Protein Profiling for Hepatocellular Carcinoma Biomarker Discovery in West African Subjects

Submitted for the Degree of Doctor of Philosophy

by

Haddy Kalatou Sulayman Fye

Wolfson College, Oxford University

Hilary Term 2013

Table of Contents

Abstract

Background: Hepatocellular Carcinoma (HCC) is the third most common cause of cancer related death worldwide and is often diagnosed by measuring serum Alpha-fetoprotein (AFP); a stand-alone biomarker with limited diagnostic proficiency. To compensate for this, AFP is commonly used in conjunction with high performance imaging and radiological methods. However, as the burden of HCC is predominantly in the developing world where such technologies are not readily available, it is imperative that efforts are made to pursue the discovery of novel, high performance, easy to measure and robust biomarkers. With the aim of improving on the diagnostic ability of AFP, our project focuses on the study of plasma proteins as identified by Mass Spectrometry (MS) in order to investigate differences seen in the respective proteomes of controls and subjects with liver cirrhosis (LC) and HCC.

Methods: Matrix Assisted Laser Desorption Ionization Time-of-Flight MS (MALDI-TOF MS) was first attempted on weak cation exchange (WCX) fractionated plasma in a pilot selection of forty subjects. On the main case-control group, quantitative MS analysis using liquid chromatography electro spray ionization quadrupole time-of-flight (LC-ESI Q-TOF) was conducted on 339 subjects using a pooled expression profiling approach. Enzyme-linked immunosorbent assays (ELISA) and 1 and 2Dimentional electrophoresis methods were performed to validate and detail candidate protein levels and modification patters in individual and pooled subjects. The human plasma used for the MS based protein discovery experiments was collected as part of a five year Liver Cancer Case-control Study (Gambia, West Africa). A smaller set of samples from subjects who formed a spectrum of non-liver disease controls, LC and HCC were obtained from the Jos University Teaching Hospital (JUTH) in Nigeria and ELISA and gel electrophoresis assays conducted on them to confirm the trends and differences seen in the Gambian subject set.

Results: Bioinformatic evaluation of MALDI-TOF data highlighted peak masses 2444m/z, 2583m/z and 2559m/z to have high diagnostic abilities based on area under curve (AUC) statistics of >0.75. Of these polypeptide fragments, one was identified as the plasma glycoprotein, alpha chain fibrinogen. Results from the large-scale label free discovery experiments indicated twenty-six proteins to be differentially expressed between the three subject groups. These prospective markers include proteins previously linked to HCC as well as novel candidates, namely glutathione peroxidase 3, serum amyloid p, carboxypeptidase N and complement factors I and H which have not been implicated in the context of HCC diagnostics. Direct measurement of Hemopexin (HPX), alpha-1-antitrypsin (α1AT), apolipoprotein A1 (Apo A1) and complement component 3 (CC3) levels confirmed their change in abundance in LC and HCC versus control patients. Further biochemical characterization of glycosylated HPX isolated from glycoprotein enriched plasma sample pools showed evidence of isoelectric point shifts, indicating differential glycosylation patterns in high mannose structures of HPX which may be disease stage linked. The direct measurements of HPX, α1AT, Apo A1 & CC3 conducted on the independent Nigerian subject group also confirmed much of the trends reported from the Gambia Liver Cancer Study (GLCS) plasma.

Conclusions: The independently validated, significant changes in the quantitative expression of ApoA1, α1AT, CC3 and HPX could be exploited for development into high-performance affordable assays, usable in the diagnosis and monitoring of HCC and LC patients. The unique signatures observed for most of these proteins, from liver disease free controls to LC and HCC suggest their involvement in independent pathways. As such, combining some or all of these four markers within a diagnostic panel could offer a much-needed boost in robustness and accuracy for AFP. The differences in the processing and molecular weight separation of these proteins also offers a novel inroad into biomarker identification. These suggested disease specific signatures could with further study offer highly specific biomarkers able to discern the key stages that predispose individuals to hepatocarcinogenesis.

Impact: This is the first MS based discovery and extensive validation study on West African subjects whose primary cause of HCC are the Hepatitis B Virus (HBV) and fungal toxins.

Abbreviations

2DE – Two-dimensional electrophoresis
α1AT / AAT – Alpha-1-antitrypsin
A1AG1 - alpha-1-acid glycoprotein
AA – Acetic Acid
AACT - alpha-1-antichymotrypsin
AAL - Aleuria aurantia lectin
AF – Aflatoxin
AFB1 – Aflatoxin B1
AFP – Alpha-fetoprotein
ALP – Alkaline Phosphatase
ALT – Alanine aminotransferase
Apo A1 – Apolipoprotein A1
Apo J – Apolipoprotein J
Apo L1 - Apolipoprotein L1
Arg (R)- Arginine
ASR – Age Standardized Incidence Rate
AST – Aspartate aminotransferase
au – Arbitrary Units
AUC – Area Under Curve
BCA - Bicinchoninic Acid
BCP – Basal Core Promoter
BD –Bruker Daltonics
Bili T – Total Bilirubin
BME – β-mercaptoethanol
BPB – Bromophenol Blue
BSA - Bovine Serum Albumin
BTB – Blotting Transfer Buffer
BTI – Breakthrough Infections
$C_2H_3NaO_2$ – Sodium Acetate
C4 – Complement 4
$CaCl_2$ – Calcium Chloride
CC3 – Complement Component 3
CC4A – Complement Component 4A
cccDNA - covalently closed circular DNA
CERU – Ceruloplasmic
CFB - Complement Factor B
CFI – Complement Factor I
CH3CN - Acetonitrile
CHAPS-3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate
CI - Confidence Interval
CIP - Calf Intestinal Phosphatase
CLD – Chronic Liver Disease
Con A - Concanavalin A
CT – Computed Tomography
DNA – Deoxyribonucleic Acid
DSL - Datura stramonium lectin
DTT-Dithiothreitol

ECL - Enhanced chemiluminescence
ECM – Extracellular Matrix
EDTA - Ethylenediaminetetraacetic acid
ELISA – Enzyme Linked Immunosorbent Assay
EPI – Expanded Programme of Immunisation
ER - Endoplasmic Reticulum
EtOH - Ethanol
FA – Formic Acid
G – Guanine
GHIS – Gambia Hepatitis Intervention Study
GLCS – Gambia Liver Cancer Study
GNCR - Gambia National Cancer Registry
GP73 – Golgi Protein 73
H_2SO_4 – Sulphuric Acid
HBcAg – Hepatitis B Core Antigen
HBeAg – Hepatitis B 'e' Antigen
HBsAg – Hepatitis B Surface Antigen
HBSP – Hepatitis B spliced protein
HBV – Hepatitis B Virus
HCC – Hepatocellular Carcinoma
HCL – Hydrochloric Acid
HCV – Hepatitis C Virus
HEPES - 4-2-hydroxyethyl-1-piperazineethanesulfonic acid
HPX – Hemopexin
HRP - Horseradish Peroxidase
HSC – Hepatic Stellate Cells
ICAT – Isotope Coded Affinity Tags
IEF – Isoelectric Focussing
IgG – Immunoglobulin G
IPA – Ingenuity Pathway Analysis
IPG – Immobilized pH Gradient
iTRAQ - Isobaric tags for relative and absolute quantitation
JUTH - Jos University Teaching Hospital
kDa - Kilodaltons
LC – Liver Cirrhosis
LC-ESI – Liquid Chromatography Electrospray Ionization
LCA - *Lens culinaris* agglutinin
LDLT - Living donor liver transplantation
m/z – Mass-to-charge
MALDI – Matrix Assisted Laser Desorption Ionization
MeOH – Methanol
$MgCl_2$ – Magnesium Chloride
MMP – Matrix Metalloproteins
$MnCl_2$ – Manganese Chloride
MOPS – 3-(N-morpholino)propanesulfonic acid
MRC – Medical Research Council
MRM – Multiple Reaction Monitoring
MS – Mass Spectrometry
N-Glycans – Asparagine Linked Glycans
Na-DOC – Sodium Deoxycholate
NaCl – Sodium Chloride

NASC – Nigeria Asymptomatic Carriers
NASH – Nonalcoholic Steatohepatitis
NCirr – Nigeria Cirrhotic
NCon – Nigeria Control
NHCC – Nigeria Hepatocellular Carcinoma
NK – Natural Killer cells
NS5A - non-structural 5A
ORF – Open Reading Frame
p – p value
PAGE – Polyacrylamide Gel Electrophoresis
PCR – Polymerase Chain Reaction
PLGS – Protein Lynx Global Server
PRDX3 - Peroxiredoxin 3
Pre-C – Pre Core
PTM – Post Translational Modification
PVDF - Polyvinylidene fluoride
Q-TOF – Quadrupole Time of Flight
qHSC – quiescent Hepatic Stellate Cells
RFLP-PCR – Restriction Fragment Length Polymorphism Polymerase Chain Reaction
RNA – Ribonucleic Acid
ROC – Receiver Operating Characteristic
ROS – Reactive oxygen species
RVTH – Royal Victoria Teaching Hospital
SCC – Scientific Co-ordinating Committee
SCCA IgM - Squamous Cell Carcinoma Antigen-Immunoglobulin M
SDS – Sodium Dodecyl Sulphate
SDS PAGE - Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEM – Standard Error of Mean
Ser (S) – Serine
SILAC – Stable isotope labelling of amino acids in cell culture
SNP – Single Nucleotide Polymorphism
SVM – Support Vector Machine
T – Thymine
T2 – Top 2
TACE - Transarterial Chemoembolization
TBST – Tris Buffered Saline with Tween 20
TCA – Trichloroacetic Acid
Temed - Tetramethylethylenediamine
TGF-β1 – Tumour Growth Factor Beta 1
Thr – Threonine
TMB - 3,3’5,5’-tetramethylbenzidine
US – Ultrasound
V-Hr – Volt Hours
WHO – World Health Organization

Acknowledgements

In the name of God the most Gracious, the most Merciful.

"Seeking knowledge is obligatory upon every Muslim, male and female." Prophet Muhammad, Peace Be Upon Him (Ibn Majah)

All praise must go to God Almighty for facilitating me getting to this point with my DPhil. The years have simply flown by, yet with all that has happened – I have never felt alone and for that I give infinite praise and thanks for HIS blessings and the guidance and best of examples set by our Beloved Prophet Muhammad, Peace Be Upon Him.

That said – there are many people whose direction, support and expertise have been integral to getting to this stage. Firstly, to my supervisory team - Dr's Maimuna Mendy & Benedikt Kessler. The former particularly for identifying the need for a study such as this based on an African population and the latter for unreservedly opening up his impressive laboratory facilities for extensive use. I am thankful to the both of you for not only challenging me but also having an open door to the many requests &

resources required in running a study such as this. I have learnt much from you both and appreciate your time, guidance and gentle encouragement throughout the years.

To the un-named and un-identified "subjects" who are indeed very real people without whom there would be no GLCS or JUTH populations. May your contributions to this and other projects be a source of good for you where-ever you may be and may this project yield fruits from which many people will benefit.

To the team of Scientists, Clinicians, Nurses, Field workers, Lab techs and other support staff who designed the GLCS and undertook the laborious task of subject recruitment in both urban and rural areas – thank you. Thanks must also go to the MRC SCC & Ethics committees for granting us access to these precious & well characterized samples. As well, to the MRC Training Office, in particular Mariatou Sallah, for securing the funds and choosing to invest in young Africans like myself in the pursuit of higher education.

I am also very grateful to Mohamed Shariff, Nimzing Ladep & Simon Taylor-Robinson at Imperial College London for donating plasma from your JUTH subjects, which allowed us to validate some of the key findings from the discovery phase of this study.

To my colleagues at both the MRC, in the Gambia and the CCMP in Oxford for the training, advice and friendship you have offered throughout the years. At this juncture, I must mention Cynthia Wright-Drakesmith who was my mentor through the first 2 years of my PhD and played a significant role in establishing our study design

and key methodologies. Also of great assistance in Oxford were Drs Nicola Ternette, Holger Kramer, Mukram Mackeen, Hong-Lei Huang, Muna El-Kasti, Kati Di Gleria, Rebecca Konietzny, Climent Casals and the entire Kessler/Casals group. Alasana Bah & Adam Jeng were instrumental in getting me set up in The Gambia, particularly at the point of fishing out samples from the freezer archives. To my "neighbours" at Whittle in the MRC – thanks for the laughs, stress relief & random chatter. The light hearted moments of birthday surprises, dinners, secret santa gifts and lunchtime cook-ups & chats at both locales also provided well needed breaks.

The guidance offered by John Townend during the MALDI pilot phase of this study really set the stage for a lot of the statistical methods that were retained throughout this project. Dr Suzi Camey came into the project within its last year but was able to offer invaluable assistance on the practical methods of combining multiple biomarker candidates given our limited sample numbers. I am very grateful to the both of them for their invaluable support. Dr Andre DaCosta was also very helpful in helping to set up the collaboration with Suzi at IARC.

The many beautiful sisters I have met and made in Oxford have been a shining light throughout these years, providing a much needed *home away from home* in many instances. Assia, Nabila, Ayah, Sara, Ghita, Lamiya, Yasmine, Amany, Rawia, Yana, Iman, Kelly, Khayrah, Amina, Sadaf and all the "summertown sisters" as well as Maryam R. and everyone in Cowley – thank you. You friendship, generosity, sage advice and warmth has meant more than you know and I value & love you all.

To Kav & Zervan for all the "PhD/Medic" chit-chat and random emails from your worldly travels… keep them coming! Aminta – for the seemingly endless, enjoyable and random stream of communication as well as those late night skype marathons! May Allah Almighty continue to open doors of good for you and yours.

I am blessed to come from a home where love is always abundantly shown and to have Parents who value the pursuit of knowledge and are always calling us to what is true and right. Mum, Dad – I appreciate and love you more than words could express (so I'm not gonna try) but know that you are always at the heart of all my prayers. To my sibs Kebba, Isha & Abdou, the sisters-in-law and my beautiful twins & Saul, I love you all so very much.

Three months into the start of my DPhil – our family were immensely tested with the loss of my oldest brother Yusupha Fye. I never ever contemplated the possibility of your death but it happened and Allah Almighty got us through it. This thesis I dedicate to you & Uncle Lewally and continue to pray that Allah Almighty grants you both the highest of Jannah. Love always; your baby sis & favourite niece.

AlhamduliLlahwaShukriLlah Wa'Sal'Allahu Alayhi wa3alaa'Aalihi waSahbihi WaBaarik waSallim.

Chapter 1: Introduction

The principal aims set out at the onset of this project were to:

1. Use proteomic fingerprinting methods to identify a single or panel of highly specific and sensitive protein biomarker candidates (at least greater than the performance of Alpha-fetoprotein (AFP)) for use in the diagnosis of patients with Hepatocellular Carcinoma (HCC).

2. Validate any observed changes by Enzyme-linked immunosorbent assay (ELISA) based measurements performed on a subset of identified markers on samples from the Gambia Liver Cancer Study (GLCS) as well as collaborators within Africa.

3. Use classical proteomics methods for biochemical characterization of validated differentially expressed proteins in various subject groups to probe for more specific molecular changes potentially triggered by disease development.

1.0 HCC & Liver Disease in the Developing World

Hepatocellular carcinoma accounts for 85-90% of all tumours emerging from the liver in high incidence areas and between 70-75% of cases in lower incidence regions. It is the fifth most common cause of cancer related mortality worldwide for males and seventh for females (Parkin et al., 2002). The burden of this disease varies greatly in different areas of the world with the most affected areas being Asia, Sub-Saharan Africa and some parts of Eastern Europe where many of its causative factors are prevalent. The major risk factors for HCC development vary

depending on geography but in the developing world, chronic infection with the Hepatitis B Virus (HBV), Hepatitis C Virus (HCV) (Stuver, 1998) and exposure to Aflatoxin B1 (AFB1) (Ross et al., 1992) play the most prominent roles. These causes, particularly HBV and AFB1 will be given primary focus in this thesis as they represent the dominant aetiological factors in the populations under study. It is worth noting however that changing social trends have led to increased reports in the last few years which show rising levels of alcohol abuse (Morgan et al., 2004) and increasingly, diet and lifestyle related effects leading to non-alcoholic steatohepatitis (NASH), as eliciting a resurgence of HCC diagnoses in Western populations (Blonski et al., 2010).

HCC is a particularly deadly condition ranking third in overall global cancer mortality (Ferlay et al., 2010). In a 2008 survey overseen by the World Health Organisation (WHO), it was reported that there were approximately 748,000 new cases of liver cancer diagnosed worldwide, in that year alone, with an estimated 695,000 reported deaths in the same period (Ferlay et al., 2010). These figures show the high mortality rate of this condition owing to multiple contributing factors; the key of which are poor monitoring of at risk populations, insufficient diagnostic resources as well as very limited treatment options, much of which require early tumour identification for any potential of curative intervention. This thesis aims to tackle one of the primary reasons for the poor prognosis of this condition which is the lack of high performance, non-invasive tests allowing for affordable, sufficient and timely diagnostics and treatment, particularly in the developing world where the disease burden is greatest.

In any discussion of HCC, one cannot ignore the issue of liver cirrhosis (LC) which occurs as a result of long-term liver damage from viral or other factors. Studies show that as many as 90% of patients with HCC are affected by LC (Monto and Wright, 2001). In fact, LC is now regarded as a pre-neoplastic condition (Okuda, 2002) with studies showing up to a 100 fold increased risk for HCC development in patients chronically infected with HBV (Beasley et al., 1981).

In the Gambia where the main subjects under study in this project were drawn from, up to 15% of the population are suspected to be chronic HBV carriers by the time they reach adulthood (Vall Mayans et al., 1990; Whittle et al., 1990) with horizontal and perinatal transmission routes playing a prominent role in how the virus is transmitted (Chang, 2007; Hsu et al., 1993). HCC is the most common cause of cancer related death in Gambian men and the second in Gambian women (Bah et al., 2001) and consequently is a huge burden on the health sector.

In Nigeria, where the secondary cohort used in this project is sourced, no nationwide studies have been conducted in order to determine the overall prevalence of HBV infection. However, population studies conducted in urban centres show a Hepatitis B Surface Antigen (HBsAg) carriage rate of 9.1% with 87% of people under 40 years showing the presence of at least one serological HBV marker. Similar to the Gambia, HBsAg transmission tends to occur earlier in life either through perinatal or horizontal transmission, leading to a significantly lowered peak age of HCC development. This is particularly highlighted in a study from the Jos University Teaching Hospital (JUTH) where 72% of HCC's diagnosed

over a 10 year period were in patients aged 21 to 50 years (Echejoh et al., 2008). No local nationwide studies surveying HCC rates have been published but GLOBOCAN reports the age-standardised incidence rate for HCC in Nigeria as 13.9 per 100,000 (Ferlay et al., 2010) with local hospital records suggesting HCC to be the most common cause of cancer related admission and death in the country (Ndububa et al., 2001; Olubuyide, 1992).

