophenol blue, 1mM EDTA). An immobilized pH gradient (IPG) strip (11cm 3-10NL, Bio-Rad Laboratories) was rehydrated with the sample overnight. For the first dimension IEF separation, the running condition was as follows: 100V for 10min, 250V for 65min, 500V for 25min, 1000V for 40min, and finally 8000V for 140min. Second dimension SDS-PAGE was performed on 4-12% Bris-Tris polyacrylamide gels (Bio-Rad Laboratories) and the proteins were separated at 200 V for 40min in ice bath. The 2D gel was then stained with silver nitrate using Amersham PlusOne silver staining kit (GE Healthcare) with some modifications to reduce the loss of proteins with MW < 10 kDa. The gel was fixed in 40% methanol/10% acetic acid for 30 min and then sensitized by thiosulfate solution with 25% w/v Glutardialdehyde. Washed with 30% ethanol for 15 min, the gel was immersed with silver solution in 30% ethanol and 37% w/v formaldehyde for 1 hr. The gel was washed with Milli-Q water for 1min for 3 times and then developed in sodium carbonate solution with 37%w/v formaldehye. The development was stopped by EDTA solution and the gel was rinsed with Milli-Q water for 3 times before performing image analysis with GS-800 calibrated densitometer (Bio-Rad Laboratories).

3.2.2 Protein identification of eluted proteins

Protein spots were excised from silver stained gels. The gel pieces were destained, reduced with 1.75% DTT, alkylated with 350 mM IAA, and digested with modified porcine trypsin overnight (Sequencing grade modified typsin from Promega, Madison, WI). The tryptic digest was harvested and cleaned up with C18 ZipTips (Millipore). The cleaned tryptic peptides were subjected to MALDI-TOF/TOF MS

(Ultraflex-III, Bruker Daltonics) with α-cyano 4-hydroxy cinnamic acid as matrix. The MS and MS/MS spectra were then processed with Data Explorer software (Flex analysis 3.0). The spectra were subjected to Gaussian smoothing with a filter width of 5 points, and the baselines were corrected with default settings. Peaks were detected based on a S/N > 15. The MS spectrum data were searched via the online Mascot search engine to obtain the protein identity by undertaking the peptide mass fingerprinting (PMF) approach. Tandem MS data were subjected to MS/MS ion search via the Mascot search engine to obtain the protein sequence of a particular peptide. For the search parameters, the 1 missed cleavage in trypsin digestion was allowed; partial oxidation of methionine, phosphorylation of serine/threonine/tyrosine, and iodoacetamide modification of cysteine residues were selected. The error tolerance values of the parent peptides and the MS/MS ion masses were 200ppm and 0.5 Da, respectively. A protein identification result was considered valid when both PMF and MS/MS ion search identified the same protein as the statistically significant hit from the NCBInr database, and/or when MS/MS ion search identified at least tryptic peptides with sequences from the same protein as the statistically significant hits.

3.2.3 Immunodepletion of apolipoprotein C-III

Protein G Agarose bead (Pierce®) was used to capture polyclonal goat anti-human apolipoprotein C-III antibody (Abcam) to deplete apolipoprotein C-III in serum. 3 conditions were set: Sepharose 4B with antibody, Protein G Agarose without and with anti-apolipoprotein antibody. The first two conditions were served as negative controls. Briefly, beads were pre-equilibrated with PBS buffer for 3 times. Then, 40μL of anti-

apolipoprotein antibody or PBS buffer (negative control) was added to the beads and incubated at 4 °C overnight. The supernant was then removed, and the beads were washed with PBS buffer for 3 times. 2 μL of serum samples diluted 100-fold in PBS was added to the washed beads and incubated at 4 °C overnight. The supernatant was finally collected for serum proteomic profiling using the developed method, as described in the previous section. Figure 21 showed the experimental flow of the immundepletion assay.

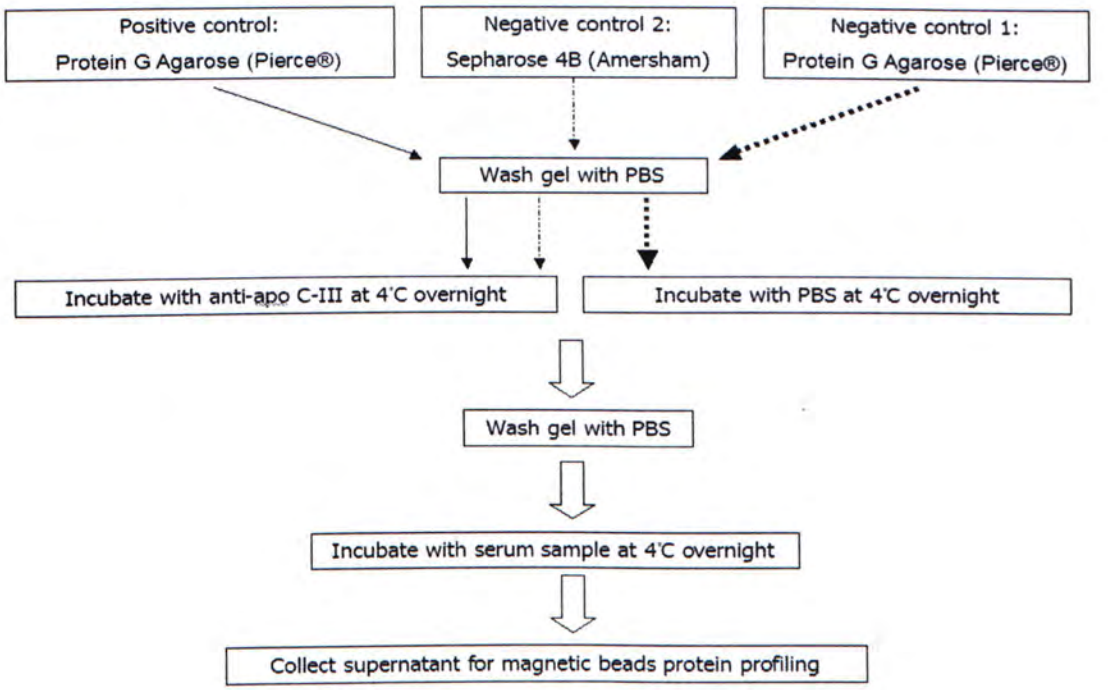


Figure 21 Experimental flow of the immunodepletion assay

3.2.4 Measurement of serum levels of apolipoproteins

Human apolipoprotein LINCOplex kit (Millipore) was used to measure the amount of apolipoproteins (apoA-I, apoA-II, apo B, apoC-II, apoC-III and apoE) in patient serum. 2μL of whole serum was diluted in 50,000 fold with assay buffer (10 mM PBS with 0.08% Sodium Azide and 1% BSA, pH 7.4). The whole procedure was followed by manufacturer's protocol. Briefly describe 10μL of the diluted samples or calibration standards were added to each well and mixed with antibody-immobilized beads for 1hr incubation. After washing with washing buffer (10 mM PBS with 0.08% Sodium Azide, and 0.05% Tween 20, pH 7.4), detection antibody was added to each well and incubated for 30min. With 3 washing steps, Streptavidin-Phycoerythrin was added to each well and incubated for another 30min. Shield fluid was finally added to each well. The concentration of the apolipoproteins was then measured by Bio-Plex 200 system (Bio-Rad Labs).

3.2.5 Statistical Analysis

Spearman rank correlation analysis was performed to investigate the correlation between concentrations of apolipoproteins and the degree of liver fibrosis in serum level. The correlation between the protein peak intensity of 9,153 m/z and the concentration of apolipoprotein C-III measured from LINCOplex kit was studied.

3.3 Results

3.3.1 Protein identification of the protein marker in the diagnostic model

It was obvious that the overall protein quantity of the low peak intensity gel (samples with minimal 9.2kDa and 12.4 kDa peak intensity) was less than that of high peak intensity gel (samples with intense 9.2kDa and 12.4 kDa peak intensity)while the gel pattern was the same (Figure 22). Gel pattern was similar between the two gels, indicating the gels were undergone similar condition and background. By studying the intensity contrast of these two gels with mass estimation from the protein gel marker, several spots were excised and protein identification work was done. The marked area on the gel photos were the potential protein spots corresponding to 9.2kDa and 12.4 kDa protein peaks. The 9.2 kDa protein selected in the predictive model was found to be apolipoprotein C-III (apo C-III). 2 peptides were matched in Mascot MS/MS database with a score of 166 (expected value = 5.13 e-13). The two peptides were m/z 1,716 and m/z 2,016 and the corresponding peptide sequences were K.DALSSVQEQVAQQAR.G and K.TAKDALSSVQESQVAQQAR. The resulting spectra and the corresponding database search result were shown in Figures 22-24. The protein identity of 12.4 kDa could not be found and was still unknown.

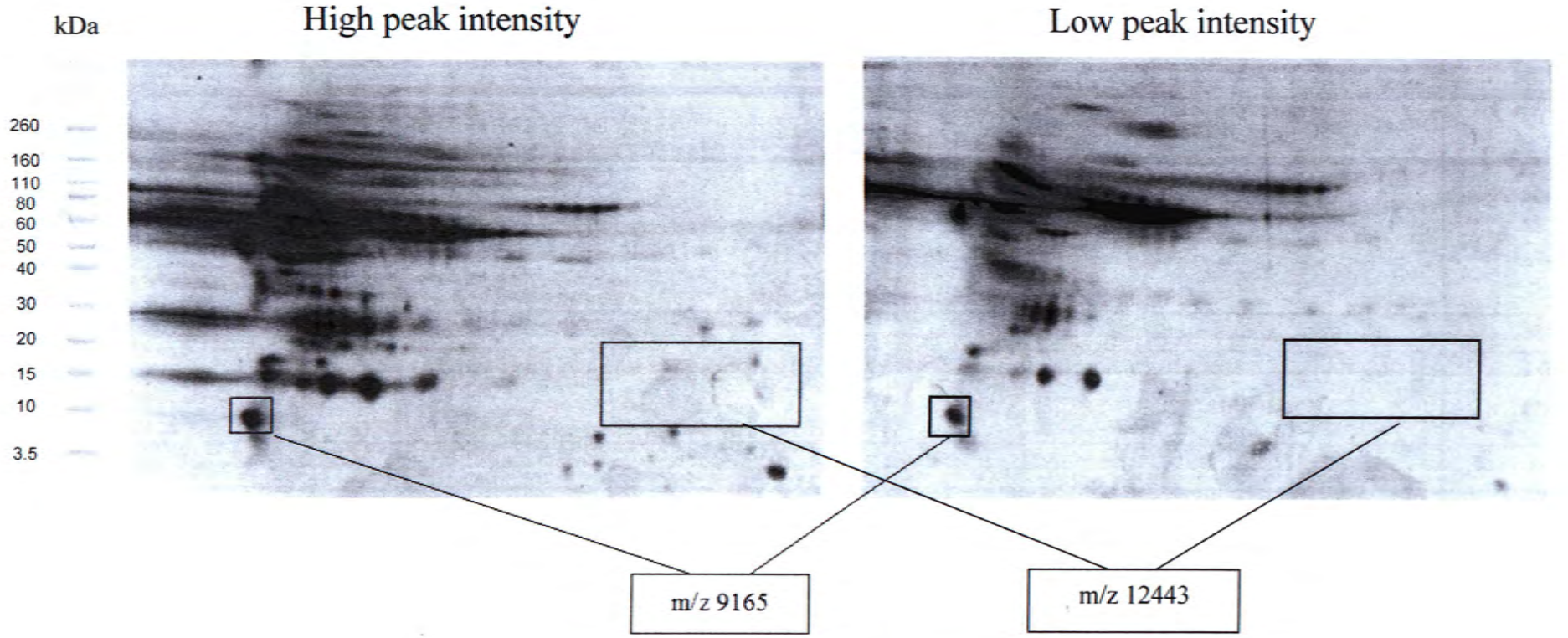
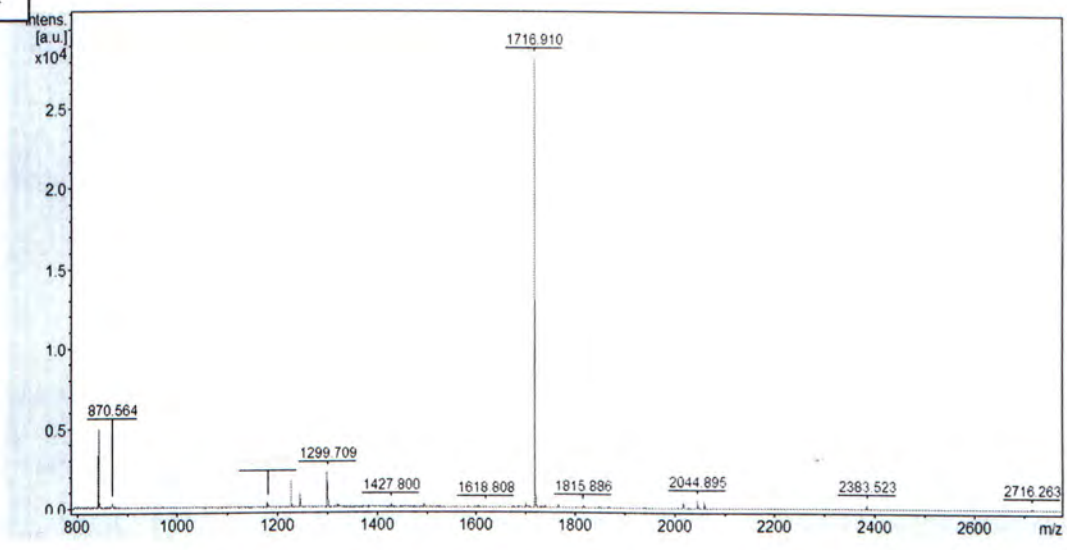