Thus our two subject populations offer similar demographics in liver cancer aetiology and spread. Plasma samples from the GLCS case-control study will serve as the discovery population with independent validations subsequently conducted using novel methods and a distinct pilot subject populace.

1.1 Viral Hepatitis & HCC

1.1.1 Impact of the HBV

The geographical distribution of HCC is largely an exact mirror of that for the chronic carriage of the hepatitis B & C viruses.

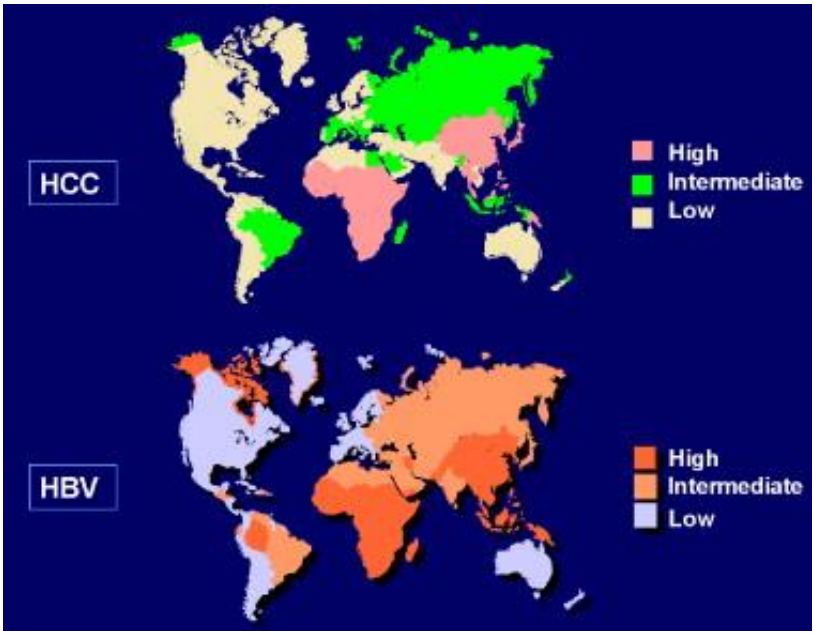


Figure 1.1: Image showing how trends in worldwide incidence of HBV infection and HCC development correlate. (Kew, 2010a)

	The Gambia	°China/◆Taiwan
HBsAg Prevalence %	≤16% (Kirk et al., 2004)	°12% (Parkin, 2006)
Anti-HCV Antibody Prevalence %	2.9% (Kirk et al., 2004)	°3.0% (Parkin, 2006)
Relative Risk HBV	13.5 (95% CI, 7.8–23.2) (Kirk et al., 2004)	°17.0 (95%CI, 4.3-99.4) (Yeh et al., 1985)
Relative Risk HCV	14.7 (95% CI, 6.3–34.4) (Kirk et al., 2004)	◆21.5 (95% CI, 9.3, 49.9) (Sun et al., 2003)
Age Standardized Incidence Rate (ASR) HCC	35.7 per 100,00 males 11.2 per 100, 000 females (Bah et al., 2001)	°37.9 per 100,000 males 14.2 per 100,000 females (Parkin et al., 2005)

Table 1.1: Summary of HBV and HCV associated risk factors for HCC in The Gambian population contrasted with Chinese and Taiwanese epidemiological data.

Chronic carriage of the HBV is a major risk factor for HCC and has been attributed as a primary cause in more than half its cases worldwide (Parkin, 2006). Globally, over a billion people have been infected with the HBV with 350 million of them developing chronic infection (Organization, 2007). HBV is an oncogenic virus; able to integrate its Deoxyribonucleic Acid (DNA) into that of the host hepatocytes in a process observed not only in hepatoma cell lines but also in as much as 80% of HBV related HCC's (Brechot et al., 1980; Chen et al., 1986). Chronic HBV infection is defined as showing persistent HBsAg positivity for at least 6 months. Children born to hepatitis B 'e antigen' (HBeAg) positive mothers are at highest risk for chronic infection with studies showing as much as 80-90% of them becoming chronic carriers (Beasley et al., 1977; Beasley et al., 1983). Horizontal infection also poses a great risk of lifelong carriage, with research showing infection prior to age 6 resulting in 23-49.2% (Beasley et al., 1982) HBV chronicity as compared to <5% in adults (Hyams, 1995). Detectable levels of serum HBV DNA have been shown to have a linear relationship with the risk of HCC development with a substantial increase occurring once HBV DNA reaches upwards of 10^4 copies/ml (Chen et al., 2006). This increased risk is independent of overall liver function as indicated by serum alanine transaminase (ALT) levels, HBeAg status or the presence or degree of cirrhosis.

Other factors associated with increased likelihood of hepatocyte carcinogenesis include infection with the genotype C strain of HBV (Chan et al., 2004), presence of the basal core promoter (BCP) A1762T and G1764A double mutation, the pre-core (pre-C) G1896A mutation (Yang et al., 2008) and pre-S mutations C1653T,

T1753V, and A1762T/G1764A (Liu et al., 2009) most of which accumulate in the virus genome during chronic infection. Other key risk factors for HCC development are elevated serum ALT as well as other demographic and lifestyle features such as, male sex, age, family history of HCC, exposure to aflatoxins and high levels of alcohol and or tobacco consumption.

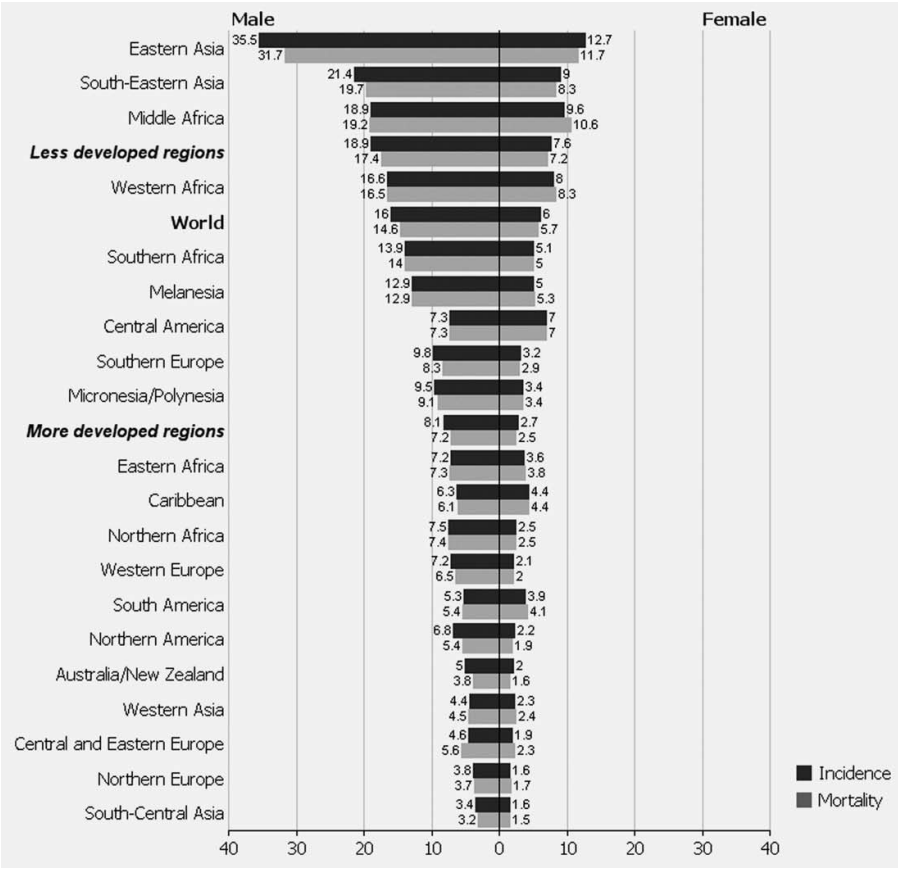


Figure 1.2: Global ASR for liver cancer showing greatest disease burden for males and in parts of Asia and Africa. Source: GLOBOCAN 2008 (Ferlay et al., 2010).

1.1.2 Hepatitis C Virus and HCC

The HCV virus, discovered in 1989 (Choo et al., 1989) is now known to have 170 million carriers worldwide and has been strongly implicated in the development

of HCC (Colombo, 1999; Di Bisceglie, 1997). Its primary mode of transmission is through intravenous drug use (IVDU), blood transfusions & haemodialysis. The role of sexual activity in HCV transmission remains ambiguous, with spousal studies in Egypt and Italy showing low to zero levels of transmission; 6% in the Egyptian cohort, (Magder et al., 2005) where the overall HCV rate is over 14%, and no evidence of spousal transmission in Italy (Vandelli et al., 2004).

HCC risk in HCV infected subjects is up to 20 times higher than in virus free individuals (Sun et al., 2003). Recent evidence suggests that between 2-5% of patients with HCV induced LC go on to develop HCC each year (Schuppan et al., 2003). The question of whether the virus triggers liver carcinogenesis directly or through indirect means such as chronic inflammation and tissue regeneration has not been definitively resolved. However, the fact that HCC rarely occurs in auto-immune hepatitis suggests that there are direct viral factors which facilitate liver carcinogenesis. The HCV core and non-structural 5A (NS5A) proteins are the most researched viral proteins linked to the development of HCC, with studies in mice showing transfection with the core protein leading to hepatic steatosis within six months and HCC after sixteen months (Moriya et al., 1997; Moriya et al., 1998). Similar results were obtained with NS5A transgenic mice, which developed a range of liver pathologies (Wang et al., 2009a).

In Africa, the WHO estimates the highest HCV prevalence rates at 8.2% in North Africa with estimates of overall occurrence in West Africa projected at around 2.2% (Parkin, 2006). Overall ASR for liver cancer ranges from 35.5 per 100,000 in

China and Eastern Asia to 3.8 per 100,000 in parts of Europe (Ferlay et al., 2010). More than 80% of the diagnoses of this cancer come from countries in the developing world. The disparities in these figures represent the differences in exposure to the key viral and environmental aetiological factors for HCC. There has, as previously mentioned however been a recent resurgence of HCC diagnoses in western populations. This trend, if not arrested may well re-instate HCC as a significant cause of morbidity and mortality in the developed world.

1.2 Hepatitis Studies in the Gambia

1.2.1 A Study of Two Villages

Liver diseases and their associated aetiologies and effects have been researched within the Gambian population for close to three decades. Early studies showed that up to 1 in 50 Gambian men were likely to die from primary HCC (Ryder, 1982) with links made to HBV as a major causative factor even prior to the availability of widespread local epidemiological data (Barin et al., 1981). Of the first major works detailing HBV prevalence in the Gambia was a study based on two neighbouring villages; Keneba and Manduar located in the Kiang West, lower river region of the country (figure 1.3).

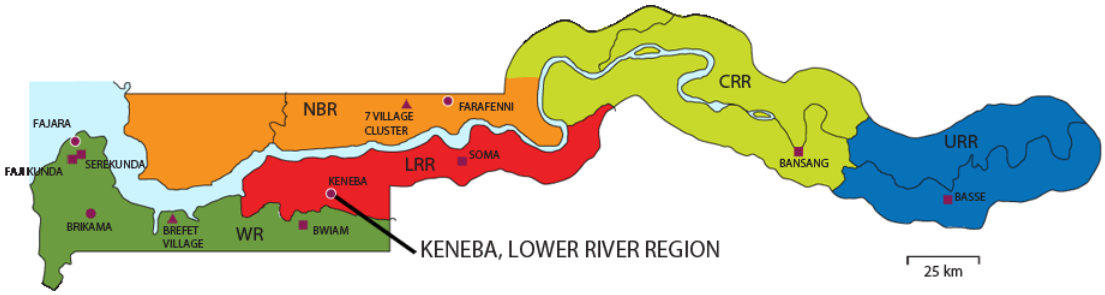


Figure 1.3: Outline of major regions of the Gambia (source: Medical Research Council (MRC) The Gambia Website www.mrc.gm)

Although the overall prevalence rates for HBV were distinct, with 27% of 2-4 year olds in the former village seen to carry the HBsAg as opposed to 65% in the latter, there were still some consistent trends which offered much needed insight into virus expression patterns in the country. HBV infection was seen to cluster in families with most of the transmission occurring horizontally through siblings; a trend distinct from earlier reports from far eastern hubs (Stevens et al., 1975; Yao, 1996) where the hepatitis viruses are a huge burden. The study of these two villages also established links in the Gambian context between the presence of the serological HBeAg marker and a higher likelihood of viral transmission. This was demonstrated in an evaluation of eleven children born to four highly infectious, HBeAg positive mothers. Ten of them were subsequently found to be positive for the HBsAg marker. Whittle and colleagues in this publication also reported the diagnoses of HCC in four adults all found to be HBsAg positive; two at a prior assessment and the other two at diagnosis (Whittle et al., 1983). This provides perhaps the earliest published link between chronic HBV carriage and hepatocarcinogenesis in Gambian individuals.

1.2.2 Establishing HBV related HCC in the Gambia

A secondary survey was repeated approximately 5 years later in Keneba and Manduar, upholding many of the primary findings detailed above, including the significant role of horizontal transmission in early infection. Male children were found to more likely carry the hepatitis B surface and "e" antigens concomitantly (Whittle et al., 1990) suggesting a lengthier phase of active viral replication. This may be one of the factors which contribute to higher rates of HCC diagnoses in males; possibly as result of the prolonged presence of detectable levels of circulatory viral DNA increasing the likelihood of its integration into the host genome (You et al., 2004). Research from Senegal conducted around the same time also looked into how age and sex related differences affect the establishment of a chronic carrier state. Their results highlighted an increased susceptibility to chronic HBV carriage by male children infected with HBV at <2years (Coursaget et al., 1987).

1.2.3 Intervention to curb transmission & high prevalence

In light of the research detailed above and the growing recognition of the impact of liver diseases in the developing world (Organization, 1983); the Gambia Hepatitis Intervention Study (GHIS) was launched. This was a randomized control trial whose primary aim was to vaccinate infants over a 4 year period and follow them up over three to four decades in order to determine key outcomes amongst which was whether a significant reduction in the sequelae of HBV associated liver diseases would be effected (Group, 1987). Four vaccination strategies were

applied which despite significant differences in their resultant antibody titres demonstrated 96% effectiveness in preventing chronic HBV infection (Whittle et al., 1991). Some degree of acute breakthrough infections (BTI) were suggested by observances of HBV core antibody positivity but these were mainly linked to low vaccine triggered antibody responses and exposure to chronically infected mothers (Whittle et al., 1991; Fortuin et al., 1994).

As markers of aflatoxin exposure became available (DL and Groopman, 1994), studies were conducted in the Gambia to examine the concomitant effects of widespread AFB1 ingestion, determined by measurement of aflatoxin-albumin adduct levels in peripheral blood and chronic HBV carriage (Wild et al., 1992). Widespread adduct presence was seen in over 95% of subjects under study with evidence of *in utero* exposure also presented following adduct detection in umbilical blood. Marked changes in AFB1 exposure were linked to seasonal variations, which in the Gambia where groundnut harvests are made annually suggested possible avenues of intervention and education of farmers with regard to long-term storage and handling of these grains. Interestingly, it was in the sub-group of children positive for the HBV that the highest aflatoxin-albumin adduct levels were recorded (Wild et al., 1992) suggesting that the presence of the virus may in some way boost or synergise with this toxin.

1.2.4 National Cancer Registry and HCC profiling

The Gambia National Cancer Registry (GNCR) established alongside the GHIS was the first population based national cancer registry established in Africa. It continues to serve as a major source of insight and outcomes on the effects of the inclusion of the HBV vaccine into the nations expanded programme of immunisation (EPI). Even at its infancy, the GNCR rapidly highlighted the large impact of liver cancer on mortality and morbidity, particularly in males but also in females (Bah et al., 1990; Bah et al., 2001). This long-term venture has projected outcomes divided into three key phases; the first of which was the inclusion of the vaccine into the nations EPI which was successfully established in 1987. The subsequent phase looked at vaccine efficacy in light of protection against infection and chronic carriage (Viviani et al., 1999; Viviani et al., 2008).

The final phase of the project, which is now well into its third decade is to follow up vaccinated children in the trial through the cancer registry to see if rates of HCC, taken as the major endpoint, would decrease. Should this endpoint be successfully reached, the project will serve as an important milestone in the implementation of a vaccine capable of curbing the effects of one of the major aetiological factors for HCC in Sub-Saharan Africa. Recent publications from the GHIS suggest a marked rise in the burden of liver cancer in females (Sighoko et al., 2011); a finding which indicates potential novel factors in the transmission of HBV in this population or perhaps an increase in the effects of newly recognised aetiological factors such as diabetes and obesity which anecdotal evidence shows to predominantly affect women in this environment.

1.2.5 Identification of unique profiles & methods for HCC Diagnostics in Gambian Subjects

Although the implementation of widespread vaccination strategies gives hope of an anticipated decline in HCC incidence, most hepatology researchers recognize that decades could pass before the full benefits of such interventions are effected. This reality combined with other largely un-tackled factors leading to HCC (figure 1.4) mean that efforts to significantly reduce the prevalence of this cancer must continue to be multi faceted (Kew, 2010a; Regimbeau et al., 2004). Of the key areas requiring renewed focus is in the identification of unique molecular signatures exploitable as highly sensitive and specific tests for HCC diagnosis (Mendy and Walton, 2009). Prior to the establishment of this study, research by Kirk and colleagues compared the prevalence of the "hotspot" 249 arginine (Arg, R) to serine (Ser, S) mutation (Bressac et al., 1991; Hsu et al., 1991; Ozturk, 1991) on the p53 gene. This substitution mutation is known to be prevalent in the HCC genomic profiles of populations with high HBV prevalence combined with large exposure to aflatoxin. To investigate this, isolates of cell-free DNA were taken from the plasma of HCC, LC and control subjects in the Gambia and compared with those of individuals of non-African origin living in France, suffering from a spectrum of liver diseases (Kirk et al., 2000). Following detection by restriction endonuclease digestion of polymerase chain reaction (PCR) products and direct sequencing and amplification of DNA, adjusted odds ratios for the presence of the mutation were calculated as 16.4 (95% CI 3.0-90.5) for the HCC subjects compared to controls. This "hotspot" mutation was not detected in any of the French sourced subjects and was observed at lowered rates in the LC and control

Gambian groups. This suggests the tracking of this alteration as a possible early predictor of hepatocarcinogenesis. An ESI mass spectrometry based study looking at the quantitative expression of the serine 249 mutation in the GLCS also drew a similar conclusion; higher plasma serine 249 concentrations existed in the HCC subjects and this significantly correlated with a diagnosis of HCC (Lleonart et al., 2005).