Figure 22 Gel photos of high and low intensity of 9.2 and 12.4 kDa. Small rectangle: 9.2kDa, Apo C-III; large rectangular: potential 12.4kDa protein marker

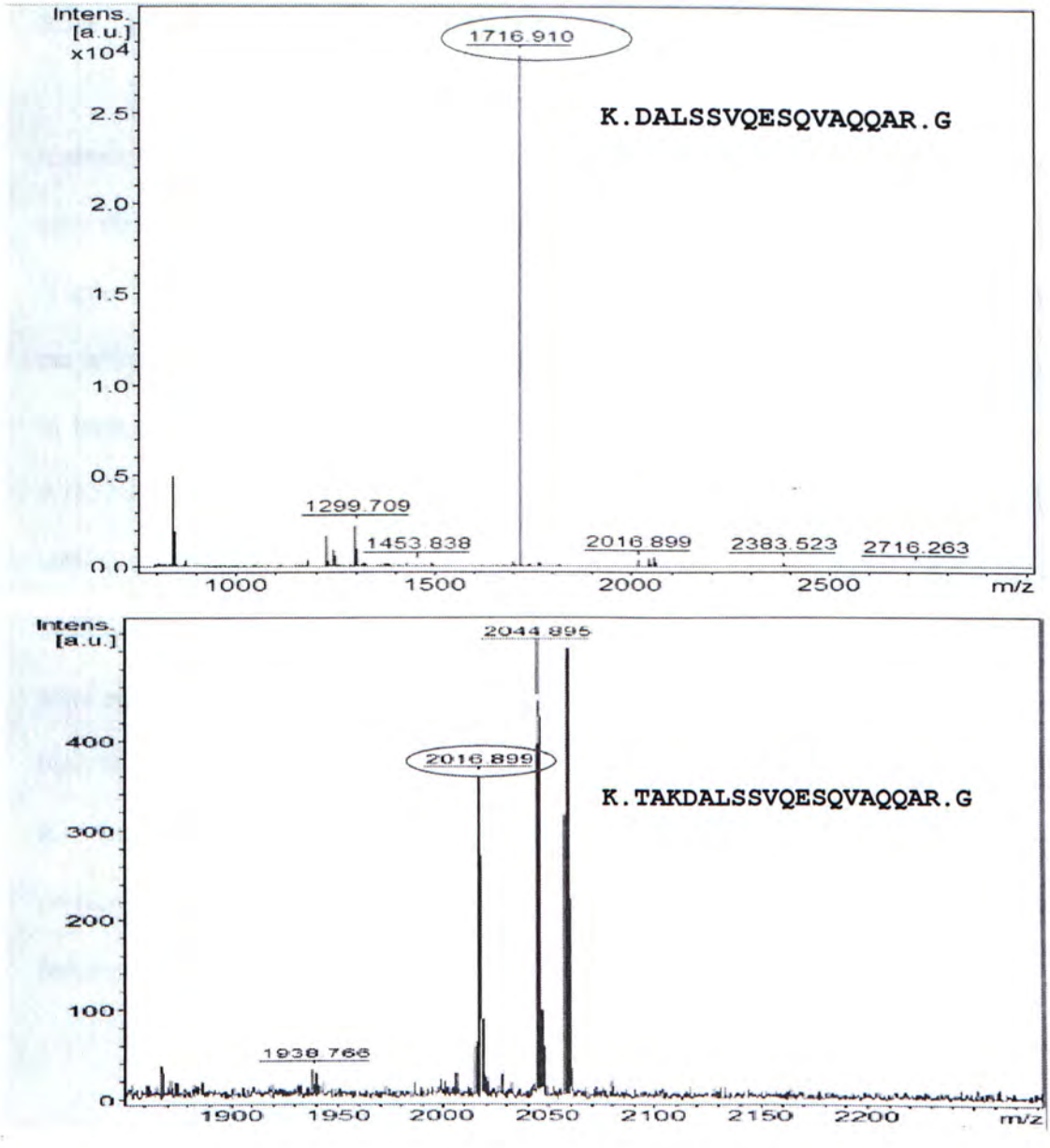
A



B

m/z	SN	Quality F	Res.	Intens.	Area	Bk. Peak
822.230	3.8	76	7833	100.91	21	false
842.516	185.5	14416	6567	4831.44	1152	true
870.564	12.3	1440	8471	317.13	64	false
1179.728	17.7	2226	9276	344.79	101	true
1226.680	90.3	11301	9076	1689.08	529	false
1245.028	45.1	2105	5441	814.39	438	false
1299.709	127.4	23629	9526	2190.23	730	false
1315.693	6.6	301	9069	109.34	42	false
1320.735	7.9	614	8086	130.45	54	false
1323.779	5.5	309	12751	93.23	27	false
1365.740	4.4	177	9974	70.88	28	true
1373.048	5.2	277	5766	80.68	48	false
1380.853	8.9	810	8140	136.14	59	false
1388.753	2.6	51	8558	38.98	17	false
1427.800	13.8	1829	9561	199.54	74	false
1453.838	3.2	66	5427	44.61	35	false
1475.892	3.0	121	7280	41.48	24	true
1493.855	11.7	1534	9605	158.55	70	true
1618.808	4.8	428	4347	57.73	68	false
1656.857	2.4	94	7674	28.93	20	false
1674.769	3.3	90	13004	38.26	16	false
1688.975	2.3	70	5388	26.01	25	false
1698.899	20.9	4231	11319	236.05	113	false
1702.253	4.9	203	5273	55.49	60	false
1707.791	4.4	201	7070	49.41	41	true
1716.910	2622.9	87870	9124	29035.91	17270	false
1738.897	7.7	250	10545	83.09	42	false
1765.767	15.0	2161	13456	157.21	66	true
1815.886	7.7	896	10477	76.19	40	false
1867.889	2.0	57	9603	19.00	12	false
1938.766	2.8	103	11837	23.96	13	false
2015.916	12.4	103	11349	96.98	56	false
2016.899	22.9	2514	12555	180.34	98	false
2044.895	47.0	4696	13683	365.16	182	false
2057.907	45.8	8125	13535	345.49	182	false
2383.523	20.3	563	14509	101.86	58	false
2716.263	4.1	85	17066	16.78	10	false

Figure 23 A) The peptide mass fingerprinting spectrum of 9.2kDa. B) Mass list of PMF spectrum of 9.2kDa.



Results List

1. APOC3_HUMAN **Mass:** 10846 **Score:** 166 **Expect:** 5.1e-13 **Queries matched:** 2
Apolipoprotein C-III OS=Homo sapiens GN=APOC3 PE=1 SV=1

Observed	Mr(expt)	Mr(calc)	ppm	Start	End	Miss	Ions	Peptide
1716.9006	1715.8933	1715.8438	28.8	45	60	0	123	K.DALSSVQESQVAQQAR.G
2016.8904	2015.8831	2016.0236	-69.68	42	60	1	28	K.TAKDALSSVQESQVAQQAR.G

Figure 24 Zoomed spectra showing the two matched peptides of apo C-III

3.3.2 Immunodepletion of apolipoprotein C-III

To further confirm the m/z 9,165 proteomic feature was apo C-III, immunodepletion assay specific to apo C-III was performed. Depletion of apo C-III was only observed when both protein G agarose and anti-apo C-III antibody were present (Figure 25). 3 peaks (9,165, 9,452 and 9,735 m/z) were depleted while other peaks were not affected. Except these 3 peaks, similar profile patterns and intensities were observed in both the experimental condition and the negative controls. The peak intensities of 9,165, 9,452 and 9,735 m/z in the negative controls were comparable to the profiles without immunodepletion, showing that the sample loss was negligible during sample manipulation. When protein G-beads were replaced with Sepharose 4B beads, 3 peaks were not depleted. This was consistent to the fact that the anti-apolipoprotein could not bind to the Sepharose 4B bead surfaces and therefore no depletion could take place in this condition. As a result, this immunodepletion experiment confirmed that m/z 9,165 proteomic feature was a variant of apo C-III. Our result also showed that proteomic features at m/z 9,452 and m/z 9,735 were other two variants of apo C-III.