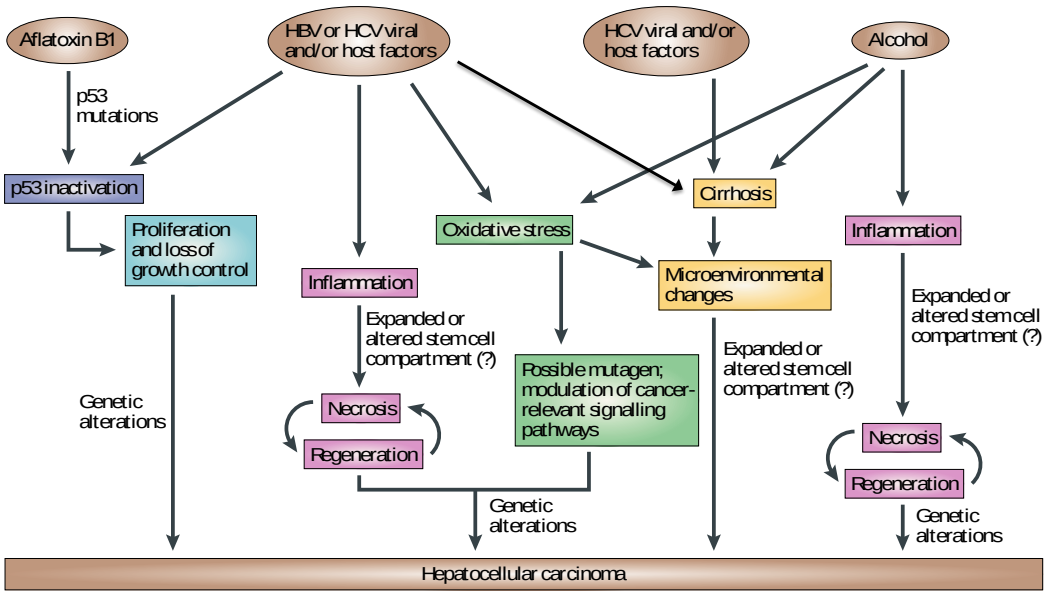


Figure 1.4: Summary of major risk factors for hepatocarcinogenesis; AFB1, HCV, HBV and excessive alcohol consumption and how they may interact and lead to genetic and histological changes which facilitate the development of HCC. Adapted from Farazi & DePihno 2006 (Farazi and DePinho, 2006).

More investigations have been conducted in the Gambia in the area of targeted diagnostics applicable to the resource-constrained settings of the developing world. A unique study looking at the determination of HBsAg status and AFP levels from blood spots dried on filter paper showed that both approaches were viable and produced reliable results when compared to established methods

(Mendy et al., 2005). Although the standalone performance of AFP in the diagnosis of liver cancer, particularly that of viral origin, is not optimal (Soresi et al., 2003) – having robust methods of sample processing which will allow for delayed analysis in field based investigations is useful and could potentially be applied to new marker panels suggested from studies like ours. These are avenues of research which need to be followed up in the developing world in order to produce tools and tests applicable to its environment and resource constraints.

1.2.6 Gambia Liver Cancer Study

Given its small population and geographical landscape, the Gambia has proven ideal and significant in its contribution to the understanding of how chronic liver diseases (CLD) are distributed and manifest in sub-Saharan Africa. Though the subjects utilised in this project have liver diseases predominantly associated with HBV and aflatoxins, research suggests that up to 50% of local HCC cases could be due to non-viral factors (Mboto et al., 2005). To help estimate the relative impact of the key viral and non-viral causes of LC and HCC, the GLCS was established. This was a large, multicentre case-control study with subject recruitment from both urban and rural populations in age groups largely unaffected by the mass HBV vaccination programmes started in the previous decade. It is this subject group that forms the backbone of much of the work in this biomarker discovery project.

The specific disease and control group definitions will be more thoroughly introduced in the subsequent chapters, however, a key factor integral to the interpretation of any results from this investigation is that of which applied cut off level for AFP was used in subject recruitment as suggestive of HCC. No accepted global diagnostic values exist – although the three most commonly used are 20ng/ml, 100ng/ml and 400ng/ml. Which of these is adapted in clinical studies however varies depending on the characteristics of the subject population as well as the research question being posed. This was demonstrated in a recent study where two AFP cut off's were investigated in relation to the expressed protein levels prior to tumour resection surgery (Zhou et al., 2012). No studies have been conducted to date comparing the effectiveness of any of these three oft-utilised cut offs' in relation to one another. Although AFP and imaging methods serve as important indicators of HCC, they are not gold standard markers and in the absence of access to subject cohorts with 100% liver biopsies conducted on all HCC and LC cases, researchers in studies such as this have to be aware of the likelihood of some degree of mis-diagnoses intrinsic to their case definitions. The presence of small tumours or possible metastases from other organs may only be reliably discernible by liver biopsy and not necessarily reflected in biochemical and radiological tests. As the characteristics of HCC may differ geographically due to changes in the tumours primary aetiological agents, it may be pertinent to conduct studies which aim to identify the most precise and effective diagnostic cut off for AFP and other potential markers under development as well as striving for more robust outlines for accepting primary HCC and LC cases.

Overall, the GLCS confirmed that HBV remained the biggest contributor to HCC risk in the Gambian population with a dose-dependent relationship seen between viral load and likelihood of LC and HCC development particularly at levels >10,000 copies/ml (Mendy et al., 2010). HCV infection was found to be pronounced in HCC subjects mainly of older age, suggesting a cohort of people likely mass infected in a similar manner. To elucidate this finding, genomic-based analyses would have to be conducted on this sub-group of subjects to see if the expressed HCV sub-types are uniform.

1.3 Aflatoxin B1 as a risk factor for HCC

Approximately half of HCC's in regions with high AFB1 exposure harbour mutations in the p53 tumour suppressor gene, the most common of which is a Guanine (G) to Thymine (T) substitution at the third position of codon 249 (AGG) resulting in an amino acid shift of Arg to Ser. Several studies have been conducted in order to investigate the possible association between exposure to this toxin and human HCC. Significant correlations were seen in research conducted across southern Africa (Peers et al., 1987; Van Rensburg et al., 1985), China (Yeh et al., 1989) and Kenya (Autrup et al., 1987) where a significant link was made between AFB1 exposure and HCC mortality. The changes triggered by AFB1 are thought to restrict the capacity of p53 to suppress tumour growth and when combined with chronic viral infections act synergistically (Qian et al., 1994) to increase the likelihood of carcinogenesis, particularly in the liver (Groopman and Kensler, 2005; Szymanska et al., 2004).

AFB1 is produced by the fungi *aspergillus flavus* and is found as a contaminant on poorly stored grains and nuts. In populations with large exposure to this toxin, there appears to be a dose-response link between the risk of HCC development and the amount of aflatoxin ingested (Wogan, 1992). The effects of this synergy between AFB1 ingestion and viral hepatitis are especially highlighted in a Chinese study showing the relative risk of HCC development as 3.4 in individuals exposed solely to AFB1 and 7.3 in those chronically infected with the HBV. In patients with both factors present, the relative risk of HCC development rose to 59.4. (Qian et al., 1994; Ross et al., 1992)

In the Gambia, levels of AFB1 exposure are ubiquitous depending on diet and seasonal availability of grains and nuts. However, studies have shown the presence of aflatoxin-albumin adducts in as much as 95% of the population (Wild et al., 1990; Wild et al., 1992) suggesting widespread exposure. As such, in the West African context where grains and groundnuts form key parts of the staple diet, the possible effects of this mycotoxin in liver cancer development cannot be ignored.

The viral and environmental causes of HCC are all important in contributing to its development, however the sustained links observed between chronic infection and HCC make infection with the "B" and "C" hepatitis viruses the most significant contributing factors to the development of this cancer. In the presence of these infections, factors such as toxin exposure and alcohol serve to further

enhance and promote the risk of HCC development but may not necessarily be causative.

1.4 Molecular mechanisms of HBV related HCC Development

The HBV belongs to the family of hepadna viruses and has unique features including its circular DNA formation, presence of partially double stranded sections, use of reverse transcription in its replication cycle and of importance to HCC development, the persistence of its DNA in infected cells either as covalently closed circular DNA (cccDNA) which acts as a template for viral ribonucleic acid (RNA) transcription or by direct integration within the chromosomes of infected hepatocytes. The HBV is a 3.2 kb, 42nm DNA virus composed of four partially overlapping open reading frames (ORF) which code for the core (core/pre-core), envelope (S/ pre-S), polymerase and HBV X proteins (figure 1.5).

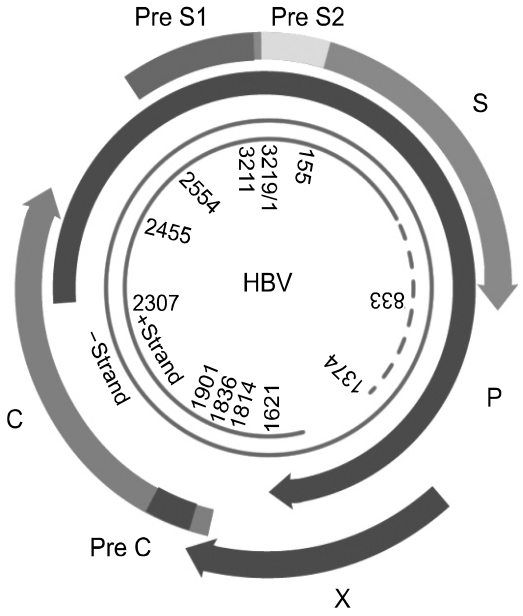


Figure 1.5: Source: publication by Park et al., 2007 - Illustration of the HBV genome in its circular format highlighting its key pre-S1, S2 and S regions as well as the polymerase, core and HBV X coding regions. Source: Molecular Pathogenesis of Hepatitis-B-virus associated HCC. (Park et al., 2007).

Primary HCC is a complex multistep disease, which arises from a myriad of environmental, host genetic and viral factors (figure 1.4). Researchers have for decades struggled to propose a well-defined mechanism of how HCC develops with uniform scientific backing of proposed mechanisms consistently lacking. What is clear is that the state of chronic inflammation of the liver, irrespective of its cause is an established precursor of LC and HCC by inducing hepatocyte proliferation and genetic alteration (Huang and Chisari, 1995). One of the major factors associated with the initiation of these key changes is chronic infection with the HBV. This is known to lead to repeated cycles of immune system mediated hepatocyte necrosis, proliferation and inflammation (Ganem and Prince, 2004). Natural killer (NK) cells, bystander lymphocytes and chemokine mediated neutrophil infiltrations have been implicated in causing extensive liver damage, with the release of cytokines and chemokines by these mediators thought to promote tumour growth in the site of chronic injury (Balkwill and Mantovani, 2001). The state of chronic inflammation of the liver which results in an up-regulated production of reactive oxygen species (ROS) causing DNA damage and mutations (Balkwill and Mantovani, 2001), also activates key signal transduction pathways that regulate apoptosis, cell proliferation and differentiation including that of NF-kapaB. With the activation of these pathways and the state of chronic inflammation and hepatocyte proliferation, oxidative stress ensues and an influx of kupffer cells promote the activation of Hepatic stellate cells (HSC) which are

the main producers of extracellular matrix (ECM) proteins in the liver, resulting in extensive, irreversible injury (Safadi and Friedman, 2002).

Research shows that up to 10% of adults who encounter the hepatitis B virus will be unable to clear it and become chronic carriers. From this, a fraction will develop HCC, with or without liver cirrhosis (Okuda et al., 1982) with a sub-set of them not displaying viral antigens in their sera (occult hepatitis) (Pollicino et al., 2004). Occult hepatitis is increasingly being recognized as a significant cause of HCC with long-term undetected infection inciting an indirect oncogenic role. These observations support suggestions that HBV is itself a direct trigger for HCC as it has been shown to incorporate into host DNA in up to 80% of cases (Matsubara and Tokino, 1990; Brechot et al., 1980; Hino et al., 1984). There also appears to be a link between HBV DNA load and HCC risk. A longitudinal study of 3653 HBsAg infected individuals in Taiwan demonstrated that elevated baseline serum HBV DNA levels were strong predictors of liver carcinogenesis and that this was independent of serum ALT level, LC or HBeAg status (Chen et al., 2006). The mechanism through which this leads to tumour development has been proposed to be via the HBV viral proteins, the most important of which is the 154 amino acid long HBV X viral protein. The surface pre-S2/S proteins and more recently the hepatitis B spliced protein (HBSP) have also been implicated (Kremsdorf et al., 2006) in the promotion of hepatocarcinogenesis.

It is the HBV X protein however that much research over the past decade has been focussed on as a potential key mediator of HBV induced

hepatocarcinogenesis. This protein is highly conserved among mammalian hepadnaviruses and is known to interfere with important cellular processes including apoptosis, proliferation, senescence and p53 mediated DNA repair. Tumour suppressor genes when functional curb unregulated cell proliferation; hyperplasia, which leads to dysplasia and ultimately tumour development. In addition to its potential inhibition by the HBV X protein, p53 has also been found to be mutated in 30-60% of patients with HCC (Staib et al., 2003) though there is some evidence suggesting that this mutation may only be involved in tumour progression, not initiation (Teramoto et al., 1994). Studies with transgenic mice have been used to investigate the impact of the presence of this protein on cancer promotion. The results have been mixed with HCC developing in some mice transfected with the gene but not all. A notable finding from these experiments was that HBV X increased HCC development upon concurrent exposure to other carcinogens suggesting some degree of synergism between viral and environmental factors.

There is also evidence to suggest that truncated forms of the HBV surface proteins have oncogenic potential and escalate hepatocyte proliferation by increasing the expression of key cyclins which control cell cycle progression (Hildt et al., 2002). These truncated surface proteins have also been linked to the promotion of oxidative stress by accumulating within the endoplasmic reticulum (ER).

Another aspect of the HBV suspected to be associated with HCC development are its expressed genotypes. HBV has been classified into eight genotypes, A – H, each with distinct geographical distributions. Genotype A is prevalent in north-western Europe, north America and South Africa, types B and C mainly centred in South-Eastern and Eastern Asia, genotype D found mainly in Southern Europe and India, genotype E predominantly restricted to West and Southern Africa, types F and H in Central and South America and genotype G mainly identified in north America and parts of Western Europe (Chemin and Zoulim, 2009). Certain forms of these viral genotypes are associated with heightened risk for chronic infection, active viral replication and HCC (Kirk et al., 2006). A study in Taiwan found infection with HBV genotype B to result in accelerated HCC development in comparison to the C subtype (Kidd-Ljunggren et al., 2002).

The mechanisms described above offer only a small window into the processes that may govern tumorigenesis in the liver. As the process of HCC development is drawn out, often taking several decades to be fully established – it is difficult to home in on key factors which are specifically causative of or promote this disease. What is more likely is that the pathogenesis from HBV which ensues following infection, is promoted by environmental and host-genetic factors triggering molecular changes that facilitate oncogenic transformations. Much research is still required to fully describe the mechanisms of these changes.

1.5 Mechanisms or Liver Fibrosis and Cirrhosis

Liver cirrhosis is an end-stage manifestation of chronic diffuse liver disease. At its onset, it is characterized by liver fibrosis, which results from acute liver damage from hepatocyte injurious agents (figure 1.4). With repeated insult, this leads to the accumulation of ECM proteins produced by HSC (figure 1.6); a characteristic signature of most CLD (Friedman, 2003). In healthy liver, a regulated balance exists between ECM synthesis and degradation as well as a balance of its constituents (Benyon and Iredale, 2000). With repeated injury, this equilibrium is lost and excess ECM accumulates with a higher proportion of the collagens I and III and fibronectin amongst its usual constituents of proteoglycans, laminin and matrix metalloproteins (MMP). This altered composition triggers increased cross-linking leading to a progressive thickening of the liver parenchyma creating a functional impediment to the flow of blood and as such compromising liver function.

One of the major drivers of this scarring mechanism is the activation of HSCs (Friedman et al., 1985), portal fibroblasts and myofibroblasts from the bone marrow (Ramadori and Saile, 2004; Forbes et al., 2004). With continued exposure to injury causing agents, the hepatocytes attempt to regenerate and replace the necrotic and apoptotic parenchymal cells in a process mediated by the conversion of quiescent HSC (qHSC) to activated HSC. This process also triggers the release of pro-inflammatory, pro-mitogenic and pro-fibrogenic cytokines such as angiotensin II, leptin and Tumour Growth Factor Beta 1 (TGF-β1) (Bataller and Brenner, 2005) which help maintain a persistent inflammatory response.

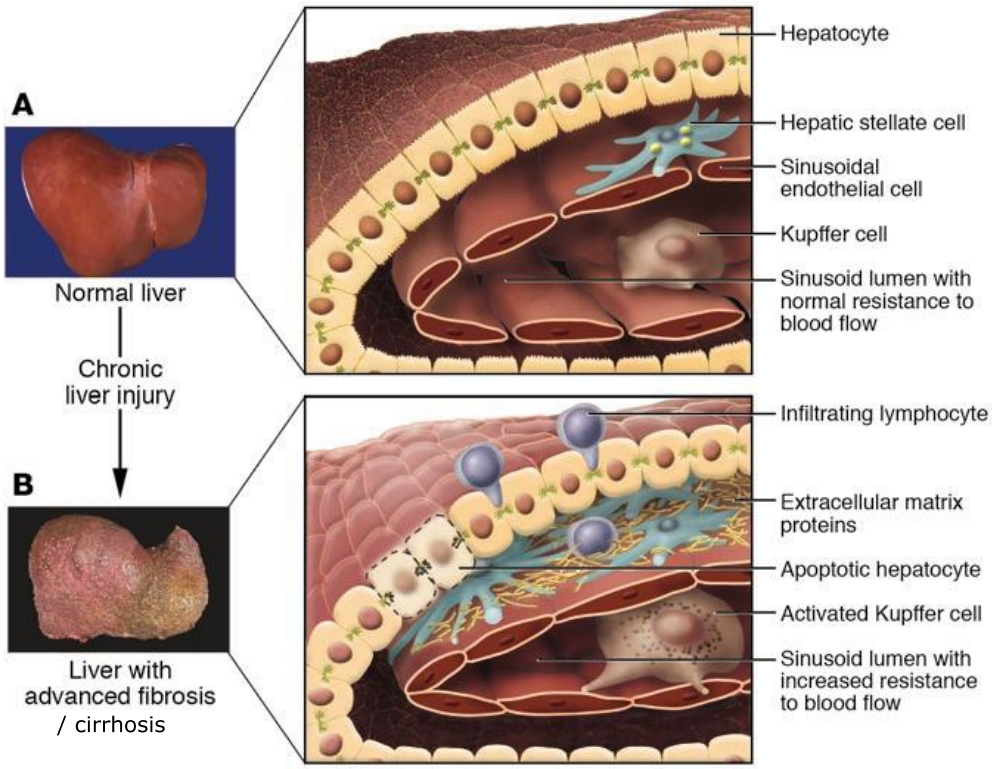


Figure 1.6: Source: (Bataller and Brenner, 2005) J Clin Invest. 2005; 115(2):209–218. Adapted image showing the series of changes which occur in the liver as a result of chronic injury. A) Healthy liver with functional hepatocytes, balanced ECM and optimal sinusoidal lumen; B) Chronically injured liver with formation of excess collagen rich ECM characteristic of advanced fibrosis.

All of these activities if uncurtailed continue to advance fibrosis to cirrhosis rendering much of the liver parenchyma structurally abnormal. These alterations are primarily characterized by the formation of dysfunctional hepatocyte nodules resistant to intrahepatic blood flow, one of the major causes of hepatic insufficiency and portal hypertension (Gines et al., 2004). These changes manifest as macronodular, micronodular or mixed macro/micro-nodular cirrhosis, a form largely associated with viral hepatitis (Okuda, 2007).

Much research is still required to elucidate and describe the mechanistic changes which predispose patients with LC to HCC. What we do know however is that there is a marked decrease in hepatocyte proliferation during liver cirrhosis (Delhaye et al., 1996) signalling the exhaustion of the regenerative mechanisms of the liver in the face of chronic exposure to injury causing agents.

Three basic mechanisms have been proposed which may result in the acceleration to HCC from LC. Evidence showing the distinct shortening of telomeres in human HCC (Plentz et al., 2004; Plentz et al., 2007) as well as its linkage to aging and limitation of the proliferative capability of hepatocytes and other primary human cells makes this one of the dominant proposed mechanisms of how LC predisposes to HCC. Impaired hepatocyte proliferation could also in itself promote carcinogenesis as seen by evidence from rat studies where exposure to chemicals which mediate the inhibition of hepatocyte proliferation accelerated liver tumorigenesis and p53 damage (van Gijssel et al., 1997). Local and systemic factors resulting in an altered milieu have also been implicated as promoting cancer risk during cirrhosis. One particular aspect is that of liver mass maintenance which is speculated to be regulated by growth factors within a feedback loop. In a scenario where due to extensive scarring and formation of fibrous tissue, liver cell mass is lost, this growth stimulation may continue and trigger tumour formation (El-Serag and Rudolph, 2007). In addition, the activation of stellate cells, a hallmark of LC, is known to lead to the generation of products of oxidative stress (Bataller and Brenner, 2005) that could also provide an optimised environment for HCC development.

There is growing evidence suggesting that both fibrosis and cirrhosis are to some degree, reversible – leading to an upsurge of interest in the targeting of their key mechanisms in drug development (Hammel et al., 2001; Bonis et al., 2001). Until these potential therapies become established and widely available, it is still vital that efforts are employed to slow or curb patient progression from fibrosis to more advanced liver disease. The presence of oral anti-viral treatments such as lamivudine, tenofovir and adefovir for chronic HBV have been shown to prevent CLD complications so long as resistance does not emerge (Yu et al., 2011). Combination treatment with pegylated interferon and ribavirin is increasingly being used in the developing world for HCV infection (Fried et al., 2002), although it is expensive and associated with significant side effects including hematologic toxicity and depression (Ward and Kugelmas, 2005).

1.6 Treatment of HCC

Following a timely diagnosis of HCC; the biggest challenge facing Physicians is the identification of viable treatment options that take into account the patients overall condition, cancer stage and disease background. As it is, prognosis following diagnosis of primary HCC is very poor, with this being particularly pronounced in the developing world. Overall, the main approaches to liver cancer treatments can be classed as surgical and non-surgical interventions, as well as the use of adjuvant therapies that aim to reduce the risk of recurrence following successful intervention.

Several factors come into play when making decisions on which treatment course to follow. Primary amongst them is an assessment of the overall health of the patients' liver. Surgical resection, the recommended first line treatment for viable candidates is only advisable in individuals without LC (Bruix and Sherman, 2011) where resection of cancerous growth will with time expectedly lead to a restoration of normal liver function. In those with advanced liver disease or a background of cirrhosis, liver transplantation is still the only curative treatment option available. The biggest limitation here is the availability of matching organs as well as that of the expertise and facilities required for such a specialist procedure. The "Milan criteria" which is commonly used to assess transplant candidates, recommends that it be conducted in patients with no more than a single tumour nodule of <5cm or 2 to 3 nodules of no more than 3cm each, with no macrovascular invasion. This definition has been put under review in the last few years to determine if it is too restrictive (Mazzaferro et al., 2009).

In cases where surgical intervention is not recommended or available; a factor of much relevance to the developing world, non-surgical approaches must be considered. These include ablative therapies that aim to destroy localized tumours by direct injection of chemicals such as ethanol (EtOH) or the use of radiofrequency based ablation. The latter is now being proposed as a treatment of choice for small HCC's following a multi-centre hospital based study which with 31 months follow up showed a sustained response in >97% of patients (Livraghi et al., 2008).

The application of chemotherapy via Transarterial Chemoembolization (TACE) targeting the blood supply network of tumours has been shown to improve survival in patients with un-resectable growths (Llovet and Bruix, 2003). The hepatic artery is the main source of oxygenated blood to the liver and consequently any tumours present within it. Injecting chemotherapeutic drugs into this artery allows the treatment to accumulate in the tumour area and trigger de-vascularization of its rich blood supply. Interest in this approach began in the 1970's and has since grown with evidence showing that it reduces tumour size (Ryder et al., 1996) although no increase in survival has been reported (Ryder, 2003).

The multikinase inhibitor, sorafenib is the first systemic therapy seen to improve survival in advanced HCC patients. It is a bi-aryl urea which targets the serine-threonine (Ser-Thr) kinase Raf-1 inhibiting pathways involved in angiogenesis and cell proliferation (Liu et al., 2006). In advanced clinical trials, treatment with sorafenib compared to placebo showed an extension of median survival time form 7.9 to 10.7 months (Llovet et al., 2008). Continued research is still needed to improve upon this, as well as to counter the high drug toxicity seen in some patient groups (Staufer et al., 2012).

Other novel and adjuvant therapies are being developed in relation to HCC including new treatment regimes applicable whilst patients await liver transplantation (Llovet et al., 2005). With long waiting times for suitable organs, intervention is often required in order to prevent rapid tumour progression to a

point where a transplant may no longer be curative. Living donor liver transplantations (LDLT) have recently been conducted with some success (Zhu et al., 2011) although it requires careful selection of recipients and accurate assessment of the potential risks to the donor (Trotter, 2002).

The commonalities between much of the treatment strategies described above is that they are at best relatively high risk with a requirement for specialist personnel and conditions. The proposed drug therapies for HCC though starting to offer some improvements from earlier suggestion still require extensive development on their life extending impacts. For there to be a significant decrease in mortality from HCC, treatment options applicable to the developing world which shoulders the brunt of the impact of this cancer need to be brought forth for widespread application. Without this being achieved, the impact of early diagnosis will be significantly diminished.

1.7 Proteomics & Mass-Spectrometry in Biomarker Discovery

Proteomics in biomarker discovery focuses on the study of protein signatures expressed in biological samples such as serum, plasma, urine or tissue sections. Changes which occur in molecular signatures measured in these samples are potentially exploitable for use in differentiating the various stages of disease progression. Homing in on key distinctions, whether due to differential expression of selected proteins or the presence of co or Post Translational Modifications (PTM) has lead to the identification of more specific changes on proteins in

disease vs. non-disease states which if developed, could out-perform current non-invasive biomarker assays. MS technologies offer the prospect of being able to study and discern these alterations in protein expression and structure on a high throughput, large-scale quantitative basis. For a disease such as HCC, this prospect is particularly promising and timely.

Accurate protein identification and quantitative methods applicable in the context of Mass Spectrometry are central to reliably carrying out biomarker discovery experiments. There essentially exist two major approaches to this; stable isotope labelling methods such as isobaric tagging for relative and absolute quantitation (iTRAQ) of peptides (Wiese et al., 2007), isotope coded affinity tags (ICAT) (Gygi et al., 1999) or detecting quantitative changes in proteins by incorporating stable isotopically labelled amino acids in cell culture (SILAC) (Ong et al., 2003).

An alternate approach to this is termed "label free" proteomics which is based on the measurement of mass peak intensities either by ionic counts (Neilson et al., 2011) or spectral counting (Trudgian et al., 2011). The key principle for quantitation used in this project is based on integrated measurements of the area under the curve (AUC) of all peptide peaks generated during an LC-MS run (Chelius and Bondarenko, 2002). Ion abundances for selected ionised peptides are measured at specific retention times without the need for spiking or labelling with an isotopic standard in a process referred to as total ion counting (Podwojski et al., 2010). In order to balance the pursuit of maximized protein identifications along

with accurate quantitation, the LCMS^E method was devised, which unlike other scanning techniques that switch between MS and MS/MS modes, uses a single data independent MS run to obtain protein identifications and quantitative data points by alternating between low and high energy collision conditions allowing for all peptides to be fragmented without selecting precursor ions. Running a Waters Q-TOF in MS^E acquisition mode allows for this experiment to be conducted (Silva et al., 2006). Classical proteomics techniques such as 1-Dimensional or 2-Dimensional polyacrylamide gel electrophoresis (PAGE) in addition offer the prospect of being able to discern and study patterns of change brought about by PTMs during disease progression.

1.8 The Diagnostic Problem

The ability to accurately and cost-effectively screen for and diagnose HCC in at risk populations is one of the major challenges faced in efforts to lower its high mortality rate. The greatest burden of liver cancer and its associated conditions is in the developing world where many of the radiological methods used to boost the diagnostic shortfalls of AFP are not readily available. As such, in order to begin to lower the mortality rates from the spectrum of liver diseases, medical researchers need to identify high performance screening and diagnostic markers capable of detecting and or predicting HCC or LC development in populations where these conditions are prevalent.

Key characteristics sought in an ideal biomarker are:

- For the target molecule to be measurable in a non-invasive sample source such as blood or urine.
- For the marker to have excellent diagnostic and or prognostic abilities for condition of interest (i.e. high sensitivity and specificity).
- Marker should be amenable for design into simple kit based assays not requiring additional equipment for analyses or interpretation.
- Assay should be cheap and thus accessible to all the populations requiring it.

An important change that needs to be discernible in order to offer patients maximized treatment options is the onset of hepatocellular carcinogenesis in a cirrhotic liver. Liver cirrhosis is now widely established as a pre-neoplastic condition (Okuda, 2002) and one of the major routes to HCC development in patients chronically infected with HBV. The "gold standard" in clinical HCC diagnosis presently is liver biopsy; but this method is often not preferred as it is expensive, invasive, of some risk and can be difficult to perform in resource-constrained settings. As such, diagnosis is often based on a combination of clinical examination, ultrasonography and AFP measurement. The latter two methods have variable sensitivity and specificity thus not making them reliable stand-alone markers. Ultrasonography (US), although being highly specific; 94% reported by Dodd et al (Dodd et al., 1992) can give variable results depending on the device sophistication, experience of the examiner and the characteristics of the patients' liver disease. In overall screening studies, US has resulted in 35% - 84% sensitivity (Peterson and Baron, 2001) making it a useful but unreliable tool. AFP, the serological marker used in conjunction with US in order to boost performance has a reported sensitivity ranging from 40 – 80% and specificity of 76 – 97% (Nguyen et

al., 2002). The performance of AFP is also dependent on whether it is used for screening or differential diagnosis.

All of these shortfalls mean that a combination of radiological and blood based assessments must be utilised for accurate non-invasive diagnosis of HCC. This is particularly problematic in the developing world due to its costs, required expertise and equipment. What would be most ideal is the proposition of a single blood test based on a single or panel of proteins able to discern early carcinogenesis in the liver, particularly in patients with backgrounds of fibrosis or LC.

1.9 Proposed HCC Biomarkers in Literature

The HCC biomarker discovery field is vast with new and on-going research continuing to propose a fast growing list of biomarker candidates, which pending further validation may be used in conjunction with AFP as the primary choice of non-invasive test for HCC determination. Biomarkers detectable in accessible body fluids such as blood, urine and saliva and to a lesser degree, tissue, have the potential to act as accurate, non-invasive indicators of the presence or progression of disease. With its virtually equal incidence and mortality rates (Ferlay et al., 2010; Parkin et al., 2005), HCC is one particular cancer that can benefit from the identification of better diagnostic markers capable of detecting primary liver tumorigenesis early enough that successful intervention could be administered.

An area of particular focus in terms of identifying unique exploitable changes for differential diagnosis are those which occur in the glycosylation of many proteins during CLD progression (Blomme et al., 2009). As Asparagine-linked (N-linked) glycosylation is the predominant form found in human sera, much research has been focussed on this particular subclass of glycoproteins. Investigations looking at the glycosylation patterns of key serum proteins associated with HCC and LC reports 19 of them as showing evidence of hyper-fucosylation (Comunale et al., 2006) in liver disease. Wang et al in a study published in 2009 compared fucosylated forms of serum kininogen and alpha-1-antitrypsin to a i) well established and ii) recently proposed marker; AFP and Golgi Protein 73 (GP73), respectively and showed a significant boost in diagnostic accuracy when differentiating between patients with non-cirrhotic HCC and those with HCC in cirrhotic backgrounds (Wang et al., 2009b). GP73 after being established in a mid sized subject population as a potential HCC marker with better performance than AFP (Marrero et al., 2005), was preliminarily investigated in the sera of a small cohort of patients previously treated for HCC. The protein showed significant potential in predicting recurrence within this pilot group (Hann et al., 2010).

The focus on alterations in glycosylation patterns of various proteins in the context of HCC biomarker discovery has continued with the proposal of some well-studied trends and changes in specifically highlighted proteins. In efforts to determine the source of the hyperfucosylation changes previously linked to liver disease, an investigation was conducted using human tissue sections sourced

from three distinct groups; HCC tumours, its adjacent cancer free sites and adults with no evidence of liver disease. Comparisons of these three groups showed that no key differences were observable in the overall levels of fucosylation between them, suggesting that hyperfucosylation changes previously reported from subject sera may not be due to up-regulation of fucosylated proteins in liver tissue. Instead, what was observed was that tetra-antennary glycan structures were notably increased in HCC tissues in comparison to all the other groups (Mehta et al., 2012).

Comunale and colleagues embarked on a study focussed on elucidating previous reports suggesting the over expression of serum clusterin, also known as apolipoprotein J (Apo J) in HCC, and its potential as a clinical indicator of liver carcinogenesis (Comunale et al., 2009; Kang et al., 2004). To assess the impact of this change, the levels of Datura stramonium lectin (DSL) enriched clusterin were measured in an independent cohort of 151 subjects classed as HCC, LC and HCV positive. This lectin which binds (β 1,4) triantenarry N-linked glycan structures allowed for the accurate distinction of subjects diagnosed with HCC from those with LC with an AUC of 0.852 (Comunale et al., 2011). When this enriched clusterin fraction was combined with AFP and the previously proposed HCC marker, fucosylated kininogen (Wang et al., 2009b) in multivariate logistical regression analyses, the level of diagnostic accuracy as indicated by AUC statistics increased to 0.944, much higher than that achievable with AFP. This suggests that the specific expression patterns of these tri- and tetra-antennary structures vary between protein groups. However, what could be exploitable is the presence of

distinct signatures that are found to be consistent across multiple subject populations.

Serum clusterin has been further associated with the diagnosis of HCC as well as being a possible indicator of the metastatic potential of primary tumours (Lau et al., 2006). Results tabulated from our label free discovery experiments shortlisted this protein as consistently, significantly differentially expressed in LC and HCC (table 3.2). Research from Egypt, which has some of the highest incidences of hepatitis infections worldwide looked at the potential of clusterin to evaluate the diagnostic and metastatic potential of virus related HCC. ELISA measurement of serum clusterin expression in HCC, LC, chronic hepatitis and healthy controls highlighted that the protein was markedly elevated in patients with poorly differentiated tumours, as well as in subjects who had some degree of hepatic portal vein or lymph node invasion. The performance of this putative marker was superior to AFP, offering 90% sensitivity and 87% specificity when used at a cut off of 128µg/ml (Nafee et al., 2012).

The current mainstream non-invasive clinical tool, AFP, has also been the subject of focus in the search for better performing HCC diagnostic and or prognostic tests. Total AFP is a combination of three glycoforms; AFP-L1, AFP-L2 and AFP-L3 which vary according to their *Lens culinaris* agglutinin (LCA) lectin binding capacities. AFP-L3 in particular has been well documented as a more sensitive and specific marker for early HCC, particularly in patients with a background of LC (Taketa et al., 1993 and Kamoto et al., 2002). High expression of this marker has

also been associated with large tumour mass, poor differentiation, declining liver function and malignant characteristics (Khien et al., 2001).

Targeted proteomic analyses using Multiple Reaction Monitoring (MRM) MS coupled with enrichment with the fucose specific aleuria aurantia lectin (AAL) was conducted by a Korean group on ten HCC cases compared to ten subjects each with the profiles; healthy controls, HBV positive and LC cases. Average peak abundances were compared between the disease and control groups with the greatest fold changes observed for alpha-1-antichymotrypsin (AACT), alpha-1-acid glycoprotein 1 (A1AG1), α1AT – which has also been implicated in its fucosylated form (Comunale et al., 2010), ceruloplasmic (CERU) and CC3 (Ahn et al., 2012). Interestingly, two of these listed proteins were identified in our study with uniform expression trends from control to HCC.

The protein-immunoglobulin complex, Squamous Cell Carcinoma Antigen-Immunoglobulin M (SCCA-IgM) has been further characterized in the context of HCC development after it was seen in a study to be elevated in all surgically resected HCC tumours. Follow up research on its expression in patient sera further validated this finding with up to 96% of HCC patients testing positive for at least one marker, SCCA-IgM or AFP when using a cut off of >20ng/ml for the latter (Pontisso et al., 2004).

Novel metabolomics approaches and Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) MS based analyses being applied to the field of HCC

biomarker discovery have yielded significant results. Using MALDI-TOF MS, HCC tumour tissue and non-cancerous adjacent sections identified peroxiredoxin 3 (PRDX3) at both the mRNA and protein levels to be over expressed in HCC's and correlated to poor cell differentiation (Qiao et al., 2012b). Metabolite based discovery has been attempted using a Ultra performance LC (UPLC^{TM})-QTOF MS, looking at metabolic profiles in the sera of LC and HCC subjects. Metabolites involved in sphingolipid metabolism and phospholipid catabolism were observed as up-regulated whilst down-regulated metabolites included those involved in cholesterol metabolism (Ressom et al., 2012). These pathways warrant further investigations to see how they may accelerate or contribute to the process of hepatocarcinogenesis.

The importance of independent validation of proposed biomarker candidates for any disease context cannot be over stated. With the plethora of protein, molecular and histological biomarkers being proposed for HCC, the focus of researchers must now shift to the independent measurement and validation of these candidates, utilizing methods independent of the applied discovery techniques and as well as distinct subject populations. Shen and colleagues attempted this in a mid-sized cohort as follow up validation of previous work suggesting Dickkopf-1 (DKK1), a secretory protein involved in embryonic development known to inhibit the Wnt pathway, as over expressed in HCC tissue compared to non-cancerous surroundings. ELISA measurement of DKK1 showed great diagnostic potential in differentiating HCC and controls particularly when used in combination with AFP. AUC scores for its performance were 0.89 (95% CI:

0.87-0.91) in the test cohort and 0.88 (95%CI: 0.86-0.92) in the validation cohort (Shen et al., 2012).