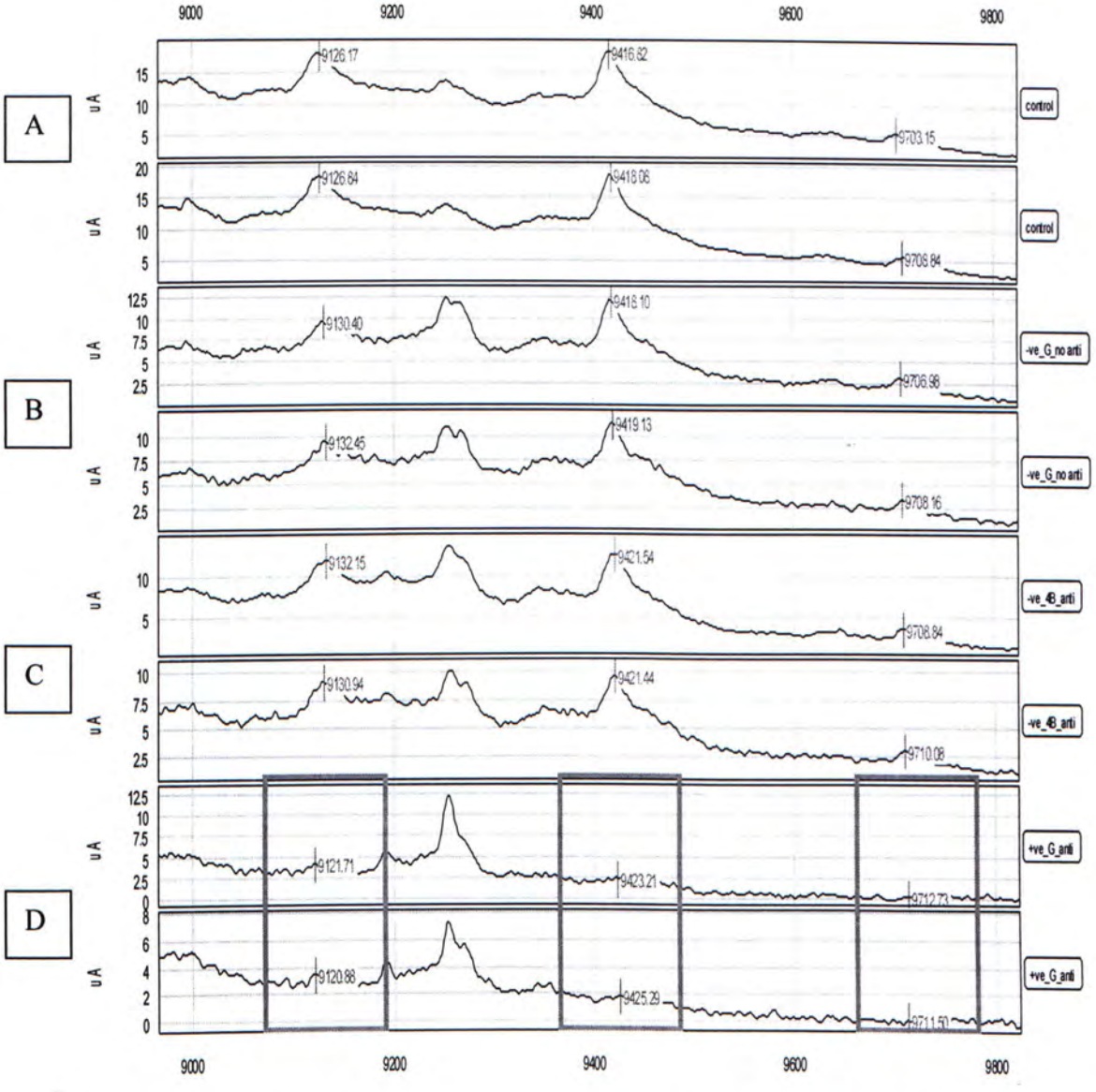


Figure 25 The proteomic profiles of a representative serum sample before and after immunodepletion of apo C-III.

A: Not subjected to immunodepletion; B: Only Protein G beads was included during immunodepletion; C: Sepharose 4B beads and anti-apo C-III antibody were included; D: Protein G beads and anti-apo C-III antibody were included.

3.3.3 Serum levels of Apolipoproteins and Their Associations with Liver Fibrosis

Serum concentrations of six apolipoproteins, including apo C-III, were measured, and correlation analysis between their serum concentrations and Ishak scores were carried out. Two apolipoproteins, apo A-II ($r = -0.156$, $p = 0.016$) and apo C-III ($r = 0.153$, $p = 0.019$), were negatively correlated with the degree of liver fibrosis. Other apolipoproteins did not show statistical significance with the progression of liver fibrosis. The correlations between six apolipoproteins and the Ishak scores were summarized in Table 13.

As expected, the intensities of the 2 protein peaks corresponding to apo C-III variants, i.e. peaks at m/z 9,165 and 9,452, were positively correlated with the apo C-III concentration obtained by immunoassay with correlation coefficient of 0.257 ($p = 0.015$) and 0.223 ($p = 0.036$) respectively. For m/z 9,735 peak, it could be considered as marginally significant. In addition, the sum of these 3 peak intensities also showed positive correlation with the apo C-III concentration ($r = 0.245$, $p = 0.020$). Table 14 showed the correlations between different peaks and the apo C-III concentration.

Table 13 Correlation between apolipoproteins and Ishak scores. (*p <0.05)

Apolipoprotein	Correlation coefficient	p-value
Apo A-I	-0.071	0.278
Apo A-II	-0.156	0.016*
Apo B	0.042	0.515
Apo C-II	-0.120	0.064
Apo C-III	-0.153	0.019*
Apo E	<0.0005	0.992

Table 14 Correlation between protein peaks intensities and apo C-III concentration (*p <0.05)

Proteomic feature	Correlation coefficient	p-value
9,165 m/z	0.257	0.015*
9,452 m/z	0.223	0.036*
9,735 m/z	0.204	0.055
Percentage of 9,165 m/z peak intensity	0.049	0.624
Percentage of 9,452 m/z peak intensity	-0.076	0.445
Percentage of 9,735 m/z peak intensity	0.073	0.462
Sum of 3 peaks	0.245	0.020*

3.4 Discussion

2-D PAGE followed by in-gel digestion was a common approach used for protein identification. 1-D SDS PAGE or 2-D PAGE could be used, depending on the complexity of the protein fraction. 2-D PAGE was used in this study was mainly because the eluted proteins consisted of many proteins with similar molecular weights. Proteins with similar molecular weights would appear as one single band in 1-D SDS PAGE, and were not suitable for protein identification. 5 samples with most intense / minimal peak intensities of 9.2 kDa and 12.4 kDa were pooled separately to find the gel spots corresponding to the interested peaks. By comparing the intensity contrast of the two gels with the help of protein marker for mass estimation, potential gel spots were selected, excised and digested for sequence identification. Apo C-III was successfully identified for m/z 9,165 peak in the diagnostic model. Unfortunately, protein identity of m/z 12,443 peak was still unknown. The failure could be explained by the following reason. It might be because the peptide sequence was not available in the database or the protein spots were mixed with other proteins so the targeted protein was masked by other proteins.

Immunodepletion of apo C-III was carried out to confirm the identity of m/z 9,165 peak. It was confirmed by depleting the 9,165 m/z peak in an experimental condition. The anti-apo C-III specifically bind to the Protein G beads for depleting apo C-III in serum but not for Sepharose CL 4B beads. Polyclonal goat apo C-III antibody with IgG isotype was used for immunodepletion. Protein G is a bacterial cell wall protein expressed in group G Streptococci. It binds to most mammalian immunoglobulins primarily through their Fc regions. Native Protein G contains two immunoglobulin

binding sites, as well as albumin and cell surface binding sites. However, since albumin appears as the major contaminant in serum, a recombinant form of protein G which albumin and cell surface binding sites are eliminated is made to reduce nonspecific binding. Sepharose CL 4B beads do not have immunoglobulin binding sites for anti-apo C-III and therefore depletion cannot occur in Sepharose CL 4B condition. After immunodepletion, other protein peaks remained unaffected in terms of peak shape and intensity, showing that depletion was so specific to the particular proteins and magnetic beads-based proteomic profiling was not affected by immuodepletion.

Together with 9,165 m/z, two other neighboring peaks (9,452 and 9,735 m/z) were also depleted by apo C-III antibody, indicating that there were 2 other isoforms of apo C-III detected in the magnetic beads proteomic profile. According to the literatures, apo C-III consists of 79 amino acid residues and exists in three isoforms depending on the number of post-translational sialyl groups (0 to 2) terminating the oligosaccharidic portions of the protein. The sugar moiety of apo C-III consists of 1 galactose, 1 N-acetyl-galactosammine and 0, 1 or 2 sialic acids molecules. [157] These three isoforms, apo $C\text{-}III_0$, apo $C\text{-}III_1$ and apo $C\text{-}III_2$ contribute 10, 55 and 35 % of total apo C-III respectively. Figure 26 showed the three peaks and their corresponding structures in the proteomic profile.

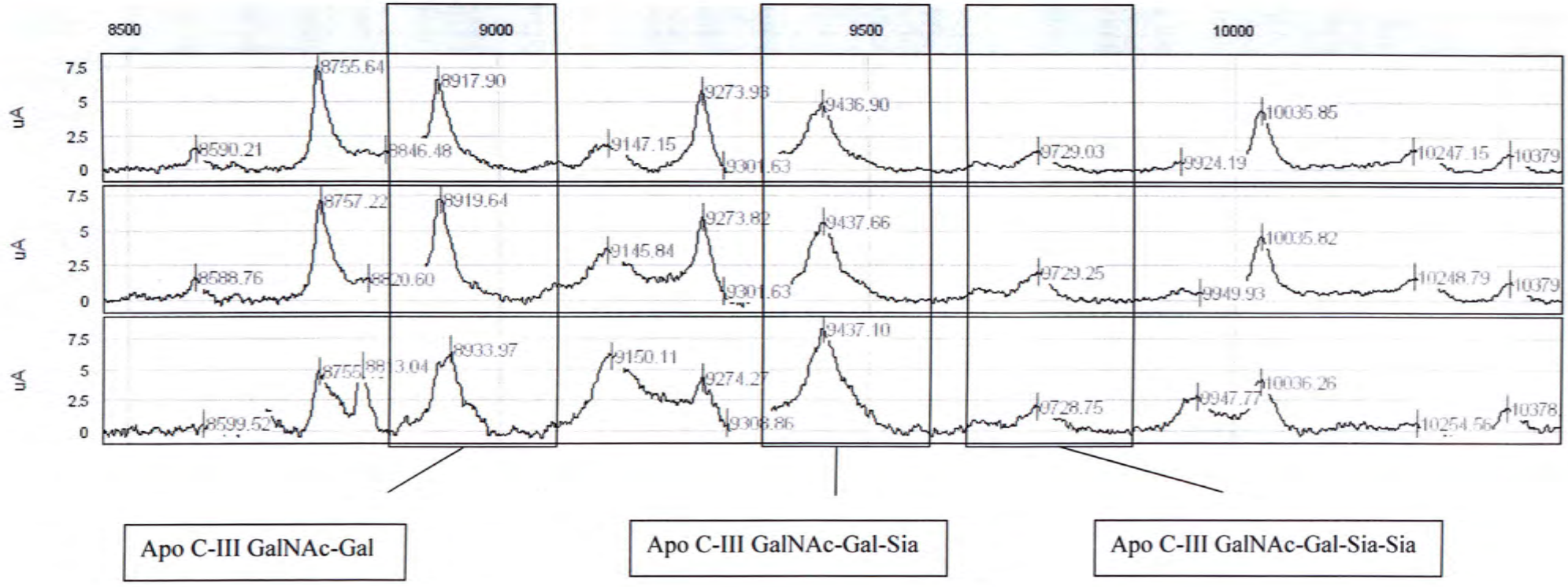


Figure 26 Representative spectra showing three apo C-III peaks and their corresponding structures in the proteomic profile.

Sialic acid is a family of N- and O-substituted derivatives of neuraminic acid which is an amino sugar. [158] Over 40 naturally occurring derivatives have been identified and the most common sialic acids found on human plasma proteins and glycolipid is 5-N-acetylneuraminic acid. The functions of sialic acids in biological system include conformation stabilization, protease resistance, charge, enhancement of water binding capacity, cellular recognition, protein targeting and developmental regulation. The sialylation of apolipoprotein is still not completely understood; however, it is known that the sialic acid content of individual apolipoproteins will vary in response to physiological conditions. Disease states may selectively alter any biosynthesis stage and passed into plasma, generating an altered glyco-isoform distribution. [159] Glycomic structures and distributions may therefore provide insight into disease mechanisms that impact on cellular events leading to protein export. Apolipoprotein sialylation is an intracellular process driven by Golgi-membrane-bound enzymes, sialyltransferases. The glycol-isoform ratio change has been studied for various diseases. Excess sialylation of apo C-III, apo C-III_2, was found in hypertriglyceridemic subjects. This may be due to its higher affinity for VLDL than apo C-III_0 and apo C-III_1. Also, apo C-III_2 is a poorer inhibitor of VLDL binding to the purported lipolysis stimulated receptor than apo C-III_0 and apo C-III_1. Harvey *et al.* demonstrated that the apo C-III_1/ apo C-III_2 ratio changed on bariatric surgery, chronic or severe liver disease. [159] A decrease in apo C-III_1/ apo C-III_2 was found after bariatric surgery while it was increased by liver disease. However, other studies showed that no significant difference was found in the distribution of these three apo C-III isoforms in a CHC study. [160] Mauger *et al.* also reported that no correlation was

observed between the relative proportion of each isoform and anthropometric variables and suggested that the degree of sialylation of apo C-III might not be related to neither plasma lipid levels nor obesity indices. [161] Our study followed the latter results. In our study, all three apo C-III isoforms were down-regulated with the degree of liver fibrosis. Furthermore, positive significant correlations were found between the three isoforms, further validating the coherent of the correlation direction trend. From our result, there was no implication that the degree of sialylation was related to liver fibrosis as the relative proportion of apo C-III was not significantly correlated with liver fibrosis. Our results might be account for our purposed mechanism rather than sialylation of apo C-III. In addition, our result was consistent with our previous pilot study. [119] Two protein peaks, 9,498 and 9,714 m/z were negatively correlated with liver fibrosis. It was believed that these 2 peaks were sialylated forms of apo C-III. Also, Molina *et al.* reported that apo C-III was a candidate biomarker in plasma associated with the resolution of HCV infection. [160] These two studies further supported our result that apo C-III was associated with liver function. It also implied that apo C-III was not only valid in CHB but also in CHC cases. Further studies were needed to examine the clinical performance of apo C-III in detecting liver fibrosis caused by different etiologies.

Apo C-III is the most abundant C-apolipoprotein in human and is found on the surface of very low density lipoprotein, low density lipoproteins, chylomicrons and high density lipoproteins (HDL). [162][163] Apo C-III is mainly synthesized in liver and intestine. It is a well known inhibitor of lipoprotein lipase and hence delays lipoprotein triglyceride lipolysis. Overexpression of apo C-III leads to the increase in plasma triglyceride levels. Apo C-III is thought to interfere with the binding of apo C-II to

lipoprotein lipase and to prevent binding of apo B particles to LDL receptors. Elevated apo C-III level is usually found in patients with metabolic syndrome and has also been associated with insulin resistance. Also, it has been implicated in coronary artery disease. It was shown that lifelong deficiency of apo C-III had a cardioprotective effect. [164]

On the contrary, in our study, apo C-III was down-regulated in CHB patients with liver fibrosis which was an opposite situation to most lipid and cardiovascular studies. The cause of this down-regulation might be associated with the dysregulation of TGF-β1 in fibrogenesis. TGF-β1, a well studied cytokine, is believed to play a key role in fibrogenesis. TGF-β1 carries out multiple biological functions including development, cell growth, differentiation, cell adhesion, migration and contribution to the regulation of the production, degradation and accumulation of ECM proteins. High levels of TGF-β1 are often found in liver fibrosis and it has been considered as a mediator of liver fibrosis. It is well known that liver fibrosis occurs under the activation of quiescent HSCs that will further enhance the growth of myofibroblasts to disrupt the liver architecture. Researches found that this consequence was associated with the overexpression and activation of TGF-β1, implying that TGF-β1 may be one of the first signals to activated quiescent HSCs. TGF-β1 enhances ECM synthesis as well as inhibits ECM degradation by suppressing expression of matrix-degrading enzymes and promoting expression of matrix metalloproteinase inhibitors (MMP). [165]

Hepatocyte nuclear factor -4α (HNF-4α), an apo C-III target gene, was inhibited by TGF-β1. [166] It was showed that TGF-β1 affected both HNF-4α mRNA and protein levels by decreasing the DNA-binding capacity of HNF-4α and inducing degradation of HNF-4α in the proteasome in HepG2 cells, respectively. HNF-4α is abundant in liver,

intestine, pancreas and kidney. It is a highly conserved member of the nuclear receptor superfamily that was initially identified as a transcriptional factor required for liver-specific gene expression. [167] It generally acts as a positive transcriptional regulator of many hepatocyte genes. HNF-4 was shown to be an important candidate for liver development; expression levels of large number of genes whose products were essential for mature liver function were disrupted in HNF-4 null livers. [168] In addition, Lucas *et al.* postulated that the loss of hepatic specific functions during the progression of liver fibrosis could be explained by the inhibition of HNF-4α by TGF-β1 and nitric oxide (NO). [166] Moreover, many studies reported that apo C-III gene was transcriptionsally dependent on HNF-4α in the liver, by interacting at least two sites of the apo C-III promoter. [169-171] Therefore, the reduction of apo C-III in liver fibrosis patients might be due to the inhibition of HNF-4α.

As apolipoproteins are mainly synthesized in liver, their contents in plasma may reflect liver function. Also, the association of apolipoproteins A1/C3/A4/A5 gene cluster that modulates plasma triglycerides reveals the inter-correlation of different apolipoproteins which may be also related to liver fibrosis. [172] Therefore, after identifying apo C-III as one of the diagnostic markers in detecting liver fibrosis, immunoassay in quantifying six apolipoproteins was carried out to study the correlation between other apolipoproteins and liver fibrosis.

Apart from apo C-III, other apolipoproteins such as apo A-I, apo B and apo E have been studied and found to have clinical values in liver diseases. Apo A-I was correlated with patients with chronic liver disease such as primary biliary cirrhosis [173] and liver fibrosis. [48] [174] It has been included in several non-invasive models in detecting liver

fibrosis like FibroTest [175] and PGAA index. [176] Apo B levels were significantly lower in patients with CHC. It was negatively correlated with steatosis and HCV viral load and became an indictor to confirm the interaction between hepatitis C infection and β-lipoprotein metabolism. [177] For apo E, Ferre *et al.* reported that hypercholesterolemic ApoE-/- mice were more susceptible to develop severe liver injury. [178] Also, the mRNA level of apo E was found to be down-regulated to the liver injury in mice model. [179]

However, among six apolipoproteins (apo A-I, A-II, B, C-II, C-III and E), only apo A-II and C-III were negatively correlated with the degree of liver fibrosis in our study. This finding was consistent with recent study. [180] D.Q. Shih *et al.* found that patients with HNF-4α mutation had a significant reduction in apo A-II and apo C-III but not in apo A-I and B though they were all HNF-4α target gene and mainly expressed in the liver. By further genotypic and phenotypic comparison, they confirmed that these reductions were due to HNF-4α haploinsufficiency rather than factors from family aggregation or diabetes. It was known that both apo A-II and C-III genes contained an HNF-4α binding site and therefore this reduction could be explained by the decreased binding capacity of HNF-4α caused by TGF-β1 in our proposed mechanism. [181]

The down-regulated correlation between apo C-III and liver fibrosis in the immunoassay further validated the clinical significance of apo C-III found in our diagnostic model. By investigating the correlation between the protein peak intensities of apo C-III (9,165, 9,452 and 9,735 m/z) and their corresponding apo C-III concentration in serum, significant positive correlation result gained the validity of these three peaks as apo C-III. However, the correlation was low and the sum of the 3

isoforms peak intensities did not show great improvement in the correlation between the apo C-III concentration. Though 3 major apo C-III isoforms were found in proteomic profile, there was still a number of variants reported in other studies [157][182][183] cannot be detected in our system. This could explain the low correlation coefficient between peak intensities of apo C-III variants and apo C-III concentration measured by immunoassay.

Though apo A-I was currently included in two common non-invasive models, there were still some studies reported that it was not correlated with liver fibrosis which was accordant with our result. Lebensztejn DM *et al.* found that apo A-I was not significantly different in patients with CHB infection compared with the controls. [184] Lu *et al.* further pointed out that there was an ethnic difference in the normal range of apo A-I. The concentration of apo A-I in controls was abnormally high compared with those in foreign countries for a CHB study in China, making it inapplicable to PGA and PGAA models. [185] Another CHB study by Selimoglu MA *et al.* also showed the same result. [186] This discrepancy may be account for the differences of ethnicity and etiology of the patients as FibroTest and PGAA index were mainly used for hepatitis C infection and alcoholic liver diseases respectively. Nevertheless, a study showed that apo A-I was a biomarker for hepatitis B virus infected liver inflammation. [187] Apolipoprotein A-I presented heterogeneous change in expression level with different isoforms specific to HBV infection. Further study on apo A-I in CHB studies against other liver diseases is needed to sort out this discrepancy. No correlation between apo B and liver fibrosis could be explained by the difference in etiology. Apo B was found to be associated with HCV in plasma. A CHC study illustrated that higher apo B-

associated cholesterol was positively correlated with CHC patients after receiving anti-viral treatment. Sherudan *et al.* proposed this might due to competition between apo B-containing lipoproteins and infectious low-density HCV lipo-viral particles for hepatocytes entry via shared lipoprotein receptors. [188] On the other hand, although several studies demonstrated that decreased expression of apo C-II was found in biliary atresia-associated liver cirrhosis patients and suggested it was a novel substrate for matrix metalloproteinases [189] [190], no correlation between apo C-II and liver fibrosis was found in our study. Despite apo C-II, like apo C-III, belongs to the C class of apolipoproteins, no recent non-invasive model or study reported apo C-II was a biomarker in detection of liver fibrosis. Further investigation is needed to elucidate its significance in clinical use. Apo E has been found to be correlated with liver function in mice model and related to lipid metabolism. However, its importance and usefulness to liver fibrosis in human remains unclear and need further investigation.

3.5 Conclusion

The m/z 9,165 proteomic feature, which was one of protein markers selected in the diagnostic model, was identified to be apo C-III by mass spectrometry. Subsequently, it was confirmed by immunodepletion. Two peaks (m/z 9,452 and 9,735) were also confirmed to be isoforms of apo C-III. Concentration of apo C-III obtained by immunoassay was negatively correlated with Ishak score. The positive correlation between the depleted protein peaks and apo C-III concentration was found. Another apolipoprotein, apo A-II, was also negatively correlated with the progression of liver

fibrosis. Magnetic bead-based proteomic profiling is able to find biomarkers in the detection of liver fibrosis in patients with CHB infection.

General discussion

After the completion of human genome project (HUGO) in 2003 [191], proteomics has become another important research area in biomarker discovery. Proteomics, which is defined as the study of proteome, has been widely applied to studies of the protein expression levels, structures and protein-protein interactions. [192] With advanced technology, complex protein mixtures could be well separated and analyzed by MS technology to study the protein expression, finding biomarkers for various diseases. Protein expression profiling has been considered as a new strategy for biomarker discovery and the disease-associated/specific "fingerprinting" or "proteomic pattern" is used as a biomarker. [193]

In our study, using a magnetic beads-based platform we successfully developed an in-house proteomic profiling method that allows both quantitative analysis and micro-scale purification in parallel. Similar to the SELDI technology, a subset of proteins with definite chemical properties were retained and screened for disease-associated proteins. Disease-associated proteins were found by comparing the protein contents between disease and control groups which was regarded as comparative proteomics. Large sample size was needed so that good statistical power could be achieved for searching biomarkers. Therefore, high-throughput technology is indispensable for a large scale study. Our developed method, magnetic-beads platform allows analyzing 96 samples in one batch which greatly reduces the experimental time and enhances reproducibility. The other concern is the specificity of the biomarkers. Serum amyloid A (SAA), an acute phase protein, was found to be a biomarker for various disease such as renal cancer [89], nasopharyngeal cancer [194], hepatocelluar carcinoma. [155] The non-

specificity of SAA made it not suitable for clinical use when is used alone as a biomarker. Furthermore, it seemed that SAA responded to inflammation rather than carcinogenesis. [195] In addition, high abundance proteins in plasma/serum are the major problems in proteomic profiling. Competition occurs during sample binding and those large and high abundance proteins usually mask low abundance proteins. Transferrin, ranked as 4th of the most abundance proteins [83], was found as biomarker for certain diseases like breast [196] and ovarian cancer. [197] The usefulness of this kind of high abundance protein is usually limited. It is believed that disease-associated markers released by a very small tumor/defected area and their microenvironment should be at very low level and probably in μg/L level like prostate specific antigen (PSA). [139] [143] Diamandis pointed out that SELDI technology could only detect proteins in mg/L level which is 1000-fold higher than the concentration of known tumor markers in the circulation. For apo C-III found in our study, the concentration was 60-180 mg/L in healthy person. Though it was not as high abundance as apolipoprotein A-I and transferrin (1,000-3,600 mg/L), it was far higher abundance than those known biomarkers. [83]

In the future, it is important to develop novel high-throughput technologies to profile the low abundance proteins in the serum/plasma samples for patients with different degrees of liver fibrosis. Penetration into the deep proteome therefore becomes the main challenge in discovering low abundance biomarkers. It is well known that high abundance proteins always mask the detection of low abundance proteins, a potential source of biomarkers under assumption that only a subtle difference may occur and is reflected to the bloodstream at the early stage of the disease. The high dynamic range of

proteins can be compressed by the use of large-bead-based library of combinational peptide ligands (Equalizer beads or ProteoMiner) to enrich the low abundance proteins. Sihlbom *et al.* investigated the feasibility of the combined use of the Equalizer beads with SELDI or 2-D DIGE to improve the protein profiling compared with the depleted method. [198] They showed that better resolving efficiency and reproducibility were achieved using beads technology compared with multiple affinity depletion columns. High abundance protein such as albumin was greatly reduced and more gel spots could be resolved from the sample depleted with Equalizer beads.