In the investigation of discriminatory markers for HCC development, some researches have chosen to focus on earlier stages of the sequelae of changes leading to CLD by attempting to home in on molecular indicators of fibrosis development. These markers may represent the earliest detectable alterations occurring in blood or tissue as a result of commencing liver disease. Some of the major proteins detected with distinct signatures at this early stage include apolipoprotein L1 (Apo L1) and complement factor H-related protein 1 (Gangadharan et al., 2007). More recently the beta chains of CC3 and complement 4 (C4) (Gangadharan et al., 2011) as well as clusterin have been implicated – all through 2-dimensional electrophoresis (2DE) analyses of serum or plasma (Gangadharan et al., 2012). Other proposed HCC biomarkers have been comprehensively reviewed in this recent publication (Behne and Copur, 2012).

It is clear from the multitude of biomarker candidates proposed from proteomic analyses of blood and tissue samples from various HCC related disease groups that there exist noteworthy changes occurring at the protein level as a result of disease aetiology and progression. In order to move these potential disease specific signatures forward and accelerate their established usage in clinics and the field, particularly in the developing world, the momentum of research must now shift from discovery to validation based experiments – focused both on expression level and mechanistic changes between clinically distinct groups. This

shift will require widespread collaborations between researchers in this field, and if successful could result in the identification of a series of biomarkers usable for assessing liver cancer prognosis, diagnosis and progression.

1.10 Aims of Thesis

The challenges described above and those faced by many clinicians and researchers in the developing world led to the design and funding of this project. Our primary aim was to use plasma collected from subjects based in HCC endemic areas with viral hepatitis and mycotoxin exposure as the primary aetiological agents and carry out discovery based proteomics using MS. The key findings from this "discovery" run were then subjected to validation experiments using independent methodologies and subject cohorts.

The stages of this investigation have been divided into chapters, all of which aim to give a clear and concise picture of the aims, background, methods used, key results and major conclusions drawn from our multi-faceted investigations. This chapter constitutes the main introduction for this thesis, aiming to offer relevant background to the subject of HCC biomarker discovery and its significance to the developing world. The subsequent four chapters all detail the key findings from this project with chapter two focusing on the initial pilot studies conducted using MALDI-TOF MS based discovery. Chapter three develops on some of the shortfalls faced with the previous methodology and reports on the experiments carried out on the main GLCS population using label-free quantitative MS. Chapter

four details the development of a subset of previously identified differentially regulated proteins; setting out to validate them using an independent population and methods. The concluding data chapter focuses on applying statistical methods of biomarker combination on four key proteins versus a routine clinical indicator of liver health, to see if any of our proposed candidates offer improved performance.

The final discussion and concluding chapters are used to wrap up this thesis, highlighting its major findings as well as the key limitations faced.

Chapter 2: Pilot Protein Profiling of HCC Subjects by MALDI-TOF MS

2.0 Introduction

The discovery of an accurate non-invasive, affordable diagnostic test for liver cancer useable in the developing world is the current focus of a significant amount of research in the field of hepatology. This project sets out with similar goals and at its inception started with a pilot investigation utilizing MALDI-TOF MS to study the protein expression profiles of fractionated plasma from a small group of subjects within the GLCS. The widespread use of MS as a clinical proteomic tool only came about over the last decade. Prior to the advent of high throughput and complex mass spectrometers; the main methods available for comparative analyses of human bio-fluid samples at the protein level were via 1 and 2-Dimentional Polyacrylamide Gel Electrophoresis. These low throughput methods are still vital in enabling researchers to compare global variations in the protein expression profiles of individual and pooled samples. Some of their key drawbacks however are that they offer limited quantitative data and for experiments probing particular proteins are still heavily reliant on the availability of suitable antibodies.

To circumvent some of the bottlenecks faced using gel based methodologies, multi platform mass spectrometers began to be applied to the identification and analyses of expression proteomics (Pusch et al., 2003). MS now allows researchers to perform protein and peptide analyses on bio-fluids from large

populations and compare obtained mass-to-charge (m/z) based spectra using various bioinformatics tools. The applicability of this workflow is vast and amenable for use in the field of liver cancer; more specifically in efforts to identify a single or combination of key plasma proteins which have their secretory patterns altered significantly by the presence of HCC.

For our pilot analyses, MALDI-TOF MS based discovery was first attempted as it did not require extensive sample preparation, allowed for individual subject based runs and permitted the export of raw data for database searches and statistical analyses using specialized software programs. The kit based enrichment methods developed by Bruker Daltonics ClinProt™ were complementary to plasma biomarker discovery; being particularly valuable in the enrichment of low molecular weight proteins and peptides of potential discriminatory value. MALDI MS uses a soft ionization approach to analyse biomolecules enriched by magnetic resin designed for specific interaction mechanisms such as reverse phase, cation or anion exchange, aiding in the efficient enrichment of proteins within biological samples (Whiteaker et al., 2007). The enriched plasma proteins react with the UV-absorbing carrier matrix and become protonated, often with a single positive charge. When these singly charged molecules enter the high vacuum of the TOF tube, the analyte proteins are irradiated with a pulsing laser and explosively released toward the analyser at a velocity proportional to their respective sizes (laser induced desorption). Targeted software, bioinformatic and statistical packages respectively allow for the visualization of spectra, peak analyses and application of statistical methods to trend peak transitions in individual samples

and disease groups allowing for the determination of key signatures amenable to biomarker discovery.

The major objectives of this study as summarized in the previous chapter were envisaged in the Gambia where HCC like in much of the developing world is a major cause of morbidity and mortality in males and females (Bah et al., 2001). Global cancer statistics highlight that eighty per cent of global HCC's occur in the developing world (Ferlay et al., 2010) where resources are constrained and access to tools that allow for early diagnosis limited or non-existent. Much work has been carried out in the Gambia on HCC, liver disease and its viral and environmental causes (section 1.2). Briefly, from decades of research targeting ways of reducing the impact of HCC and associated liver conditions on populations - one of the key bottlenecks identified was the lack of adequate non-invasive tools for early diagnosis. Whereas in developed nations, imaging technologies have been brought in to boost and in some cases replace the shortfalls of AFP, much of the developing world is still reliant on the marker despite its limited ability to determine early hepatocarcinogenesis (Sherman, 2001). Previous work has shown that in the critical stage at which potentially curative intervention is still viable, AFP is an inadequate indicator of disease (Sherman, 2001). The 2010 guidelines for HCC surveillance and diagnosis published by the American Association for the Study of Liver Diseases (AASLD) concludes that AFP with regard to HCC "lacks adequate sensitivity and specificity for effective surveillance and for diagnosis" (Bruix and Sherman, 2011) and recommends surveillance of high risk groups using bi-annual ultrasounds. In the

Gambia, observations from healthcare facilities show that most HCC's are diagnosed at an advanced stage by which point little more can be offered apart from palliative care.

A sizeable body of research exists utilizing MALDI-TOF MS methods to probe human sera for signatures able to differentiate HCC patients from non-liver disease controls as well as those with other CLD (Liu et al., 2011; Poon et al., 2003; Qiao et al., 2012a). The suggestions in literature for proteins that are differentially regulated and or expressed in this spectrum of conditions are vast. As such, one of the key aims of this project has been to take proposed markers forth for validation using independent methods and cohorts. This, we feel will be a more robust contribution to the field of biomarker research in liver diseases and offer curated candidates a stronger base from which they could be taken forward for extensive validation and characterization.

2.1 Materials & Methods

Care Weak Cation Exchange (WCX) Kit purchased from Bruker Daltonics (BD), Billerica, USA; α-Cyano-4-hydroxycinnamic acid and EtOH were purchased from SIGMA–ALDRICH; Acetone was purchased from Fisher Scientific.

2.1.1 Gambia Liver Cancer Study Pilot Samples

A subgroup of forty plasma samples consisting of 13 controls, 12 LC's and 15 HCC's were selected at the initial phase of this project. These were chosen for pilot MALDI-TOF MS based analysis at the onset of this project in order to:

a) Check for sample viability following, storage, transport and processing.
b) Provide an initial snapshot of whether any interesting differences were observable between the three groups.

The full details of how these subjects were recruited and diagnosed are described in the main discovery and validation chapters in sections 3.1.2, and 4.1.2.1, however a summary of the recruitment process and case definitions are provided in figure 2.1.

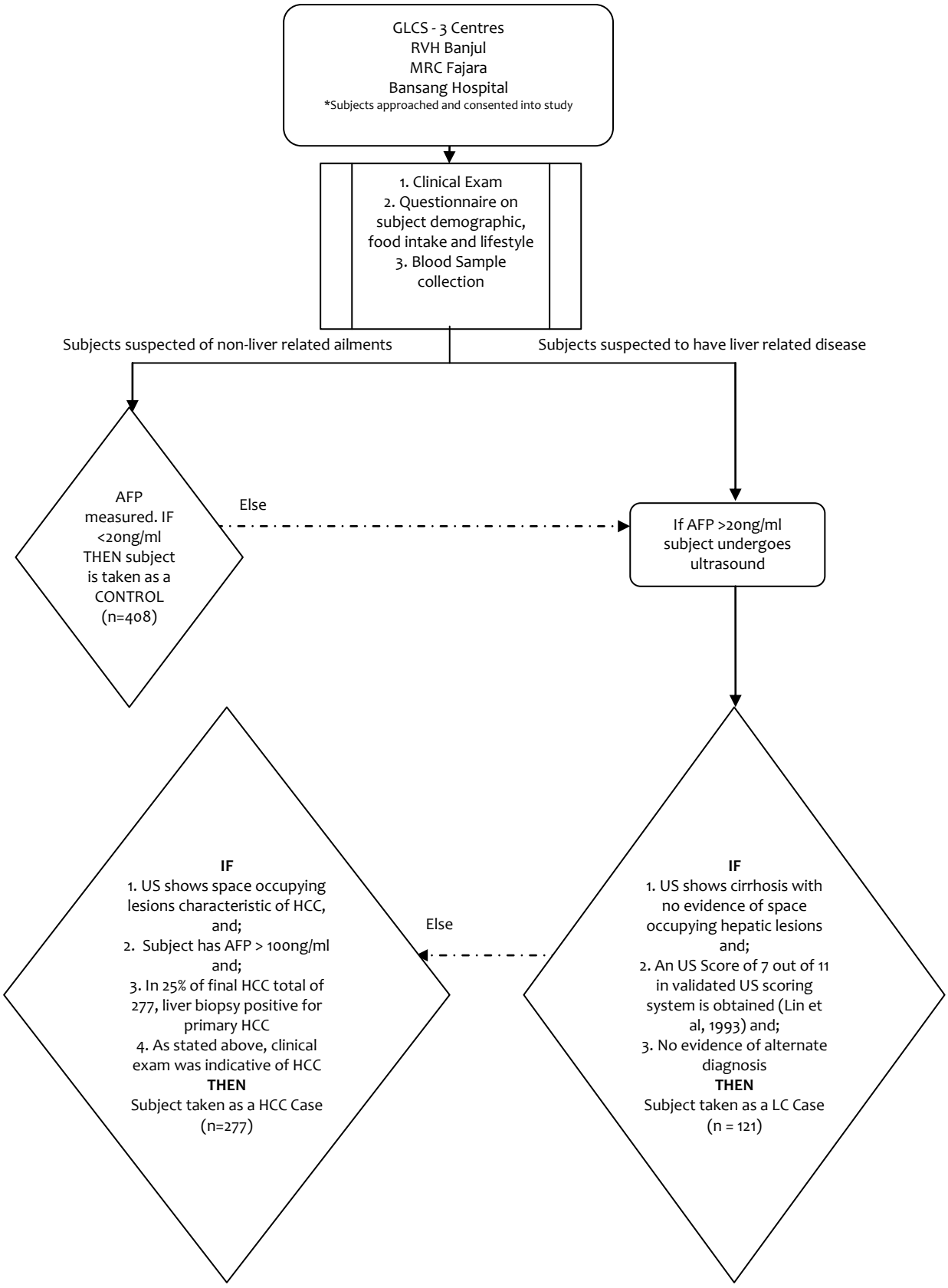


Figure 2.1: Scheme summarizing how GLCS subjects were categorised into the three clinical groups following informed consent.

2.1.2 Sample Size Determination

Prior to the application for full MRC scientific steering & ethics committee approval of this study; pilot data generated from the MALDI-TOF MS based analyses was used to determine a viable sample size on which subsequent discovery or validation experiments conducted could have enough statistical power to yield significant results. This approximation was based on results from MALDI spectra exported using the Flex Analysis software (version 2.4, BD). The full spectrum of peaks exported from the 40-subject pilot suggested a standard deviation for the receiver operating characteristic (ROC) curves of around 0.5 for most protein or peptide masses. AUC, in this context, is a measure of the overall ability of a protein peak to be used to diagnose HCC or LC. Peaks with masses less than 3500 m/z with at least one sample where it was measured at an intensity >80 arbitrary units (a.u) were prioritized as they offered greatest likelihood of successful sequencing. It was projected that around 50 peaks would be found in this range. For comparison, we theorized that AFP as a diagnostic tool for HCC has an AUC of approximately 0.7. Thus, a sample size of 100 cases + 100 controls would give us 80% power to detect peaks as significantly better than 0.7 if the AUC value is 0.85 or greater (one-sided $p<0.0005$, equivalent to a one-sided test at $p<0.025$ adjusted for 50 multiple comparisons). From the original pilot samples, only the standard deviation of the ROC curves was used to determine the design of the study, thus serving as an internal pilot, i.e. included within the 120 HCC & control and 100 LC samples required for the next round of analyses, in order to not further diminish the sample numbers available. As they were available, an excess of the minimum required sample numbers were taken in the control and

HCC groups. All available LC cases were taken for follow up experiments, these numbered 99.

2.1.3 Weak Cation Exchange Preparation

Magnetic bead based WCX fractionation kits were used in sample preparation according to manufacturer's instructions for the analysis of endogenous peptides in the 1000 – 10,000 Da molecular weight fraction of plasma. Briefly, 5ul of subject plasma were incubated with a binding buffer and WCX magnetic bead mix. Using a magnetic separator, the beads were collected on thin walled micro-tubes and the supernatant removed. Three washing steps were then carried out on the sample bound beads using the wash buffer provided. Using the kit elution buffer, the beads were again mixed thoroughly and collected at the walls with a magnetic separator. The resultant clear supernatant was aspirated and transferred to a fresh tube and frozen at -20°C pending analysis.

2.1.4 MALDI-TOF Analyses & Peak Export

On the day of MALDI-TOF analysis, 1ul of the eluted peptide solution was mixed with 9ul of freshly prepared matrix solution (0.3g/l α-Cyano-4-hydroxycinnamic acid in a 2:1 solution of EtOH and acetone). 1µl each of this mixture was spotted in quadruplicate to an approximate diameter of 600µm on a clean AnchorChip™ target plate, allowing it to air dry prior to insertion into the instrument. The samples were run in linear mode at 80% laser intensity resulting in a total 1200 shots per spot. Spot calibrations were performed for each location to ensure

uniform sampling starting from the centre, moving outwards. Calibration of the instrument was carried out using supplied standards ranging from 1 to 12 kilodaltons (kDa). Peptide sequencing was performed in reflectron mode as described in this publication (Goldman et al., 2007). Samples were chosen for peptide sequencing when seen to hold protein fragments of interest at a sufficiently high intensity (at least >80 a.u).

Spectra importing and viewing from the MALDI instrument was performed using the BD software program Flex Analysis (version 2.4). For further bioinformatics analysis, the ClinPro Tools™ (version 2.0) software, also from BD was used for spectra normalization, baseline subtraction as well as analysis using the support vector machine (SVM) algorithm. This method evaluates protein fragments, comparing changes in intensity across sample groups, highlighting those which best differentiate between the various disease states.

2.1.5 Peak List Analysis by STATA

To determine the diagnostic ability of expressed peptides, the statistical package STATA (version 9) was used to calculate AUC statistics for differentially expressed peaks with significant p values ($p<0.05$), see tables 2.1 and 2.2. This was done for the comparisons Control vs. HCC and HCC vs. LC.

The peak lists obtained from the completed MALDI-TOF runs were exported for analysis from ClinPro Tools (version 2.0) to the statistical software. Using STATA,

ROC curves were plotted for peak m/z's which were seen to be either significantly up or down-regulated between controls and either of the two disease groups. ROC curves present the relationship between sensitivity and specificity over all possible cut-offs and when combined with AUC values offer an insightful measure of test performance (Zweig and Campbell, 1993)

2.2 Results

2.2.1 Changes in peak mass intensity between distinct subject groups suggest identification of fragments discriminatory toward liver disease.

This section displays the results of a series of approaches used to demonstrate differences in the expression of WCX fractionated low molecular weight plasma protein fragments, detected in the 40-subject GLCS pilot. Figure 2.2 shows the overall normalized and background subtracted spectra from the 13 control and 15 HCC subjects. The general trend observable is that there are multiple mass regions at which noticeable differences in peak mass intensities occur; the most prominent of these are at approximately 2660 m/z where the average peak intensity is much higher in the controls than HCC's and at approximately 2560 m/z where the opposite is true. Figure 2.3 shows results on a two-dimensional plane following the application of a SVM model generation algorithm. It highlights two peak masses; 2761 m/z and 2776 m/z found to best classify subjects into their three distinct diagnostic groups. A definite separation is seen between the control versus LC and HCC groups, however, using these two markers, the LC and HCC have a large degree of overlap which is not optimal. When the same analysis is

repeated in the absence of the LC sub-group (figure 2.4), distinct separation between the HCC and control clusters is seen with the selection of novel markers, suggesting that the combined peptide masses, 1330 m/z and 2445 m/z, offer enhanced distinction of disease (HCC) from liver disease free controls.