Traditionally, the high abundance proteins were depleted by immunodepletion based on dye-ligands or specific antibodies. [199] However, it is subject to some limitations. It is a low throughput and time consuming process as the abundance proteins are depleted accordingly. Certain amount of low abundance proteins are co-depleted out with the removal proteins and information on these low abundance proteins may be lost. Furthermore, the large dynamic range is only reduced by removing the high abundance proteins; there is no concentration step on low abundance proteins which still limits the chance of detecting the low species. Indeed, this depletion step will further dilute the sample as larger elution volume is obtained compared with the original sample. [200] A rather new method, Equalizer bead, has been introduced to compress the high abundance proteins and enrich the low species simultaneously with no further dilution on the elution sample. Its principle is based on the use of combinatorial ligand libraries of hexapeptides bound to a chromatographic support. [201] The ligand libraries consist of 10^9-10^{12} unique peptide ligands which greatly enhance the number and variety of proteins being captured. In the past, due to the relative abundance issue, high abundance

proteins had a higher chance to be captured and analyzed than low abundance proteins. With the new technology, the high abundance proteins stop being captured when the binding sites of the specify beads are saturated.

In the future, combined use of Equalizer beads and our magnetic beads technology should allow the discovery of low abundance proteins as potential biomarkers. The serum can be first treated by Equalizer beads to enrich the low abundance proteins. The fraction is then subjected to magnetic beads protein profiling for biomarker discovery. It is believed that the relative concentration differences of the low abundance proteins between disease and control groups should remain approximately unchanged if these low abundance proteins were not saturated on the beads. [200] The main drawback of this technology is large volume of plasma/serum (e.g. 1mL) is needed for the enrichment of low abundance proteins, and the current technology can only be performed manually. If the magnetic property is added to the Equalizer beads, our developed method can be adopted so that automation of enrichment together with purification can be done on the Kingfisher 96 platform. Besides applying 2-D PAGE for protein identification, gel-free shotgun proteomics approach can be used instead. The eluted fraction can be subjected to 2-D LC for protein identification using MudPIT approach. The fraction was first digested and then separated by SCX and reversed phase columns. This approach can greatly reduce sample loss and high coverage of protein can be obtained.

Although our current magnetic beads-based profiling technology and the SELDI-TOF technology cannot provide quantitative information of the low abundance proteins, the identification of disease-specific variants may help to improve the specificity of the proteomic fingerprints for disease diagnoses. Both SELDI and magnetic bead profiling

are capable of differentiating low molecular weight protein variants having different post-translation modification(s). Some variants could be disease-associated. For example, des-Ala-fibrinopeptide A (m/z 1,466), was found to be a potential liver cancer marker [202] and fibrinogen α–chain (m/z 2,664) was a potential biomarker for oral cancer. [84] These protein variants are too small to be analyzed using traditional 2-D PAGE technique.

In our study, apo C-III variants were detected as a biomarker in detecting liver fibrosis which was undergone sialylation. Though three variants were correlated with the degrees of liver fibrosis, the hyposialylated apo C-III variant appeared as the best marker. It is well known that it is difficult to develop a glycoform-specific assay using the conventional immunoassay technology. If one wants to only quantify a specific apo C-III variant, one will need to develop glycosylation immunosorbent assay [203], MS immunoassay [204], or multiple reaction monitoring assay. [205] MRM can be done for further investigation on isoform characterization and relative quantitation of apo C-III isoforms.

In clinical point of view, our result has suggested that our diagnostic model can be used as a screening test for liver fibrosis. Blood test is cheaper and more versatile compared with other non-invasive tests such as FibroScan. As FibroScan is proved as a stable and accurate method in predicting liver fibrosis, it may be a good alternative to reduce liver biopsy. However, specialist is required to perform the FibroScan test, leading to lower turn around time and higher running cost. Therefore, in order to shorten the waiting time and running cost, blood test can be introduced as the first line screening test. Patients can first do a routine blood test in the clinics for routine check. Those

patients with blood test result falling in the grey area will be followed up with FibroScan. Finally, the highly suspected cases will be confirmed by histological examination of liver biopsies. This approach can remarkably decrease the hospital expense.

Conclusion

In conclusion, a novel magnetic beads-based proteomic profiling method was successfully developed, and proven to be a useful tool in biomarker discovery. With the developed methodology, a proteomic fingerprint composed of 2 protein markers was found in patients with CHB-associated liver fibrosis, and could be used as a diagnostic marker with good accuracy. The predictive model was treat-independent which would not be affected by anti-viral treatment. It could be considered as a screening test for detecting liver fibrosis. With our classification strategy, those patients in high risk and low risk group can be excluded from the examination of liver biopsy. Liver biopsy and hospital expense can be reduced under this strategy. Apo C-III was identified to be one of the protein markers. Subsequently, we showed that serum Apo C-III and Apo A-II concentration were decreased in the CHB patients with liver fibrosis. As both apo A-II and apo C-III were regulated by HNF-4α, it was believed that HNF-4α maybe a key factor for liver fibrosis.

Reference

1. Manning DS, Afdhal NH. Diagnosis and quantitation of fibrosis. Gastroenterology. 2008, 134, 6, 1670-81.
2. Friedman SL. Mechanisms of disease: Mechanisms of hepatic fibrosis and therapeutic implications Nat Clin Pract Gastroenterol Hepatol. 2004, 1, 2, 98-105.
3. WTO. Fact sheets hepatitis B. http://www.who.int/mediacentre/factsheets/fs204
4. Fung J, Lai CL, But D, Wong D, Cheung TK, *et al.* Prevalence of fibrosis and cirrhosis in chronic hepatitis B: implications for treatment and management. Am J Gastroenterol. 2008, 103, 6, 1421-6.
5. Kershenobich Stalnikowitz D, Weissbrod AB. Liver fibrosis and inflammation. A review. Ann Hepatol. 2003, 2, 4, 159-63.
6. Parola M, Robino G. Oxidative stress-related molecules and liver fibrosis. J Hepatol. 2001, 35, 2, 297-306.
7. Benyon RC, Iredale JP. Is liver fibrosis reversible? Gut. 2000, 46, 4, 443-6.
8. Gutierrez-Reyes G, Gutierrez-Ruiz MC, Kershenobich D. Liver fibrosis and chronic viral hepatitis. Arch Med Res. 2007, 38, 6, 644-51.
9. Shigeki T, Christopher JP , Richard A. Mechanisms of liver fibrosis. Clin Chim Acta. 2004, 364, 33-60.
10. Eng, F. J., Friedman, S. L. Fibrogenesis I. New insights into hepatic stellate cell activation: the simple becomes complex. Am. J. Physiol. Gastrointest. Liver Physiol. 2000, 279: G7-G11.
11. Mak KM, Lieber CS. Lipocytes and transitional cells in alcoholic liver disease: a morphometric study. Hepatology 1988, 8, 1027-33.

12. Masuda H, Fukumoto M, Hirayoshi K, Nagata K. Coexpression of the collagen-binding stress protein HSP47 gene and the alpha 1(I) and alpha 1(III) collagen genes in carbon tetrachloride-induced rat liver fibrosis. J Clin Invest. 1994, 94, 6, 2481-8.

13. Friedman SL. Cytokines and fibrogenesis. Semin Liver Dis. 1999, 19, 2, 129-40.

14. Leask A, Abraham DJ. The role of connective tissue growth factor, a multifunctional matricellular protein, in fibroblast biology. Biochem Cell Biol. 2003 Dec;81(6):355-63.

15. Bonner JC. Regulation of PDGF and its receptors in fibrotic diseases. Cytokine Growth Factor Rev. 2004, 15, 4, 255-73.

16. Visse R, Nagase H. Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. Circ Res. 2003, 92, 8, 827-39.

17. Iredale JP, Benyon RC, Arthur MJ, Ferris WF, Alcolado R, *et al.* Tissue inhibitor of metalloproteinase-1 messenger RNA expression is enhanced relative to interstitial collagenase messenger RNA in experimental liver injury and fibrosis. Hepatology. 1996, 24, 1, 176-84.

18. Baffy G. Kupffer cells in non-alcoholic fatty liver disease: the emerging view. J Hepatol. 2009, 51, 1, 212-23.

19. Sebastiani G, Alberti A. Non invasive fibrosis biomarkers reduce but not substitute the need for liver biopsy. World J Gastroenterol. 2006, 12, 23, 3682-94.

20. Brunt EM. Grading and staging the histopathological lesions of chronic hepatitis: the Knodell histology activity index and beyond. Hepatology. 2000, 31, 1, 241-6.

21. Knodell RG, Ishak KG, Black WC, Chen TS, Craig R, *et al.* Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. Hepatology. 1981, 1, 5, 431-5.

22. Scheuer PJ. Liver biopsy in chronic hepatitis: 1968-78. Gut. 1978, 19, 6, 554-7.

23. Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. Hepatology. 1994, 19, 6, 1513-20.

24. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, *et al.* Histological grading and staging of chronic hepatitis. J Hepatol. 1995, 22, 6, 696-9.

25. Hui AY, Chan HL, Wong VW, Liew CT, Chim AM, *et al.* Identification of chronic hepatitis B patients without significant liver fibrosis by a simple noninvasive predictive model. Am J Gastroenterol. 2005, 100, 3, 616-23.

26. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology. 1996, 24, 2, 289-93.

27. Cadranel JF, Lahmek P, Causse X, Bellaiche G, Bettan L, *et al.* Epidemiology of chronic hepatitis B infection in France: risk factors for significant fibrosis--results of a nationwide survey. Aliment Pharmacol Ther. 2007, 26, 4, 565-76.

28. Gressner OA, Weiskirchen R, Gressner AM. Biomarkers of liver fibrosis: clinical translation of molecular pathogenesis or based on liver-dependent malfunction tests. Clin Chim Acta. 2007, 381, 2, 107-13.

29. Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, *et al.* Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol. 2002, 97, 10, 2614-8.

30. Colloredo G, Guido M, Sonzogni A, Leandro G. Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: the smaller the sample, the milder the disease. J Hepatol. 2003, 39, 2, 239-44.

31. Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. Nature. 2003, 425, 6961, 905.

32. Petricoin EF, Belluco C, Araujo RP, Liotta LA. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. Nat Rev Cancer. 2006, 6, 12, 961-7.