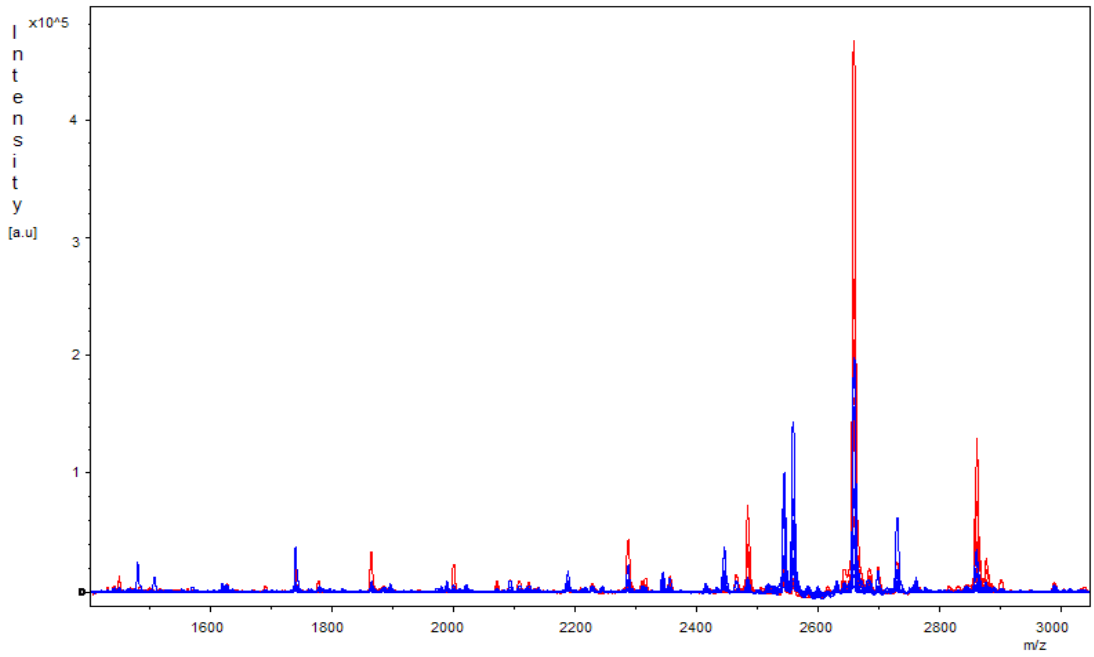


Figure 2.2: Figure showing the overall average spectra (up to 3000 m/z) of HCC (blue) and Control (red) patients as obtained from the pilot study. It shows several peak masses to be up/down regulated between the two groups. The fragment intensity, measured in arbitrary units is plotted on the y axis and fragment mass to charge ratios on the x axis.

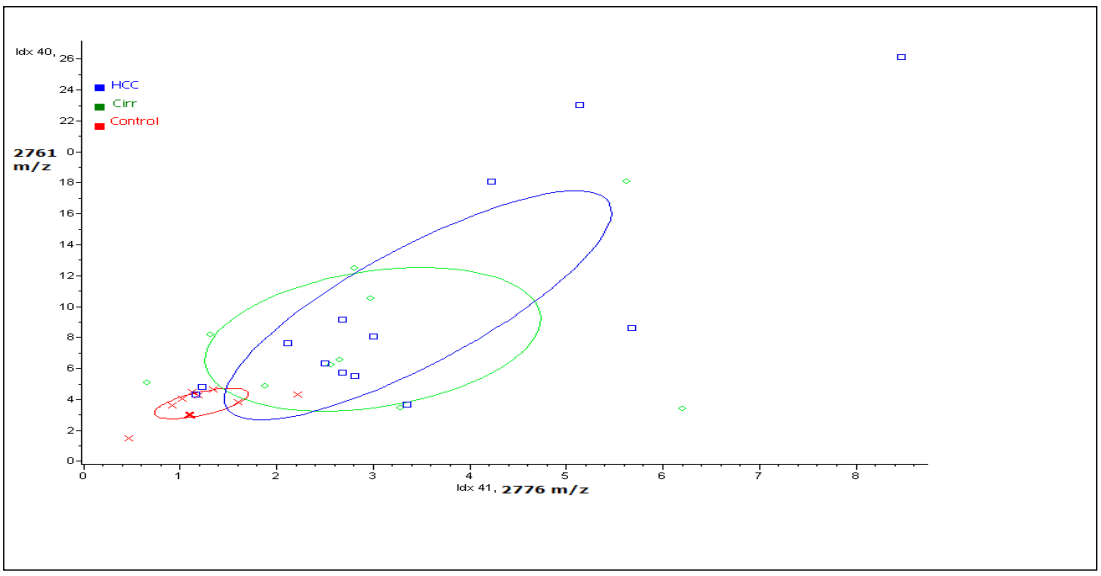


Figure 2.3: Analysis using model generation methods to identify two peak masses, which, following WCX based fractionation, separate best the three subject groups. The relative abundance of the peak masses 2776 m/z and 2761 m/z are plotted on the x and y axes respectively and individual subject abundances indicated in order to view separation patterns. Good separation is seen between the two disease groups and Controls (red) but there is a large degree of overlap between the Cirrhosis (green) and HCC (blue) patients.

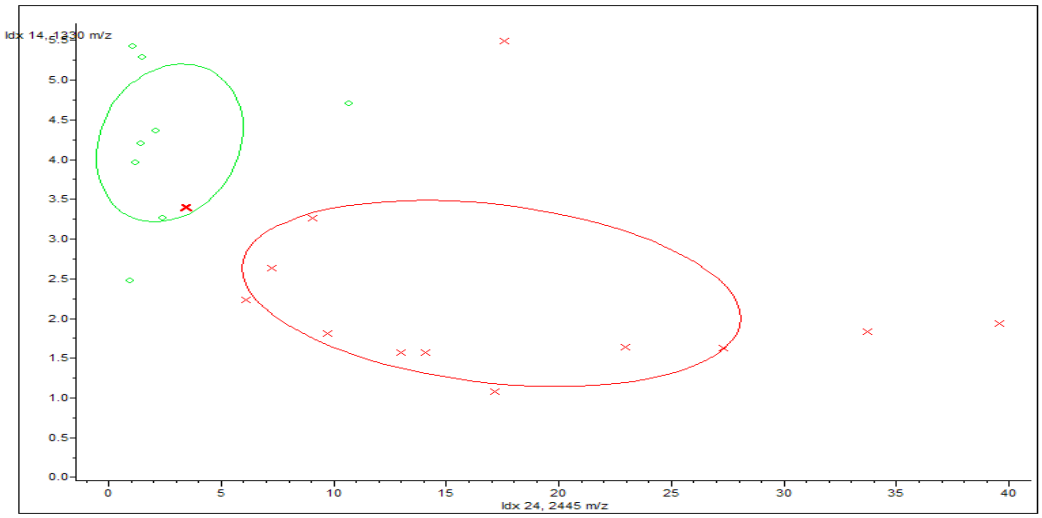


Figure 2.4: Same two dimensional viewing method used on the two groups showing more distinct separation of HCC (red) and Control (green) groups with peak masses 1230 m/z and 2445 m/z.

2.2.2 Statistical analyses of peak mass distributions identify protein fragment masses with high diagnostic potential for LC and HCC.

Following statistical analyses generating AUC, 95% Confidence Interval (CI) and p-value statistics, multiple peaks were seen to be significantly differentially expressed with three particular peptide fragments; 2444 m/z, 2583 m/z and 2559 m/z (tables 2.1 & 2.2), appearing to be diagnostic for both LC and HCC. AUC values of as high as 0.952 were attained for peaks that differentiated between controls and HCC's, whilst the highest value attained in the two less discernible groups was 0.880 (figure 2.5). Dot-plots of the trend in intensity between the three groups also show the recurring pattern of distinct separation of control and HCC patients, with those with LC falling between the two extremes (figure 2.6).

Mass (CON vs. HCC)	Area Under Curve (95% CI)	p
2444.98	0.952 (0.852 -1.000)	0.0000754
2583.83	0.894 (0.685 – 1.000)	0.0004595
2517.38	0.942 (0.824 – 1.000)	0.0008036
3139.95	0.933 (0.812 -1.000)	0.0008802
2559.85	0.894 (0.685 -1.000)	0.0012593

Table 2.1: List of peak masses, AUC's and respective p values (p) associated with their ability to distinguish between Control and HCC patients using MALDI-TOF MS based profiling.

Mass (HCC v.s. LC)	Area Under Curve (95% CI)	P – Value
2583.89	0.880 (0.731 – 1.000)	0.0029623
2444.97	0.795 (0.604 – 0.986)	0.0212315
2559.86	0.752 (0.509 – 0.995)	0.0488438

Table 2.2: List of peak masses, AUC's and respective p values associated with their ability to distinguish between HCC cases and Cirrhosis patients using MALDI-TOF MS based profiling.

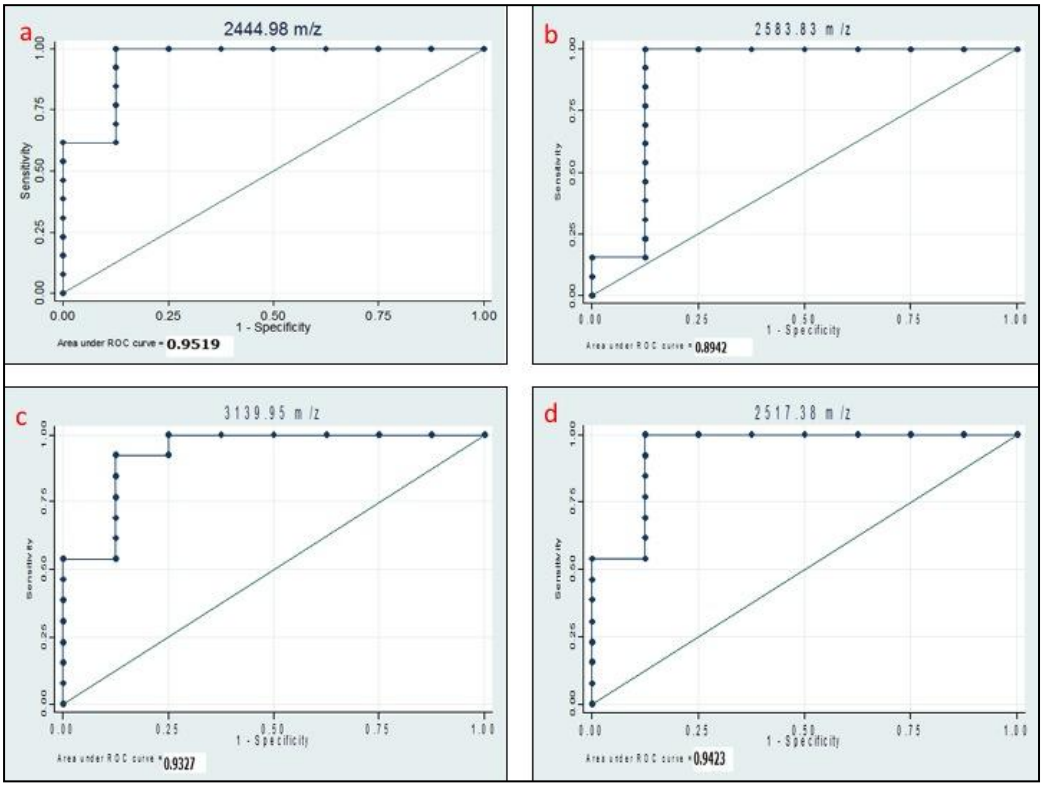


Figure 2.5 a-d: ROC curves of peak masses found by MALDI-TOF MS to have greatest diagnostic ability. Some appear have the potential to be diagnostic for both liver cirrhosis and HCC (2444 m/z and 2583 m/z).

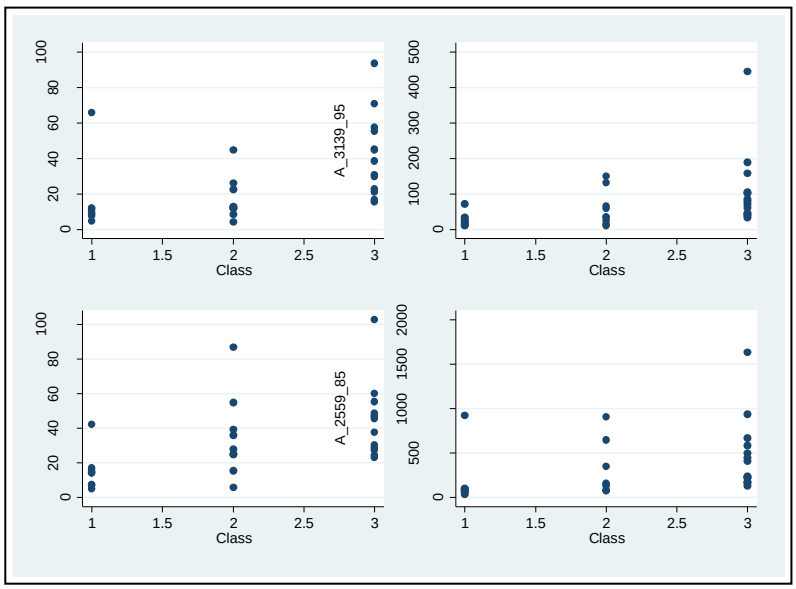


Figure 2.6: Peak intensity comparisons illustrated by dot plots showing trends in the expression of 4 peptide masses found to be significantly differentially expressed in comparisons between at least two of the three patient groups under study. Class 1 = Controls, Class 2 = LC, Class 3 = HCC. The y axis shows the intensity in arbitrary units of the labelled peptide.

2.2.3 Protein fragment highlighted in statistical analyses as discriminatory between three subject groups identified as Fibrinogen – alpha chain.

Protein ID's have not been attained for most of the peptide fragments mentioned in the MALDI-TOF MS result section as attempts to sequence discriminating peaks were largely unsuccessful. To date, only the peptide of mass 2559 m/z (figure 2.7) has been successfully sequenced and identified using the MASCOT (version 2.3, Matrix Science) software, searched against the SwissProt Database.

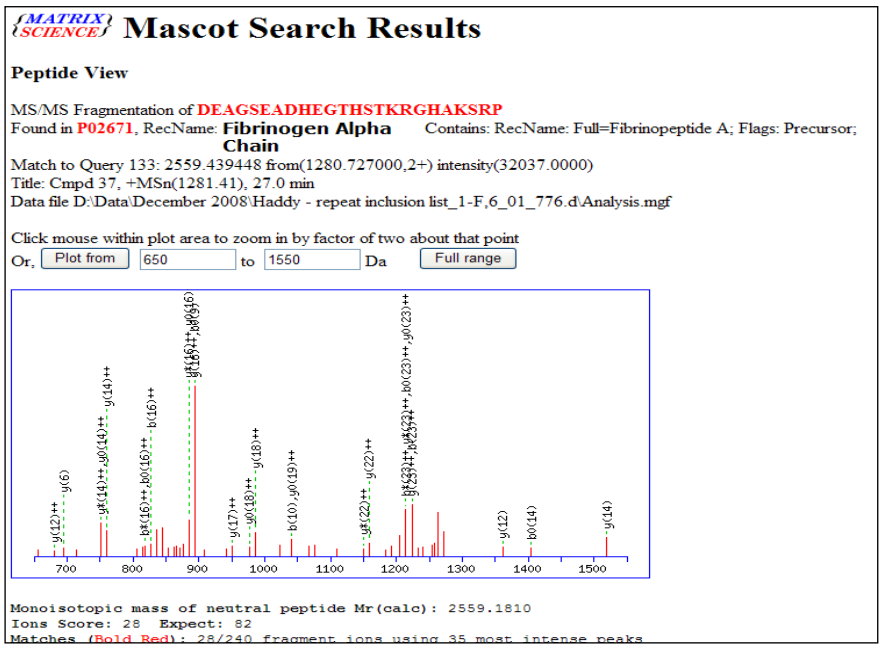


Figure 2.7: MASCOT search result and MS/MS spectra for Protein ID of fragment 2559 m/z found to be a potential marker for HCC and LC. The generated fragments from tandem MS analysis are plotted with annotations for respective b and y fragment residues and charge states.

2.3 Discussion

Major advances in the utilization of MS in the study of proteins set off when Karas, Hillenkamp and colleagues in 1988 discovered that laser application could cause large protein molecules to ionize without significant fragmentation when complexed with an appropriate holding matrix (Karas and Hillenkamp, 1988). The direct pathways via which this occurs have not been well described. However, the primary mechanisms proposed are based on photochemical ionization (Vertes et al., 1993), cluster ionization (Karas et al., 2000) and more recently a pseudo-proton transfer model (Chang et al., 2007). The rapid development of MS from this point, to a method of robust protein and peptide identification now enables us to compare protein expression spectra on a large number of samples, and

using various bioinformatics tools, analyse their results to pinpoint full proteins or peptide fractions which facilitate the optimal detection of the presence or progression of disease.

The primary drawback of traditional 1 or 2DE methods, particularly in their application to multi-subject, plasma based biomarker discovery is that they are low throughput, often poor at distinguishing proteins with extremes in molecular weight and do not directly offer quantitative data on protein expression. 2DE in particular does however enable researchers to compare qualitative variations in protein expression at high-resolution (O'Farrell, 1975; Schwartz and Cantor, 1984) and can thus serve as a valuable tool in the detailed study of some of the biochemical characteristics of proteins in disease and non-disease states. MALDI-TOF MS based protein biomarker discovery in opposition to this carries some key advantages; firstly it allows for individual subject based analyses, does not require extensive sample preparatory steps and allows for MS data on peak masses and intensities to be exported for database searches and analyses. Our focus on the enriched low molecular weight region of the plasma proteome also aids the identification of high-performing protein fragments. Despite their discovery however, the major bottleneck of this method is the ability to successfully sequence and identify the protein ID's of highlighted fragments.

HCC development is a progressive condition that often occurs from a background of liver cirrhosis. In areas where HBV is endemic, patients with cirrhotic livers have a 15% 5-year cumulative risk for HCC development, as opposed to a 10% risk in

Western countries (Fattovich et al., 2004). As such, a key focus in biomarker discovery has become the identification of proteins or peptide fragments which can discern early HCC from LC and other CLD's. The results obtained from comparing the MALDI-TOF profiles of fractionated plasma of control, LC and HCC subjects highlight five protein fragments which significantly discriminate ($p<0.05$) between controls and at least one of the two disease groups. Three of these (2444 m/z, 2559 m/z and 2583 m/z) are of particular interest as they have been shown to significantly differentiate between controls and both disease states at sensitivity and specificity of >75%. Such candidates if identified and validated, could significantly improve upon the levels of diagnostic accuracy attainable with sole AFP measurement. Attempts to assign these putative markers to known proteins resulted in the identification of peak mass 2559 m/z, which following sequencing using an independent LC-MS/MS system (HCTplus^{TM} BD) (Batycka et al., 2006), was successfully identified as Fibrinogen Alpha (FGA). No reports have been published to date proposing FGA as a diagnostic tool for HCC or LC. The gamma chain of the same protein, Fibrinogen Gamma (FGG), has however been reported at the mRNA level to be significantly higher in HCC tissues compared to adjacent non-cancerous sites. The same trend was observed at the plasma protein level where FGG expression was found to be higher in HCC versus healthy controls (Zhu et al., 2009).

One of the other key objectives of running a pilot set of experiments was to check for the viability of the GLCS subject plasma following long-term storage. As no plans were finalized during the subject recruitment period for proteomics based

biomarker studies, all sample collections were conducted using standard clinical protocols following good laboratory practice. All the indications from careful scrutiny of our results and comparison with previous and concurrent results from lab colleagues were that the samples were in comparable form and usable for proteomic analyses (El-Kasti et al., 2011; Sideso et al., 2010). The observed signal intensities of the fragments were sufficiently high with low overall signal to noise ratios. Following bioinformatics analyses, the pilot results were also indicative of significant disease specific differences worthy of further study; as such, a novel approach based on label free proteomics was targeted as the main method to be utilised on the selected full subject population.