33. Gutman S, Kessler LG. The US Food and Drug Administration perspective on cancer biomarker development. Nat Rev Cancer. 2006, 6, 7, 565-71.

34. Grigorescu M. Noninvasive biochemical markers of liver fibrosis. J Gastrointestin Liver Dis. 2006, 15, 2, 149-59.

35. Sporea I, Popescu A, Sirli R. Why, who and how should perform liver biopsy in chronic liver diseases. World J Gastroenterol. 2008, 14, 21, 3396-402.

36. Gebo KA, Herlong HF, Torbenson MS, Jenckes MW, Chander G, *et al.* Role of liver biopsy in management of chronic hepatitis C: a systematic review. Hepatology. 2002, 36(5 Suppl 1), S161-72.

37. Shaheen AA, Myers RP. Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis C-related fibrosis: a systematic review. Hepatology. 2007, 46, 3, 912-21.

38. Murawaki Y, Ikuta Y, Nishimura Y, Koda M, Kawasaki H. Serum markers for connective tissue turnover in patients with chronic hepatitis B and chronic hepatitis C: a comparative analysis. J Hepatol. 1995, 23, 2, 145-52.

39. Becker L, Salameh W, Sferruzza A, Zhang K, Ng Chen R, *et al.* Validation of hepascore, compared with simple indices of fibrosis, in patients with chronic hepatitis C virus infection in United States. Clin Gastroenterol Hepatol. 2009, 7, 6, 696-701.

40. Mehta P, Ploutz-Snyder R, Nandi J, Rawlins SR, Sanderson SO, *et al.* Diagnostic accuracy of serum hyaluronic acid, FIBROSpect II, and YKL-40 for discriminating fibrosis stages in chronic hepatitis C. Am J Gastroenterol. 2008, 103, 4, 928-36.

41. Körner T, Kropf J, Gressner AM. Serum laminin and hyaluronan in liver cirrhosis: markers of progression with high prognostic value. J Hepatol. 1996, 25, 5, 684-8.

42. Oberti F, Valsesia E, Pilette C, Rousselet MC, Bedossa P, *et al.* Noninvasive diagnosis of hepatic fibrosis or cirrhosis. Gastroenterology. 1997, 113, 5, 1609-16.

43. Johansen JS, Christoffersen P, Møller S, Price PA, Henriksen JH, *et al.* Serum YKL-40 is increased in patients with hepatic fibrosis. J Hepatol. 2000, 32, 6, 911-20.

44. Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, *et al.* A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology. 2003, 38, 2, 518-26.

45. Shaheen AA, Myers RP. Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis C-related fibrosis: a systematic review. Hepatology 2007, 46, 3, 912-21

46. Lin CS, Chang CS, Yang SS, Yeh HZ, Lin CW. Retrospective evaluation of serum markers APRI and AST/ALT for assessing liver fibrosis and cirrhosis in chronic

hepatitis B and C patients with hepatocellular carcinoma. Intern Med. 2008, 47, 7, 569-75.

47. Forns X, Ampurdanès S, Llovet JM, Aponte J, Quintó L, *et al.* Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. Hepatology. 2002, 36(4 Pt 1), 986-92.

48. Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, *et al.* Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. Lancet. 2001, 357, 9262, 1069-75.

49. Naveau S, Gaudé G, Asnacios A, Agostini H, Abella A, *et al.* Diagnostic and prognostic values of noninvasive biomarkers of fibrosis in patients with alcoholic liver disease. Hepatology. 2009, 49, 1, 97-105.

50. Kim BK, Kim SA, Park YN, Cheong JY, Kim HS, *et al.* Noninvasive models to predict liver cirrhosis in patients with chronic hepatitis B. Liver Int. 2007, 27, 7, 969-76.

51. Mohamadnejad M, Montazeri G, Fazlollahi A, Zamani F, Nasiri J, *et al.* Noninvasive markers of liver fibrosis and inflammation in chronic hepatitis B-virus related liver disease. Am J Gastroenterol. 2006, 101, 11, 2537-45.

52. Zeng MD, Lu LG, Mao YM, Qiu DK, Li JQ, *et al.* Prediction of significant fibrosis in HBeAg-positive patients with chronic hepatitis B by a noninvasive model. Hepatology. 2005, 42, 6, 1437-45.

53. Pinzani M, Vizzutti F, Arena U, Marra F. Technology Insight: noninvasive assessment of liver fibrosis by biochemical scores and elastography. Nat Clin Pract Gastroenterol Hepatol. 2008, 5, 2, 95-106.

54. de Lédinghen V, Vergniol J. Transient elastography (FibroScan). Gastroenterol Clin Biol. 2008, 32(6 Suppl 1), 58-67.

55. Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. J Hepatol. 2008, 48, 5, 835-47.

56. Friedrich-Rust M, Ong MF, Martens S, Sarrazin C, Bojunga J, *et al.* Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology. 2008, 134, 4, 960-74.

57. Fontaine H, Petitprez K, Roudot-Thoraval F, Trinchet JC; Association francaise pour l'etude du foie, *et al.* Guidelines for the diagnosis of uncomplicated cirrhosis. Gastroenterol Clin Biol. 2007, 31, 5, 504-9.

58. Sandrin L, Fourquet B, Hasquenoph JM, Yon S, Fournier C, *et al.* Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol. 2003, 29, 12, 1705-13.

59. Chan HL, Wong GL, Choi PC, Chan AW, Chim AM, *et al.* Alanine aminotransferase-based algorithms of liver stiffness measurement by transient elastography (Fibroscan) for liver fibrosis in chronic hepatitis B. J Viral Hepat. 2009, 16, 1, 36-44.

60. Grant SG, Blackstock WP. Proteomics in neuroscience: from protein to network. J Neurosci. 2001, 21, 21, 8315-8.

61. Banks RE, Dunn MJ, Hochstrasser DF, Sanchez JC, Blackstock W, *et al.* Proteomics: new perspectives, new biomedical opportunities. Lancet. 2000, 356, 9243, 1749-56.

62. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science. 2001, 294, 5543, 853-8.

63. Jose BC, Jose LL. A better understanding of molecular mechanisms underlying human disease. Proteomics Clin. Appl. 2007, 1, 983–1003.

64. Elzbieta P, Stella BSi, Malgorzata C, Richard IS. Proteomics in human cancer research Prot. Clin. Appl. 2007, 1, 1, 4-17.

65. Lewis TS, Hunt JB, Aveline LD, Jonscher KR, Louie DF, *et al.* Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. Mol Cell. 2000, 6, 6, 1343-54.

66. Shin J, Lee W, Lee W. Structural proteomics by NMR spectroscopy. Expert Rev Proteomics. 2008, 5, 4, 589-601

67. Xue H, Lu B, Lai M. The cancer secretome: a reservoir of biomarkers. J Transl Med. 2008, 6, 52.

68. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, *et al.* HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics. 2005, 5, 13, 3262-77

69. Sven B, Uta C, Georg MF, Jan L, Alexander L, *et al.* Standardized Approach to Proteome Profiling of Human Serum Based on Magnetic Bead Separation and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. Clin Chem 2005, 51, 973-980.

70. Marshall J, Kupchak P, Zhu W, Yantha J, Vrees T, *et al.* Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. J Proteome Res. 2003, 2, 4, 361-72.

71. Ayache S, Panelli M, Marincola FM, Stroncek DF. Effects of storage time and exogenous protease inhibitors on plasma protein levels. Am J Clin Pathol. 2006, 126, 2, 174-84

72. Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik. 1975, 26, 3, 231-43.

73. Iwadate Y. Clinical proteomics in cancer research-promises and limitations of current two-dimensional gel electrophoresis. Curr Med Chem. 2008, 15, 23, 2393-400.

74. Berggren K, Chernokalskaya E, Steinberg TH, Kemper C, Lopez MF, *et al.* Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis. 2000, 12, 2509-21.

75. Wang J, Jiang D, Zhang H, Lv S, Rao H, *et al.* Proteome responses to stable hepatitis B virus transfection and following interferon alpha treatment in human liver cell line HepG2. Proteomics. 2009, 9, 6, 1672-82

76. Spano D, Cimmino F, Capasso M, D'Angelo F, Zambrano N, *et al.* Changes of the hepatic proteome in hepatitis B-infected mouse model at early stages of fibrosis. J Proteome Res. 2008, 7, 7, 2642-53.

77. Wildgruber R, Harder A, Obermaier C, Boguth G, Weiss W, *et al.* Towards higher resolution: two-dimensional electrophoresis of Saccharomyces cerevisiae proteins using overlapping narrow immobilized pH gradients. Electrophoresis. 2000, 21, 13, 2610-6.

78. Zhou G, Li H, DeCamp D, Chen S, Shu H, *et al.* 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. Mol Cell Proteomics. 2002, 1, 2, 117-24.

79. Unlü M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis. 1997, 18, 11, 2071-7.

80. Han X, Aslanian A, Yates JR 3rd. Mass spectrometry for proteomics. Curr Opin Chem Biol. 2008, 12, 5, 483-90.

81. Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. Rapid Commun. Mass Spectrom. 1993, 7, 576-580.

82. Righetti PG, Castagna A, Antonioli P, Boschetti E. Prefractionation techniques in proteome analysis: the mining tools of the third millennium. Electrophoresis. 2005, 26, 2, 297-319.

83. Hortin, G.L. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. Clin. Chem. 2006, 7, 1223-37.

84. Cheng AJ, Chen LC, Chien K Y, Chen Y J, Chang J T, *et al.* Oral cancer plasma tumor marker identified with bead-based affinity-fractionated proteomic technology. Clin. Chem. 2005, 12, 2236-44.

85. Honda K, Hayashida Y, Umaki T, Okusaka T, Kosuge T, *et al.* Possible detection of pancreatic cancer by plasma protein profiling. Cancer Res. 2005, 22, 10613-22.

86. Kawada N. Cancer serum proteomics in gastroenterology. Gastroenterology 2006, 6, 1917-19.

87. Huang JT, Wang L, Prabakaran S, Wengenroth M, Lockstone HE, *et al.* Independent protein-profiling studies show a decrease in apolipoprotein A1 levels in schizophrenia CSF, brain and peripheral tissues. Mol. Psychiatry 2008, 12, 1118-28.

88. Zhang X, Jin M, Wu H, Nadasdy T, Nadasdy G, *et al.* Biomarkers of lupus nephritis determined by serial urine proteomics. Kidney Int. 2008, 6, 799-807.

89. Tolson J, Bogumil R, Brunst E, Beck H, Elsner R, *et al.* Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid alpha in renal cancer patients. Lab. Invest. 2004, 7, 845-56.

90. Jayanthi S, Buie S, Moore S, Herning RI, Better W, *et al.* Heavy marijuana users show increased serum apolipoprotein C-III levels: evidence from proteomic analyses. Mol Psychiatry. 2008 May 13.

91. Poon TC, Sung JJ, Chow SM, Ng EK, Yu AC, *et al.* Diagnosis of gastric cancer by serum proteomic fingerprinting. Gastroenterology 2006, 6, 1858-64.

92. Diamandis EP..Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol. Cell. Proteomics 2004, 4, 367-78.

93. Gundry RL, White MY, Nogee J, Tchernyshyov I, Van Eyk JE. Assessment of albumin removal from an immunoaffinity spin column: critical implications for proteomic examination of the albuminome and albumin-depleted samples. Proteomics. 2009, 9, 7, 2021-8.

94. Ward DG, Cheng Y, N'Kontchou G, Thar TT, Barget N, *et al.* Changes in the serum proteome associated with the development of hepatocellular carcinoma in hepatitis C-related cirrhosis. Br J Cancer. 2006, 94, 2, 287-92.