These results, although based on a small group of subjects suggest that there are condition specific signatures identifiable at this level worthy of further evaluation. Due to the difficulties faced in assigning peptide fragments to known proteins through database searches; MALDI-TOF MS based discovery was not applied to the full discovery subject population as efforts to validate differentially expressed proteins could be significantly hampered by not being able to confirm the full identities of highlighted peak masses.

2.4 Conclusions

The pilot MALDI-TOF MS biomarker discovery experiments and analyses highlighted that disease specific protein signatures were detectable in this subject population and could be exploited to home in on protein candidates for

subsequent validation. The quality of the plasma samples being used was also ascertained and found to be suitable for further proteomic based analyses. To counter the difficulties faced with protein identification, a label free platform offering direct sequencing and data on quantitative expression was opted for going forward, thus forming the core approach from which well selected biomarker candidates were proposed.

Chapter 3: Protein Profiling of Hepatocellular Carcinoma and Liver Cirrhosis by Label-Free Quantitative Proteomics in West African Subjects.

3.0 Introduction

The poor prognosis associated with a diagnosis of HCC, is a significant contributor to its ranking as the overall third highest cause of cancer related mortality worldwide (Parkin, 2001). The impact of the lack of a low cost, reliable diagnostic and screening test, useable in the developing world where access to high performance compensatory diagnostic aids are severely limited can be directly linked to poor prognosis and survival for HCC. As previously detailed, the main objective of this study was to make significant contributions towards alleviating this problem by conducting a vast and comprehensive scan of subject plasma collected from a West African population for differentially expressed proteins potentially exploitable in biomarker applications. Thus, following the MALDI-TOF MS based pilot profiling of plasma proteins and the collation of a full set of 339 well characterised plasma samples from the GLCS, a novel approach for comprehensive, quantitative proteomic analyses was adapted. This new platform, termed label free MS^E was applied to the comprehensive profiling of plasma proteins in pooled subject groups of healthy controls, LC and HCC.

The principle of applying MS technologies to the field of biomarker discovery in bio-fluids has significantly aroused the attention of researchers working to

continually improve contemporary methods of analyses. The biggest bottle-necks faced in this effort are in improving instrument sensitivity, allowing for higher throughput analyses combined with maximized dynamic range coverage and accurate comparative quantitation of identified proteins. Label free quantitative "shotgun proteomics" has arguably been the most successful alternative proposed to labelling methods and has comparatively simplified sample processing workflows with greater cost-effectiveness and linearity at both the peptide and protein levels (Bantscheff et al., 2007). The applicability of statistical analyses to this approach further encourages the discovery of small but significant changes in protein and peptide expression patterns that can be followed up by lower throughput traditional proteomic and immuno-assay based methods.

The key quantitative principle of the label-free quantitative MS^E approach used in this study is based on integrated measurements of the AUC of all peptide peaks generated during an LC-MS run (Chelius and Bondarenko, 2002). Ion abundances for selected ionised peptides are measured at specific retention times without the need for spiking or labelling with an isotopic standard, in a process referred to as total ion counting (Podwojski et al., 2010). In order to balance the pursuit of maximized protein identifications with accurate quantitation, the $LCMS^E$ method was devised, which unlike other scanning techniques that switch between MS and MS/MS modes, uses a single data independent MS run to obtain protein identifications and quantitative data points by alternating between low and high

energy collision conditions allowing for all peptides to be fragmented without selecting precursor ions (Silva et al., 2006).

The tandem MS^E data acquisition mode alternates in applying low and high collision energy; the first of which does not result in precursor ion fragmentation. In the secondary MS/MS scan, the applied collision energy is ramped up linearly, inducing the fragmentation of any molecular species present within the collision cell. This two-tiered system of sample analysis allows for accurate quantitation through both the spectral counting methodology which compares the number of attained MS/MS spectra from the same sample across multiple LC MS/MS runs and that based on changes in ion intensity as reflected by variations in peak height or area. Proteins of high abundance typically yield more proteolytic peptides following digestion, leading to greater overall sequence coverage and total identified MS/MS spectra; these combined spectral counts are seen to strongly correlate with relative protein abundance (Liu et al., 2004; Washburn et al., 2001). In ion intensity based quantitation, applied to this GLCS label free dataset, research shows great consistency between signal intensity from ESI and total ionic concentration. A study looking at myoglobin digests analysed by LC MS/MS demonstrates a clear trend of increase in peak area correlated with higher concentrations of injected myoglobin digests. This correlation was retained even when myoglobin was spiked into a complex mixture of human serum (Chelius and Bondarenko, 2002).

Following extensive preparatory and analytical steps of the GLCS plasma pools, twenty-six differentially expressed proteins were identified with six of these further validated in individual subject plasma and four displaying results suggestive of their plausibility as novel biomarker candidates for HCC. The decision for which shortlisted markers to take forth for extensive, individual subject based validation was dependent on literature reports of the proteins in relation to HCC, commercial kit and antibody availability as well as representation of various protein families. As this work represents the first extensive protein biomarker study conducted on a West African population, it was important that shortlisted differentially expressed proteins which overlapped with previous reports in other populations were prioritized for analysis.

3.1 Materials & Methods

Acetonitrile (CH_3CN) was purchased from Merck Millipore; C18 Sep-Pak® Cartridges from Waters, USA; Dithiothreitol (DTT), Formic Acid (FA) and Sodium Chloride (NaCl) from Alfa Aesar®; the EZQ® Protein Quantitation Kit, Protein A Agarose and SYPRO Ruby from Invitrogen®; Trypsin (sequencing grade) was purchased from Promega; Bromophenol Blue (BPB), Glu-fibrinopeptide , Glycerol, HEPES (4-2-hydroxyethyl-1-piperazineethanesulfonic acid), Iodoacetamide, Sodium Dodecyl Sulphate (SDS), Trizma Base and Urea were purchased from SIGMA®; Acetic acid (AA), β-mercaptoethanol (BME), Chloroform, EtOH, Methanol (MeOH) and Thiourea were purchased from SIGMA-ALDRICH®; the Enzyme linked immunosorbent assay (ELISA) kits for α1AT, Apo A1 & HPX were purchased from

Immunology Consultants Laboratory, Inc. (Portland, OR, USA); the CC3 ELISA kit was purchased from Abnova (Taiwan).

3.1.1 Study Population

The plasma samples used in this project were collected as part of a five-year liver cancer case-control study. The participants were recruited from three major health care facilities within the Gambia; Royal Victoria Teaching Hospital (RVTH), Bansang Hospital and the MRC Unit, Fajara. The study consists of patients who were identified by tertiary referral of cases by local Physicians after displaying signs of liver disease as well as those who were actively sought through field worker surveillance of wards and clinics.

3.1.2 Subject Selection Method & Diagnosis

Eligible individuals were approached regarding the study and those who were willing gave informed consent to be included (see figure 4.2 for schematic on how subjects were diagnosed). Subjects were bled in Ethylenediaminetetraacetic acid (EDTA) tubes and in the urban health facilities had their blood samples sent to the on site Serology lab for immediate separation followed by storage at -20°C for working samples and -80°C for all remaining aliquots. The samples from Bansang Hospital were separated immediately following bleeding and frozen at -80°C on site pending transport on dry ice to the MRC Serology Lab in Fajara. Working aliquots were taken for lab analyses, which were used to profile recruited subjects as controls, LC or HCC cases based on clinical examination, AFP measurement, ultrasonography and liver biopsy. An ultrasound scoring system (Hung et al.,

2003; Lin et al., 1993) was used in the diagnosis of LC & HCC in the GLCS subjects. This was based on ultrasonographic examinations of the liver surface, parenchyma, vascular structure and spleen compared against the gold standard of liver biopsy. In HBV related LC, sensitivity and specificity values of 77.8% and 92.5,% respectively have been attained. Approximately 24% of the subjects diagnosed as having HCC in the GLCS sample set underwent a liver biopsy (table 3.1).

Scientific and Ethical approvals (MRC Scientific Steering Committee (SCC) No. 1154) were granted for single aliquots of 0.5-1ml to be removed from the frozen archives for utilization in this study. The GLCS subjects were not tested for any viral infections aside from HBV and HCV.

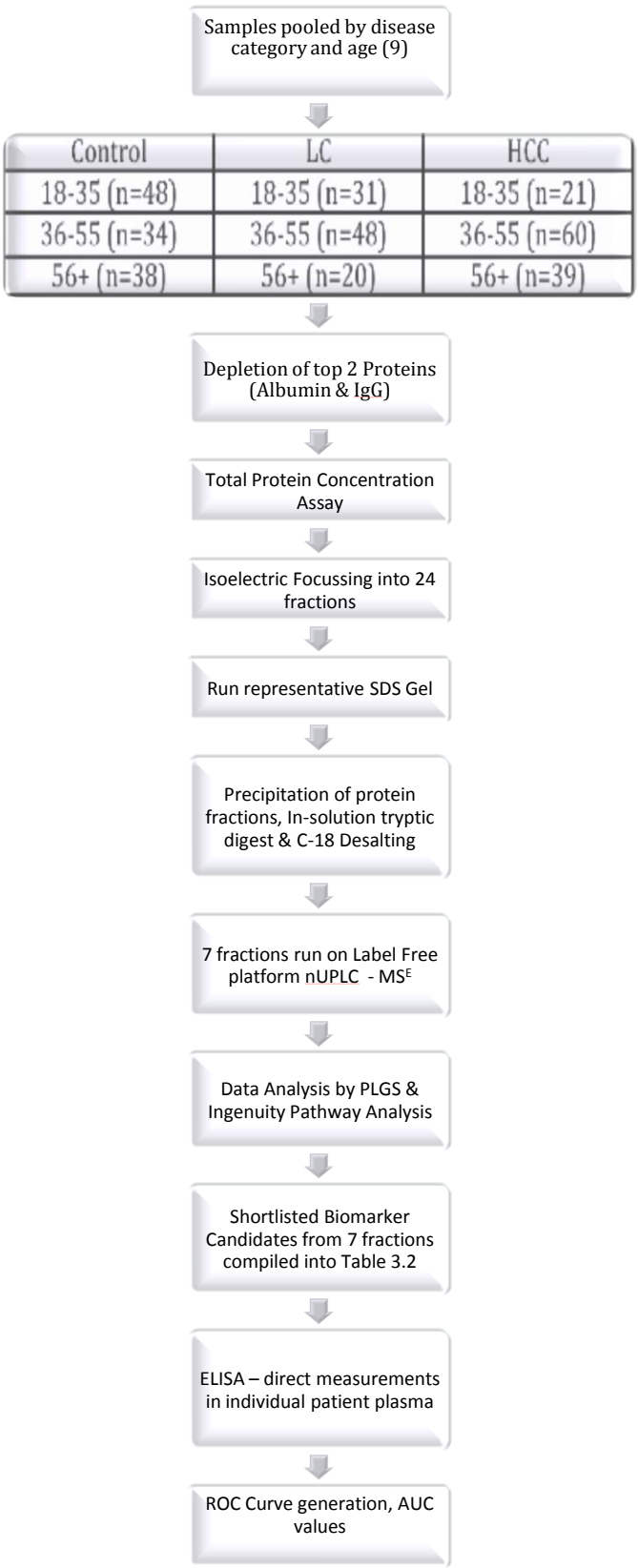


Figure 3.1: Scheme of workflow detailing major steps undertaken from discovery and quantitative proteomics to independent validation.

3.1.3 Sample Pooling

Suspected aetiology of liver disease and age were considered as the basis for stratifying and pooling samples but the latter was chosen as it offered a more even distribution of patients, per sub-group. This approach also allowed for a more distinct separation strategy, as many of the subjects were positive for more than one of the key aetiological factors (HBV, HCV infection or p53 mutation). There is also a growing body of work which suggests that there are age related changes which occur in the liver resulting in its reduced volume, decline in metabolism of certain drugs, changes in protein expression and lowered hepatobiliary functions. Other more subtle changes have also been linked to the ageing liver, particularly at the DNA level and are believed to result in lowered rates of DNA repair, which may have direct implications in virus, or toxin induced mutations that lead to hepatocarcinogenesis (Namieno et al., 1995; Schmucker, 2005). Pooling the subjects by age and disease categorization (figure 3.1) provided not only more balanced sample numbers per sub-group but also allowed for the grouping of patients with expected similar levels of basal liver functionality.

As a first step, 20µl of plasma was taken from subjects in each age and disease category and pooled together. From this, 200µl of pooled plasma was taken forth into the Immunoglobulin G (IgG) and albumin depletions and downstream protocols as described below.

3.1.4 T2 Depletion & Protein Quantitation

The nine sample pools were each subjected to Top 2 (T2) depletion of albumin and IgG (Goding, 1978; Hjelm et al., 1972; Fu et al., 2005). A starting volume of 125µl whole plasma was firstly delipidated by a 15 minute, 15,000G centrifugation step at room temperature. Concurrently, 200µl of Protein A beads (Goding, 1978; Hjelm et al., 1972) were washed three times with a 100mM NaCl, 10mM HEPES pH 7.4 buffer with intermittent inversion and vortexing. The mixture was spun down at the end of each wash cycle in order to remove the holding liquid, with careful attention given to ensure the aspiration of all remaining solution after the concluding wash. Following delipidation, 100µl of plasma was removed from below the lipid layer and diluted 3:1 with 150mM NaCl. The mixture was added to the washed beads and allowed to co-incubate on rotation for 1 hour at room temperature, after which the IgG depleted fraction was transferred to a fresh tube. The Protein A beads were given a final wash with 100mM NaCl, 10mM HEPES – pH 7.4 and the collected supernatant combined with the previously collected fraction. To this was then added chilled EtOH to a final concentration of 42% v/v followed by a one hour incubation at 4°C. The precipitated plasma proteins were collected by centrifugation at 15,000G at 4°C for 45 mins and the resultant pellet resuspended in off-gel fractionation buffer (6.3M Urea, 2.4M thiourea, 45mM Dithiothreitol (DTT), 6% v/v glycerol and pH 3-10 carrier ampholytes).

Representative 1µl aliquots were taken from each depleted pool and subject to a protein quantitation assay using the EZQ Protein Quantitation Kit as per stated instructions.

3.1.5 Off-gel Fractionation & Gel Viewing

Isoelectric focusing (OFF-GEL Fractionator, 3100 Agilent) into 24 fractions was performed on high-resolution pH 3-10 strips as per manufacturer's protocol. For comparative purposes, aliquots from all 24 off-gel fractions were separated by Sodium Dodecyl Sulphate PAGE (SDS-PAGE) using a 4-12% Criterion precast Biorad system run at 200V for 1 hour. The gels were fixed in 40% MeOH and 10% acetic acid then stained overnight with SYPRO Ruby. A destain was conducted the next day with 10% MeOH and 7% Acetic Acid. The gel images were viewed on a UV transilluminator (AutoChemi System) with image acquisition using the LabWorks 4.6 software. The gel images were compared for all 24 fractions and three disease groups and used to determine which fractions would be taken forward for analysis by mass spectrometry. The chosen fractions are highlighted in figure 3.2a and were primarily picked based on the amount of protein material present, absence of an overwhelmingly abundant protein band and lane uniformity across the respective age groups.

3.1.6 Protein Precipitation, Digestion, Desalting & MS Analysis

The seven selected off-gel fractions were taken forth for purification by MeOH chloroform precipitation of proteins (Wessel and Flugge, 1984). Briefly, this

method adds to 200µl of isoelectrically focussed sample with intermittent vortexing, MeOH to the ratio 4:1; 1:5 parts chloroform and 1:3 parts millipore H_2O followed by rapid centrifugation and supernatant removal. A pellet is acquired after this step and is given a final wash with methanol, spun down by centrifugation and all excess liquid removed.

To digest, the protein pellets from the previous step were each dissolved by sonication and vortex in a 6M Urea, 100mM Tris pH 7.8 buffer and incubated in a shaker, at room temperature with a 200mM DTT, pH 7.8 reducing solution for 1 hour. This was followed by a 1 hour incubation with an alkylating solution (200mM iodoacetamide, pH 7.8) and subsequent re-incubation with the reducing agent in order to consume any excess iodoacetamide. The urea concentration was then reduced by diluting the reaction mixture with 775µl water. A digestion step was carried out overnight with 2.0µg/µl of sequencing grade trypsin at 37°C. The digested samples were then desalted by a Sep-Pak C-18 (Waters) column. The columns were at first pre-equilibrated by washes with two buffers, solution A: 98% Water (H2O), 1.9% Acetonitrile (CH3CN) & 0.1% Formic Acid (FA) & solution B: 80% CH_3CN, 19.9% H2O and 0.1% FA. The digested sample was passed through the column, washed with solution A and the bound peptides eluted with 2ml of solution B. The resultant purified peptides were evaporated in a vacuum centrifuge, frozen at -80°C and resuspended in 50µl of solution A on the day of analysis.

The samples generated from this extensive workflow were analysed by nano-UPLC™ ESI-MS/MS in triplicate on a 75µm inner diameter 25cm length C18 nano-Acquity™ UPLC™ column coupled to a Q-TOF premier mass spectrometer (Waters), with 1.7µm particle size on a 90 minute gradient of 2-45% solution B. Data acquisition was performed in high-definition MS^E mode which utilizes high and low collision energy switching, every 1.5 seconds (Xu et al., 2008).

3.1.7 Data Analysis and Quantitation

Data processing including normalization, deisotoping and deconvolution was performed using the Protein Lynx Global Server (PLGS v2.2.5, Waters). MS/MS spectra were reconstructed by combining all precursor and fragment masses with identical retention times. The mass accuracy of the raw data was corrected using Glu-fibrinopeptide (200 fmol/µl; 700 nl/min flow rate; 785.8426 Da [M + 2H]2+) that was infused into the mass spectrometer as a lock mass during analysis (every 30 seconds). Peptides and regarding proteins were identified by searching the peak lists against a UniProt/SwissProt database [version 2009.04.23; 19713 human sequence entries] using the following parameters: Minimum fragment ion matches per peptide: 3; minimum fragment ion matches per protein: 7; minimum peptide matches per protein: 1; maximum protein mass: 250000 Da; primary digest reagent: trypsin; missed cleavages: 1; fixed modifications: Carbamidomethyl; variable modifications: Oxidation. All protein hits that were identified with a confidence of >95% were included in the quantitative analysis. Identical peptides from each triplicate set per sample were clustered based on

mass precision (<10 ppm, typically ca 5 ppm) and a retention time tolerance of <0.25 min using the clustering software included in PLGS v2.2.5. If two or more distinct proteins shared an identical peptide but were found to be regulated differently, then the quantitation algorithm did not include the peptide in question. In order to allow for this, peptide probabilities are always softened by the PLGS software slightly prior to quantitation. Because of this, the contributions from peptides with even 100% probability of presence were suppressed in order to avoid potential errors in quantitation. Normalisation was performed using the PLGS "auto-normalization" function. The statistical significance of relative expression ratios was calculated using a Monte-Carlo algorithm and expressed as $p < 0.05$ for down-regulated and $1\text{-}p > 0.95$ for up-regulated proteins, respectively. Only proteins that were identified in two out of three replicate analysis runs were selected for further analysis (replicate filter two). Multiple protein entries that had controversial regulation values throughout different fractions were not considered for shortlisting. A list of differentially expressed proteins was from PLGS imported into the Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems®), and the Biomarker discovery tool was consulted for the short-listing of proteins as potential marker candidates (table 3.2).