95. Göbel T, Vorderwülbecke S, Hauck K, Fey H, Häussinger D, *et al.* New multi protein patterns differentiate liver fibrosis stages and hepatocellular carcinoma in chronic hepatitis C serum samples. World J Gastroenterol. 2006, 12, 47, 7604-12.

96. Tucholska M, Bowden P, Jacks K, Zhu P, Furesz S, *et al.* Human serum proteins fractionated by preparative partition chromatography prior to LC-ESI-MS/MS. J Proteome Res. 2009, 8, 3, 1143-55.

97. Yang X, Clifton J, Huang F, Kovac S, Hixson DC, *et al.* Proteomic analysis for process development and control of therapeutic protein separation from human plasma. Electrophoresis. 2009, 30, 7, 1185-93.

98. Gao M, Deng C, Yu W, Zhang Y, Yang P, Zhang X. Large scale depletion of the high-abundance proteins and analysis of middle- and low-abundance proteins in human liver proteome by multidimensional liquid chromatography. Proteomics. 2008, 8, 5, 939-47.

99. Cagney G, Park S, Chung C, Tong B, O'Dushlaine C, *et al.* Human tissue profiling with multidimensional protein identification technology. J Proteome Res. 2005, 4, 5, 1757-67.

100. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001, 19, 3, 242-7.

101. Zhang J, Xu X, Gao M, Yang P, Zhang X. Comparison of 2-D LC and 3-D LC with post- and pre-tryptic-digestion SEC fractionation for proteome analysis of normal human liver tissue. Proteomics. 2007, 7, 4, 500-12.

102. Wiesner J, Premsler T, Sickmann A. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics. 2008, 8, 21, 4466-83.

103. Whitelegge JP, Zabrouskov V, Halgand F, Souda P, Bassilian S, *et al.* Protein-Sequence Polymorphisms and Post-translational Modifications in Proteins from Human Saliva using Top-Down Fourier-transform Ion Cyclotron Resonance Mass Spectrometry. Int J Mass Spectrom. 2007, 268, 2-3, 190-197.

104. Parks BA, Jiang L, Thomas PM, Wenger CD, Roth MJ, *et al.* Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers. Anal Chem. 2007, 79, 21, 7984-91.

105. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002, 1, 5, 376-86.

106. Chen N, Sun W, Deng X, Hao Y, Chen X, *et al.* Quantitative proteome analysis of HCC cell lines with different metastatic potentials by SILAC. Proteomics. 2008, 8, 23-24, 5108-18.

107. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999, 17, 10, 994-9.

108. Ahrends R, Pieper S, Kühn A, Weisshoff H, Hamester M, *et al.* A metal-coded affinity tag approach to quantitative proteomics. Mol Cell Proteomics. 2007, 6, 11, 1907-16.

109. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. 2007, 389, 4, 1017-31.

110. Schmidt C, Urlaub H. iTRAQ-Labeling of In-Gel Digested Proteins for Relative Quantification. Methods Mol Biol. 2009;564:207-26.

111. Kitteringham NR, Jenkins RE, Lane CS, Elliott VL, Park BK. Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. J Chromatogr B Analyt Technol Biomed Life Sci. 2009 May 1;877(13):1229-39.

112. Latterich M, Abramovitz M, Leyland-Jones B. Proteomics: new technologies and clinical applications. Eur J Cancer. 2008, 44, 18, 2737-41.

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

114. Richard K, Chris B, Lucy R, Balwir MB, Pamela B, *et al.* Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. Rapid Commun. Mass Spectrom. 2008, 22, 3255–60.

115. Unwin RD, Griffiths JR, Whetton AD. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple

reaction monitoring-initiated detection and sequencing (MIDAS). Nat Protoc. 2009;4(6):870-7.

116. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics. 2006, 5, 4, 573-88.

117. Zhu XD, Zhang WH, Li CL, Xu Y, Liang WJ, *et al.* New serum biomarkers for detection of HBV-induced liver cirrhosis using SELDI protein chip technology. World J Gastroenterol 2004, 10, 16, 2327-9.

118. Cui J, Kang X, Dai Z, Huang C, Zhou H, *et al.* Prediction of chronic hepatitis B, liver cirrhosis and hepatocellular carcinoma by SELDI-based serum decision tree classification. J Cancer Res Clin Oncol 2007, 133, 825-834

119. Poon TC, Hui AY, Chan HL, Ang IL, Chow SM, *et al.* Prediction of liver fibrosis and cirrhosis in chronic hepatitis B infection by serum proteomic fingerprinting: a pilot study. Clin Chem. 2005, 51, 2, 328-35

120. Sundsten T, Eberhardson M, Goransson M., Bergsten P., The use of proteomics in identifying differentially expressed serum proteins in humans with type 2 diabetes. Proteome Sci. 2006, 4, 22.

121. Pang RT, Poon TC, Chan KC, Lee NL, Chiu RW, *et al.*, Serum proteomic fingerprints of adult patients with severe acute respiratory syndrome. Clin. Chem. 2006, 3, 421-429.

122. Poon TC, Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. Expert Rev. Proteomics 2007, 1, 51-65.

123. Alterovitz G, Aivado M, Spentzos D, Libermann TA, Ramoni M, *et al.* Analysis and robot pipelined automation for SELDI-TOF mass spectrometry. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2004, 4, 3068-3071.

124. Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, *et al.*, Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clin. Chem. 2005, 51, 102-112.

125. Villanueva J, Lawlor K, Toledo-Crow R, Tempst P, Automated serum peptide profiling. Nat. Protoc. 2006, 2, 880-891.

126. Sorrell MF, Belongia EA, Costa J, Gareen IF, Grem JL,*et al.* National Institutes of Health consensus development conference statement: management of hepatitis B. Hepatology. 2009, 49(5 Suppl), S4-12.

127. Lok AS, McMahon BJ. Chronic hepatitis B. Hepatology. 2007, 45, 2, 507-39.

128. Michael FS, Edward A B, Jose C, Ilana FG, Jean LG, *et al.* National Institutes of Health consensus development conference statement: Management of hepatitis B. hepatology, 2009, 49, 5, S96-102.

129. Halfon P, Munteanu M, Poynard T. FibroTest-ActiTest as a non-invasive marker of liver fibrosis. Gastroenterol Clin Biol. 2008, 32(6 Suppl 1, 22-39.

130. Adams LA, Bulsara M, Rossi E, DeBoer B, Speers D, *et al.* Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. Clin chem. 2005, 51, 10, 1867-73.

131. van Winden AW, Gast MC, Beijnen JH, Rutgers EJ, Grobbee DE, *et al.* Validation of previously identified serum biomarkers for breast cancer with SELDI-TOF MS: a case control study. BMC Med Genomics. 2009, 2, 4.

132. Wu FX, Wang Q, Zhang ZM, Huang S, Yuan WP, *et al.* Identifying serological biomarkers of hepatocellular carcinoma using surface-enhanced laser desorption/ionization-time-of-flight mass spectroscopy. Cancer Lett. 2009, 279, 2, 163-70.

133. Hammoud ZT, Dobrolecki L, Kesler KA, Rahmani E, Rieger K, *et al.* Diagnosis of esophageal adenocarcinoma by serum proteomic pattern. Ann Thorac Surg. 2007, 84, 2, 384-92.

134. White IR, Patel K, Symonds WT, Dev A, Griffin P, *et al.* Serum proteomic analysis focused on fibrosis in patients with hepatitis C virus infection. J Transl Med. 2007, 5, 33.

135. Pang RT, Poon TC, Chan KC, Lee NL, Chiu RW, *et al.* Serum amyloid A is not useful in the diagnosis of severe acute respiratory syndrome. Clin Chem 2006, 52, 1202-4.

136. Janssen HL, van Zonneveld M, Senturk H, Zeuzem S, Akarca US, *et al.* Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. Lancet. 2005, 365, 9454, 123-9.

137. Lau GK, Piratvisuth T, Luo KX, Marcellin P, Thongsawat S, *et al.* Peginterferon Alfa-2a, lamivudine, and the combination for HBeAg-positive chronic hepatitis B. N Engl J Med. 2005, 352, 26, 2682-95.

138. Marcellin P, Lau GK, Bonino F, Farci P, Hadziyannis S, *et al.* Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. N Engl J Med. 2004, 351, 12, 1206-17.

139. Diamandis EP. Point: Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? Clin Chem. 2003, 49, 8, 1272-5.

140. Madan A, El-Ferzli G, Carlson SM, Whitin JC, Schilling J, *et al.* A potential biomarker in the cord blood of preterm infants who develop retinopathy of prematurity. Pediatr Res. 2007, 61, 2, 215-21.

141. O'Gorman D, Howard JC, Varallo VM, Cadieux P, Bowley E, *et al.* Identification of protein biomarkers in Dupuytren's contracture using surface enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI-TOF-MS). Clin Invest Med. 2006, 29, 3, 136-45.

142. Poon TC, Sung JJ, Chow SM, Ng EK, Yu AC, *et al.* Diagnosis of gastric cancer by serum proteomic fingerprinting. Gastroenterology. 2006, 130, 6, 1858-64.

143. Diamandis EP. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst. 2004, 96, 5, 353-6.

144. Pang RT, Poon TC, Chan KC, Lee NL, Chiu RW, *et al.* Serum proteomic fingerprints of adult patients with severe acute respiratory syndrome. Clin chem. 2006, 52, 3, 421-9

145. Rossi E, Adams L, Prins A, Bulsara M, de Boer B, *et al.* Validation of the FibroTest biochemical markers score in assessing liver fibrosis in hepatitis C patients. Clin Chem. 2003, 49, 450-4

146. Hongbo L, Xiaohui L, Hong K, Wei W, Yong Z. Assessing routine and serum markers of liver fibrosis in CHB patients using parallel and serial interpretation. Clin Biochem. 2007, 40, 8, 562-6.

147. Poynard T, Bedossa P. Age and platelet count: a simple index for predicting the presence of histological lesions in patients with antibodies to hepatitis C virus. METAVIR and CLINIVIR Cooperative Study Groups. J Viral Hepatol 1997, 4, 199-208.

148. Kim BK, Kim SA, Park YN, Cheong JY, Kim HS, *et al.* Noninvasive models to predict liver cirrhosis in patients with chronic hepatitis B. Liver Int. 2007, 27, 7, 969-76.

149. Poynard T, Morra R, Halfon P, Castera L, Ratziu V, *et al.* Meta-analyses of FibroTest diagnostic value in chronic liver disease. BMC Gastroenterol. 2007, 7, 40.

150. Poynard T, Ngo Y, Marcellin P, Hadziyannis S, Ratziu V, *et al.* Impact of adefovir dipivoxil on liver fibrosis and activity assessed with biochemical markers (FibroTest-ActiTest) in patients infected by hepatitis B virus. [J Viral Hepat. 2009, 16, 3, 203-13.

151. Poynard T, Zoulim F, Ratziu V, Degos F, Imbert-Bismut F, *et al.* Longitudinal assessment of histology surrogate markers (FibroTest-ActiTest) during lamivudine therapy in patients with chronic hepatitis B infection. Am J Gastroenterol. 2005, 100, 9, 1970-80.

152. Wu C, Wang Z, Liu L, Zhao P, Wang W, *et al.* Surface enhanced laser desorption/ionization profiling: New diagnostic method of HBV-related hepatocellular carcinoma. J Gastroenterol Hepatol. 2009, 24, 1, 55-62.

153. Liu X, Li L, Zhang G, Sheng G, Xu W. An approach to the characterization of serum low-molecular weight proteins/peptides in liver injury using SELDI-TOF MS and factor analysis. Clin Biochem. 2007, 40, 16-17, 1266-71.

154. Kanmura S, Uto H, Kusumoto K, Ishida Y, Hasuike S, *et al.* Early diagnostic potential for hepatocellular carcinoma using the SELDI ProteinChip system. Hepatology, 45, 4, 948-956

155. He QY, Zhu R, Lei T, Ng MY, Luk JM, *et al.* Toward the proteomic identification of biomarkers for the prediction of HBV related hepatocellular carcinoma. J Cell Biochem. 2008, 103, 3, 740-52.

156. Ward DG, Cheng Y, N'Kontchou G, Thar TT, Barget N, *et al.* Preclinical and post-treatment changes in the HCC-associated serum proteome. Br J Cancer. 2006, 95, 10, 1379-83.

157. Bondarenko PV, Cockrill SL, Watkins LK, Cruzado ID, Macfarlane RD, *et al.* Mass spectral study of polymorphism of the apolipoproteins of very low density lipoprotein. J Lipid Res. 1999, 40, 3, 543-55.

158. Millar JS. The sialylation of plasma lipoproteins. Atherosclerosis. 2001, 154, 1, 1-13.

159. Harvey SB, Zhang Y, Wilson-Grady J, Monkkonen T, Nelsestuen GL, *et al.* O-glycoside biomarker of apolipoprotein C3: responsiveness to obesity, bariatric surgery, and therapy with metformin, to chronic or severe liver disease and to mortality in severe sepsis and graft vs host disease. J Proteome Res. 2009, 8, 2, 603-12.

160. Sonia M, Dorothee M, Stephane R, Stephanie B,Jean-P C, *et al.* Identification of apolipoprotein C-III as a potential plasmatic biomarker associated with the resolution of hepatitis C virus infection. Proteomics Clin. Appl. 2008, 2, 751–61

161. Mauger JF, Couture P, Bergeron N, Lamarche B. Apolipoprotein C-III isoforms: kinetics and relative implication in lipid metabolism. J Lipid Res. 2006, 47, 6, 1212-8.

162. Bobik A. Apolipoprotein CIII and atherosclerosis: beyond effects on lipid metabolism. Circulation. 2008, 118, 7, 702-4

163. Gangabadage CS, Zdunek J, Tessari M, Nilsson S, Olivecrona G, *et al.* Structure and dynamics of human apolipoprotein CIII. J Biol Chem. 2008, 283, 25, 17416-27.

164. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, *et al.* A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. Science. 2008, 322, 5908, 1702–5.

165. Liu X, Hu H, Yin JQ. Therapeutic strategies against TGF-beta signaling pathway in hepatic fibrosis. Liver Int. 2006, 26, 1, 8-22.

166. Lucas Sd S, López-Alcorocho JM, Bartolomé J, Carreño V. Nitric oxide and TGF-beta1 inhibit HNF-4alpha function in HEPG2 cells. Biochem Biophys Res Commun. 2004, 321, 3, 688-94.

167. Sladek FM, Zhong WM, Lai E, Darnell JE Jr. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes Dev. 1990, 4(12B), 2353-65.

168. Watt AJ, Garrison WD, Duncan SA. HNF4: a central regulator of hepatocyte differentiation and function. Hepatology. 2003, 37, 6, 1249-53.

169. Lavrentiadou SN, Hadzopoulou-Cladaras M, Kardassis D, Zannis VI. Binding specificity and modulation of the human ApoCIII promoter activity by heterodimers of ligand-dependent nuclear receptors. Biochemistry, 1999, 38, 964–975.

170. Taylor DG, Haubenwallner S, Leff T. Characterization of a dominant negative mutant form of the HNF-4 orphan receptor. Nucleic Acids Res. 1996, 24, 2930–2935.

171. Ogami K, Hadzopoulou-Cladaras M, Cladaras C, Zannis VI. Promoter elements and factors required for hepatic and intestinal transcription of the human ApoCIII gene. J Biol Chem. 1990, 265, 17, 9808-15.

172. Chien KL, Chen MF, Hsu HC, Su TC, Chang WT, *et al.* Genetic association study of APOA1/C3/A4/A5 gene cluster and haplotypes on triglyceride and HDL cholesterol in a community-based population. Clin Chim Acta. 2008, 388, 1-2, 78-83.

173. O'Kane MJ, Lynch PL, Callender ME, Trimble ER. Abnormalities of serum apo A1 containing lipoprotein particles in patients with primary biliary cirrhosis. Atherosclerosis. 1997, 131, 2, 203-10.

174. Myers RP, Benhamou Y, Imbert-Bismut F, Thibault V, Bochet M, *et al.*Serum biochemical markers accurately predict liver fibrosis in HIV and hepatitis C virus co-infected patients. AIDS. 2003, 17, 5, 721-5.

175. Poynard T, Imbert-Bismut F, Munteanu M, Messous D, Myers RP, *et al.* Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV

FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. Comparative Hepatology 2004, 3, 8

176. Naveau S, Poynard T, Benattar C, Bedossa P, Chaput JC. Alpha-2-macroglobulin and hepatic fibrosis. Diagnostic interest. Dig Dis Sci 1994, 39, 2426–32.

177. Petit JM, Benichou M, Duvillard L, Jooste V, Bour JB, *et al.* Hepatitis C virus-associated hypobetalipoproteinemia is correlated with plasma viral load, steatosis, and liver fibrosis. Am J Gastroenterol. 2003, 98, 5, 1150-4.

178. Ferré N, Martínez-Clemente M, López-Parra M, González-Périz A, Horrillo R, *et al.* Increased susceptibility to exacerbated liver injury in hypercholesterolemic ApoE-deficient mice: potential involvement of oxysterols. Am J Physiol Gastrointest Liver Physiol. 2009, 296, 3, G553-62.

179. Pandurò A, Lin-Lee YC, Chan L, Shafritz DA. Transcriptional and posttranscriptional regulation of apolipoprotein E, A-I, and A-II gene expression in normal rat liver and during several pathophysiologic states. Biochemistry. 1990, 29, 36, 8430-5.

180. Shih DQ, Dansky HM, Fleisher M, Assmann G, Fajans SS, *et al.* Genotype/phenotype relationships in HNF-4alpha/MODY1: haploinsufficiency is associated with reduced apolipoprotein (AII), apolipoprotein (CIII), lipoprotein(a), and triglyceride levels. Diabetes. 2000, 49, 5, 832-7.

181. Cardot P, Chambaz J, Cladaras C, Zannis VI. Regulation of the human ApoA-II gene by the synergistic action of factors binding to the proximal and distal regulatory elements. J Biol Chem.1991, 266, 36, 24460-70.

182. von Eckardstein A, Holz H, Sandkamp M, Weng W, Funke H, Assmann G. Apolipoprotein C-III(Lys58----Glu). Identification of an apolipoprotein C-III variant in a family with hyperalphalipoproteinemia. J Clin Invest. 1991, 87, 5, 1724-31.

183. Ferrell RE, Kamboh MI, Majumder PP, Valdez R, Weiss KM. Genetic studies of human apolipoproteins. XIII. Quantitative polymorphism of apolipoprotein C-III in the Mayans of the Yucatán Peninsula. Hum Hered. 1990, 40, 3, 127-35.

184. Lebensztejn DM, Skiba E, Tobolczyk J, Sobaniec-Lotowska ME, Kaczmarski M. Diagnostic accuracy of serum biochemical fibrosis markers in children with chronic hepatitis B evaluated by receiver operating characteristics analysis. World J Gastroenterol. 2005,11, 45, 7192-6.

185. Lu LG, Zeng MD, Mao YM, Li JQ, Qiu DK, *et al.* Relationship between clinical and pathologic findings in patients with chronic liver diseases. World J Gastroenterol. 2003,9, 12, 2796-800.

186. Selimoglu MA, Yagcl RV, Yüce G. Low plasma apolipoprotein A-I level is not a reliable marker of fibrosis in children with chronic hepatitis B. World J Gastroenterol. 2004, 10, 19, 2864-6.

187. He QY, Lau GK, Zhou Y, Yuen ST, Lin MC, *et al.* Serum biomarkers of hepatitis B virus infected liver inflammation: a proteomic study. Proteomics. 2003, 3, 5, 666-74.

188. Sheridan DA, Price DA, Schmid ML, Toms GL, Donaldson P, *et al.* Apolipoprotein B-associated cholesterol is a determinant of treatment outcome in

patients with chronic hepatitis C virus infection receiving anti-viral agents interferon-alpha and ribavirin. Aliment Pharmacol Ther. 2009, 29, 12, 1282-90.

189. Lee CW, Lin MY, Lee Wen-Chien, Chou MH, Hsieh CS, *et al.* Characterization of plasma proteome in biliary atresia. Clin Chim acta 2007, 375, 104-9.

190. Andersson Y, Majd Z, Lefebvre AM, Martin G, Sechkin AV, *et al.* Developmental and pharmacological regulation of apolipoprotein C-II gene expression. Comparison with apo C-I and apo C-III gene regulation. Arterioscler Thromb Vasc Biol. 1999, 19, 1, 115-21.

191. Gyorgy MV. Proteomics principles and challenges. Pure Appl. Chem., 2004, 76, 4, 829–837.

192. Poon TCW, Philip JJ. Proteome analysis and its impact on the discovery of serological tumor markers. Clinica Chimica Acta. 2001, 313, 231-239.

193. Hortin GL. Can mass spectrometric protein profiling meet desired standards of clinical laboratory practice? Clin Chem., 2005, 51,1, 3-5

194. Cho WC, Yip TT, Yip C, Yip V, Thulasiraman V, *et al.* Identification of serum amyloid a protein as a potentially useful biomarker to monitor relapse of nasopharyngeal cancer by serum proteomic profiling. Clin Cancer Res. 2004, 10, 43-52.

195. Zakynthinos E, Pappa N. Inflammatory biomarkers in coronary artery disease. J Cardiol. 2009, 53, 3, 317-33.

196. Gonçalves A, Esterni B, Bertucci F, Sauvan R, Chabannon C, *et al.* Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. Oncogene. 2006, 25, 7, 981-9.

197. Kozak KR, Su F, Whitelegge JP, Faull K, Reddy S, Farias-Eisner R. Characterization of serum biomarkers for detection of early stage ovarian cancer. Proteomics. 2005, 5, 17, 4589-96.

198. Sihlbom C, Kanmert I, Bahr H, Davidsson P. Evaluation of the combination of bead technology with SELDI-TOF-MS and 2-D DIGE for detection of plasma proteins. J of Proteome Res. 2008, 7, 4191–98.

199. Björhall K, Miliotis T, Davidsson P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics. 2005, 5, 1, 307–17

200. Righetti PG, Boschetti E, Lomas L, Citterio A. Protein Equalizer Technology: the quest for a "democratic proteome". Proteomics. 2006, 6, 3980–92.

201. Boschetti E, Righetti PG. The art of observing rare protein species in proteomes with peptide ligand libraries. Proteomics. 2009, 9, 1492–1510.

202. Orvisky E, Drake SK, Martin BM, Abdel-Hamid M, Ressom HW, *et al.* Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma. Proteomics. 2006, 6, 9, 2895-902.

203. Poon TC, Mok TS, Chan AT, Chan CM, Leong V, *et al.* Quantification and utility of monosialylated alpha-fetoprotein in the diagnosis of hepatocellular carcinoma with nondiagnostic serum total alpha-fetoprotein. Clin Chem. 2002, 48, 7, 1021-7.

204. Ueda M, Misumi Y, Mizuguchi M, Nakamura M, Yamashita T, *et al.* SELDI-TOF mass spectrometry evaluation of variant transthyretins for diagnosis and pathogenesis of familial amyloidotic polyneuropathy. Clin Chem. 2009, 55, 6, 1223-7.

205. Nakagawa K, Oak JH, Higuchi O, Tsuzuki T, Oikawa S, *et al.* Ion-trap tandem mass spectrometric analysis of Amadori-glycated phosphatidylethanolamine in human plasma with or without diabetes. J Lipid Res. 2005, 46, 11, 2514-24.

Original Data

The original raw data obtained from this investigation will be available on request.