3.1.8 ELISA Assays

ELISA based measurements were carried out using commercial kits for four proteins; α1AT, Apo A1, HPX, and CC3. In brief, antibodies from the protein of interest were adsorbed to the base of a microtitre well plate, reacting with and

binding specific antigens in the diluted plasma. Any unbound protein was removed by four consecutive wash steps followed by incubation with Horseradish Peroxidase (HRP) conjugated antibodies to the protein of interest. This was followed by a final quadruple wash step, after which the chromomeric substrate 3,3'5,5'-tetramethylbenzidine (TMB) was added and allowed to develop to optimal intensity, for approximately 10 minutes and the reaction then stopped by addition of 0.3M Sulphuric Acid (H_2SO_4) solution. In their specific measurements, samples for α1AT were diluted 1:50,000; Apo A1 1:20,000 and HPX 1:30,000.

The CC3 ELISA kit varied from the previously described method and is based on a quantitative competitive enzyme immunoassay principle. Polyclonal human CC3 antibodies, raised against the full-length protein were adsorbed to a 96 well microtitre plate. Our optimised protocol started with the addition of plasma samples (diluted 1:400) to each well along with biotinylated CC3 for a competitive co-incubation step. Following this, any subsequently unbound material after two hours of reaction was removed in 5 wash steps. This was followed by further incubation with a streptavidin-peroxidase conjugate and a final wash step prior to the addition of TMB and a 0.5M Hydrochloric Acid (HCL). Stop solution.

All the plates for all four proteins had their absorbance measured at 450nm with a 4 parameter logistic regression standard curve generated to extrapolate select protein levels in individual samples, taking into account the respective dilution factors.

3.1.9 Protein Level Extrapolation & ROC Curves

To determine the diagnostic ability of these four proteins, the statistical software Graph Pad Prism (version 5) was used to generate ROC curves. The area under these ROC's is well recognised as a measure of diagnostic ability for candidate biomarkers (Akobeng, 2007). All the duplicate measurements from the known standard and unknown sample readouts were exported to Graph Pad Prism. The mean of the duplicate measurements for each subject were taken and Log10 transformed followed by the performance of a nonlinear regression sigmoidal dose response analysis (Motulsky and Christopoulos, 2004). This step extrapolated protein concentrations for the unknown samples. The antilog (10x) of the obtained values was then calculated followed by multiplication by the dilution factor in order to obtain accurate analyte levels. ROC curves were generated on the same software by listing the measured protein levels in any two disease conditions. All ROC's were reported with readings for AUC, 95% CI, p values and sensitivity and specificity changes with their associated likelihood ratios (Motulsky, 2003) (table 3.3).

3.2 Results

3.2.1 Main causative factors for Chronic Liver Disease represented in GLCS population in a condition-specific pattern.

The GLCS consists of well-defined patient groups representing the main causative factors of CLD; HBV, HCV and or AF exposure. In the control group the HBsAg positive rate was measured as 8.4% with none of these individuals showing 'e'

Antigen positivity, a marker sometimes indicative of active viral replication. 5.8% of the 120 controls were HCV antibody positive. In the LC group, 14.4% of patients were HCV antibody positive, 54.5% HBsAg positive and from these 31.7% showed the presence of the HBV 'e' antigen. In the HCC sub-group, the HCV positive rate was found to be 31.7%, HBsAg carriage stood at 55.6% and the amount of 'e' antigen positive individuals amongst them at 22.6%. The presence of Single Nucleotide Polymorphisms (SNPs) in the p53 gene were also investigated in the GLCS subjects using Restriction Fragment Length Polymorphism Polymerase Chain Reaction (RFLP-PCR) and this showed the presence of at least one mutation in 4.4% of controls, 12.4% of LC's and 36.2% of subjects diagnosed with HCC. Almost all the point mutations detected were at the codon 249 (table 3.1).

Variable	Control No.	Control %	LC No.	LC %	HCC No.	HCC %
Biopsy	n/a	n/a	5	5.1	29	24.2
Clinical Exam	120	100	99	100	120	100
Ultrasound	n/a	n/a	99	100	120	100
AFP > 100ng/ml	0	0	20 of 95 tested	21.1	106	88.3
Mean ALT (IU/L)	6.58	n/a	12.4	n/a	14.3	n/a
Mean AST (IU/L)	19.3	n/a	64.5	n/a	158.4	n/a
ALT > 56 IU/L	1	94.2 tested	0	92.3 tested	4	92.5 tested
AST > 40 IU/L	8	93.3 tested	46	94.0 tested	83	94.2 tested
Male	84	70	63	63.6	92	76.7
Female	36	30	36	36.4	28	23.3
Age (mean)	44.4	n/a	43.7	n/a	49.1	n/a
HbsAg Positive	10 of 119 tested	8.4	54 of 99 tested	54.5	65 of 117 tested	55.6
Tp53 Mut Positive	5 of 113 tested	4.4	11 of 89 tested	12.4	38 of 105 tested	36.2
E Antigen Positive	0 of 10 positive	0	13 of 41 tested	31.7	12 of 53 tested	22.6
HCV Positive	7 of 120 tested	5.8	14 of 97 tested	14.4	38 of 120 tested	31.7

Table 3.1: Key clinical parameters of the GLCS.

3.2.2 Label-free quantitation of differentially expressed proteins in plasma between LC, HCC and control subjects.

For the proteomics workflow, samples in each age and disease sub group (figure 3.1) were pooled and subject to albumin and IgG depletion. The levels of protein

present in each of the nine pools were found to be of the same order of magnitude thus providing a similar baseline for preparative isoelectric focussing (figure AF 1.0). SDS-PAGE analysis demonstrated fraction-by-fraction profiles, which appeared to be similar between the different disease and healthy control groups with no presence of predominating protein bands. Seven of the total twenty-four isoelectrically focussed fractions containing the majority of protein material were subjected to in-solution trypsin digestion and analysed by Q-TOF MS^E. This yielded 1307 unique protein identifications for the combined 7 processed fractions, with the highest number of protein identifications made at the lanes corresponding approximately to pH 5.5-6.5; (fractions 13, 14 and 15 – figure 3.2a). These results include proteins with condition-specific changes in expression ratios as high as 21 fold (fraction 13) and 22 fold (fraction 15) as well as other proteins exclusively found in specific disease states. Stringent filtering steps, as described in the Methods section were applied to these fractions with only proteins identified with >50% confidence and observed in at least two of three analytical replicates being exported to assemble a shortlist of uniformly differentially expressed proteins.

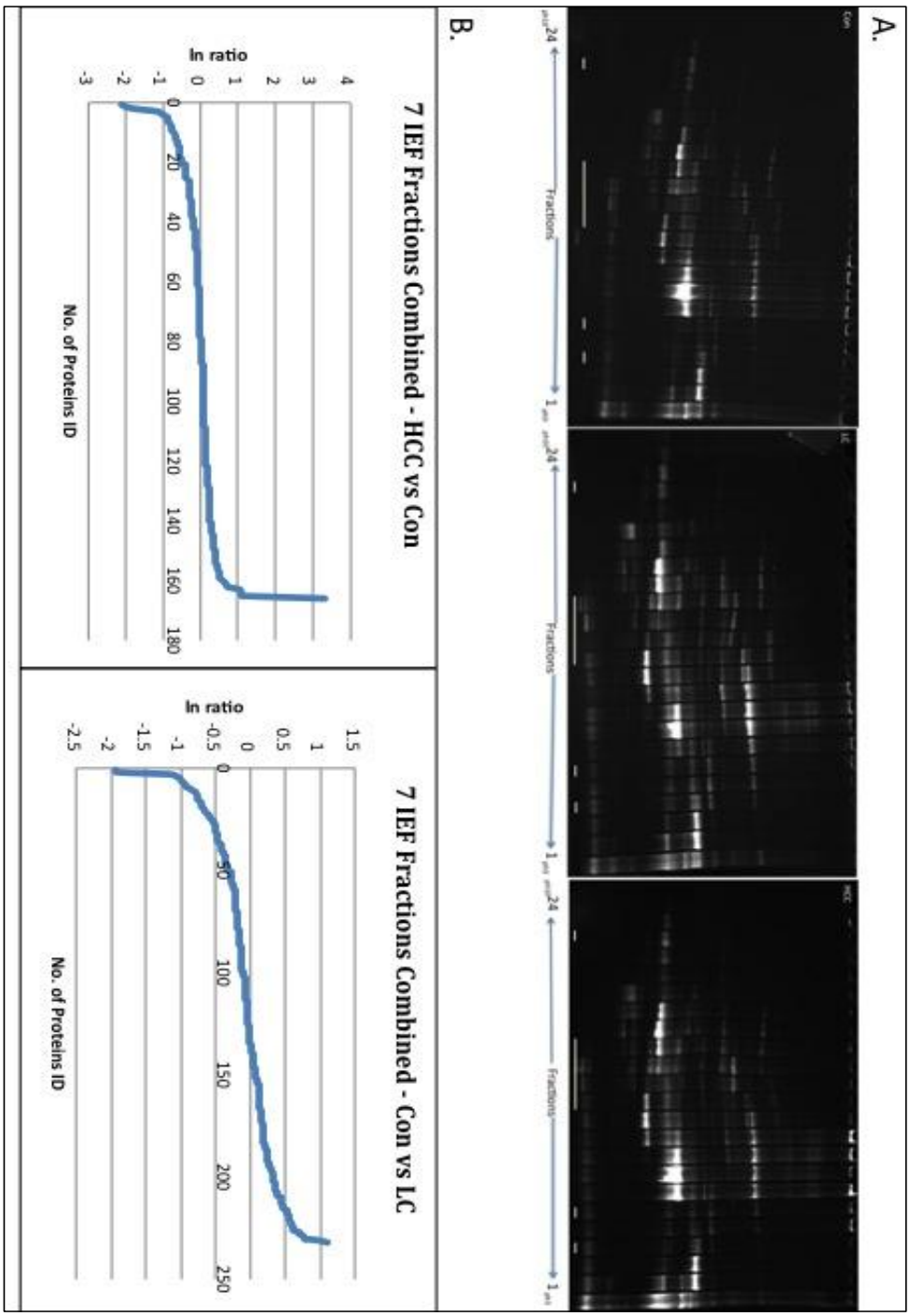


Figure 3.2: *Quality control of fractionation for MS analysis*. (**A**) Representative SDS-PAGE gel analysis of isoelectric focussing (IEF) fractions from each disease group following IgG & albumin depletion and off-gel fractionation of subject pools. Visualisation was performed with Sypro Ruby Red staining showing healthy control (Con, left panel), cirrhosis (middle panel) and hepatocellular carcinoma (right panel). (**B**) Natural log distribution plots of proteins showing their expression changes (ln ratio) for non-liver ailment controls versus HCC and liver cirrhosis cases.

3.2.3 Stringent filtering results in the identification of 26 differentially expressed proteins.

A set of 21 extracellular and plasma proteins were identified and shortlisted following rigorous filtering as differentially expressed in HCC subjects versus individuals with no evidence of liver related ailments. In workflows comparing controls and patients with active liver cirrhosis, 22 proteins were shortlisted as possible biomarker candidates. The overlap between the two groups means that a total of 26 proteins (table 3.2) were with the GLCS population found to have condition-specific signatures over the spectrum of non-liver related ailments to the development of liver cancer. Of these 26, alpha-2-macroglobulin, complement factor I (CFI), carboxypeptidase N (polypeptide 2) and caspase 8 showed expression patterns suggesting their diagnostic potential for HCC only, whilst the MS^E data for complement component 4A (CC4A), haptoglobin, complement factor B (CFB), α1AT and serum amyloid p suggested the changes in these proteins as unique to subjects with LC. As the diagnosis of early stage HCC in a cirrhotic liver is of significant clinical importance, all validated proteins, regardless of which progressive condition they seemed significant or unique for, were measured in all three sub-groups.

Differentially Expressed Proteins	Fraction	Swissprot ID	PLGS Score	Seq. Coverage. %	No. of Peptides	HCC vs. CON Ratio	CON vs. LC Ratio
α-2-macroglobulin	14	P01023	1385.68	67.5	83	1.04*, 1.32*, 1.04*	***HCC vs. CON***
apolipoprotein E	14	P02649	391.2	71.3	27	1.42*, 1.60*, 1.40*	0.92*, 0.74*, 0.76*
complement component 4 binding protein,**α**	14	P04003	91.8	40.4	20	1.07*, 1.68*, 0.96*	0.97•, 0.84*, 0.90*
complement factor I	14	P05156	61.02	36.7	20	1.07•, 1.07•, 1.04•	***HCC vs. CON***
glutathione peroxidase 3	14	P22352	61.26	42	10	1.27*, 1.45*, 1.04•	0.66*, 0.61*, 0.72*
apolipoprotein H	15	P02749	49.75	41.4	10	1.04•, 1.20*, 1.06•	2.39*, 1.77*, 1.80*
complement component 4B	6	P0C0L5	384.56	17.1	22	1.02•, 1.68*, 1.04•	1.30*, 1.36*, 1.21*
α-1-antichymotrypsin	6	P01011	524.43	37.6	13	1.65*, 1.34*, 1.28*	0.84*, 0.51*, 0.96•
carboxypeptidase N, polypeptide 2	4	P22792	206.51	38	14	1.19•, 1.05•, 1.06•	***HCC vs. CON***
leucine-rich **α**-2-glycoprotein 1	4	P02750	170.54	49.6	15	1.86*, 1.40*, 1.11•	0.51*, 0.84*, 0.79*
complement component 3*	14	P01024	569.1	52.1	79	0.88*, 0.84*, 0.78*	1.46*, 1.56*, 1.19*
apolipoprotein A-I	15	P02647	110.79	65.9	17	0.55*, 0.68*, 0.66*	1.52*, 1.23*, 1.73*
complement factor H	15	P08603	566.61	62.5	63	0.90*, 0.69*, 0.90*	1.15*, 1.42*, 1.25*
haptoglobin-related protein	15	P00739	191.45	59.2	21	0.64*, 0.65*, 0.99•	1.43*, 1.38*, 1.21*
hemopexin	13	P02790	333.07	59.3	23	0.56*, 0.45*, 0.68*	1.57*, 1.92*, 1.60*
α-1-microglobulin/bikunin precursor	6	P02760	158.24	31	7	0.75*, 0.71*, 0.73*	1.39*, 1.27*, 1.42*
paraoxonase 1	6	P27169	146.28	39.4	9	0.54*, 0.81•, 0.54*	1.27*, 1.52*, 1.78*
clusterin	4	P10909	102.59	39.2	17	0.76*, 0.48*, 0.52*	1.52*, 1.67*, 2.30*
complement component 4A	15	P0C0L4	82.92	19.6	28	***CON vs. LC***	1.28*, 1.30*, 1.22*
haptoglobin	13	P00738	286.29	60.1	22	***CON vs. LC***	1.54*, 1.35*, 1.17*
complement factor B	21	P00751	544	52.2	36	***CON vs. LC***	2.36*, 1.97*, 1.20*
α-1-antitrypsin	4	P01009	186.57	39.5	14	***CON vs. LC***	0.39*, 0.29*, 0.87•
caspase 8	13	Q14790	16.81	12.5	7	0.80•, 0.66*, 0.93•	***HCC vs. CON***
fibrinogen **α** chain	13	P02671	75.6	27.5	19	0.90*, 0.84*, 0.51*	1.12*, 1.11•, 1.52*
amyloid P component, serum	13	P02743	94.24	36.8	10	***CON vs. LC***	2.83*, 1.84*, 1.40*
CD5 molecule-like	12	O43866	650.63	61.7	23	1.17*, 1.32*, 1.19*	0.61*, 0.54*, 0.53*

Table 3.2: Differentially expressed proteins in plasma from controls, LC and HCC subjects identified by quantitative MS. * p: </=0.05•p: >0.05

3.2.4 Alpha-1-Antitrypsin, Complement Component 3, Apolipoprotein A1 and Hemopexin show same expression trends in MS and ELISA based analysis.

Validation by ELISA allowed for direct measurements of target proteins in plasma in order to verify the expression changes suggested by mass spectrometry. To achieve this, we selected four shortlisted proteins based on representation of various protein families interpreted as suggestive of their involvement in different pathways, implications in literature and availability of quality commercial antibody based methods for direct measurements and took them forth for more comprehensive validation. Plots showing the relative expression trends for three representative peptides in triplicate allocated to α1AT, CC3, Apo A1 and HPX identified with high scores in all three disease categories were selected and compared (figures 3.3e & 3.4a). With the exception of a few outliers, the trends in expression seen at the peptide level, between the three disease groups were identical to those seen in direct measurement of specific proteins in individual subjects by ELISA.

3.2.5 ELISA validation of four markers identified by label-free proteomics suggests diagnostic potential.

The four proteins selected for validation were measured in individual subject plasma across the three disease groupings using ELISA based assays. Measurement of α1AT showed a net increase in the proteins overall level with the most significant change observed between mean amounts of plasma α1AT in the control versus HCC groups (figure 3.4); an AUC of 0.84 (95% CI: 0.78-0.90),

p<0.0001 calculated by ROC analysis suggests a high discriminatory potential (table 3.3). HPX levels in patient plasma showed an overall trend of lowered expression in those with active cirrhosis compared to controls, with a "bounce back" effect seen as the protein levels are somewhat restored to control levels in HCC patients. The highest reported AUC was for the change in HPX levels of Con vs. LC; 0.82 (95%CI: 0.76-0.87), p<0.0001. The trend for CC3 expression was similar to that of HPX with its levels dropping from control to LC and rising again in the presence of a liver tumour. However with this protein, the biggest observable difference was seen between cirrhotic and HCC patients with an AUC of 0.70 (95% CI: 0.63-0.78) p<0.0001. The last protein from our shortlist which was validated by ELISA was the high density lipoprotein component, Apo A1. The trend with this candidate was a significantly lowered expression as patients develop LC and HCC. The biggest shift in expression was observable between the control and HCC groups with a reported AUC of 0.83 (95%CI: 0.76-0.90) p<0.0001. The strong suppression in ApoA1 levels seen between controls and LC subjects also gave a significant AUC reading of 0.79 (95%CI: 0.71-0.87) p<0.0001 (table 3).

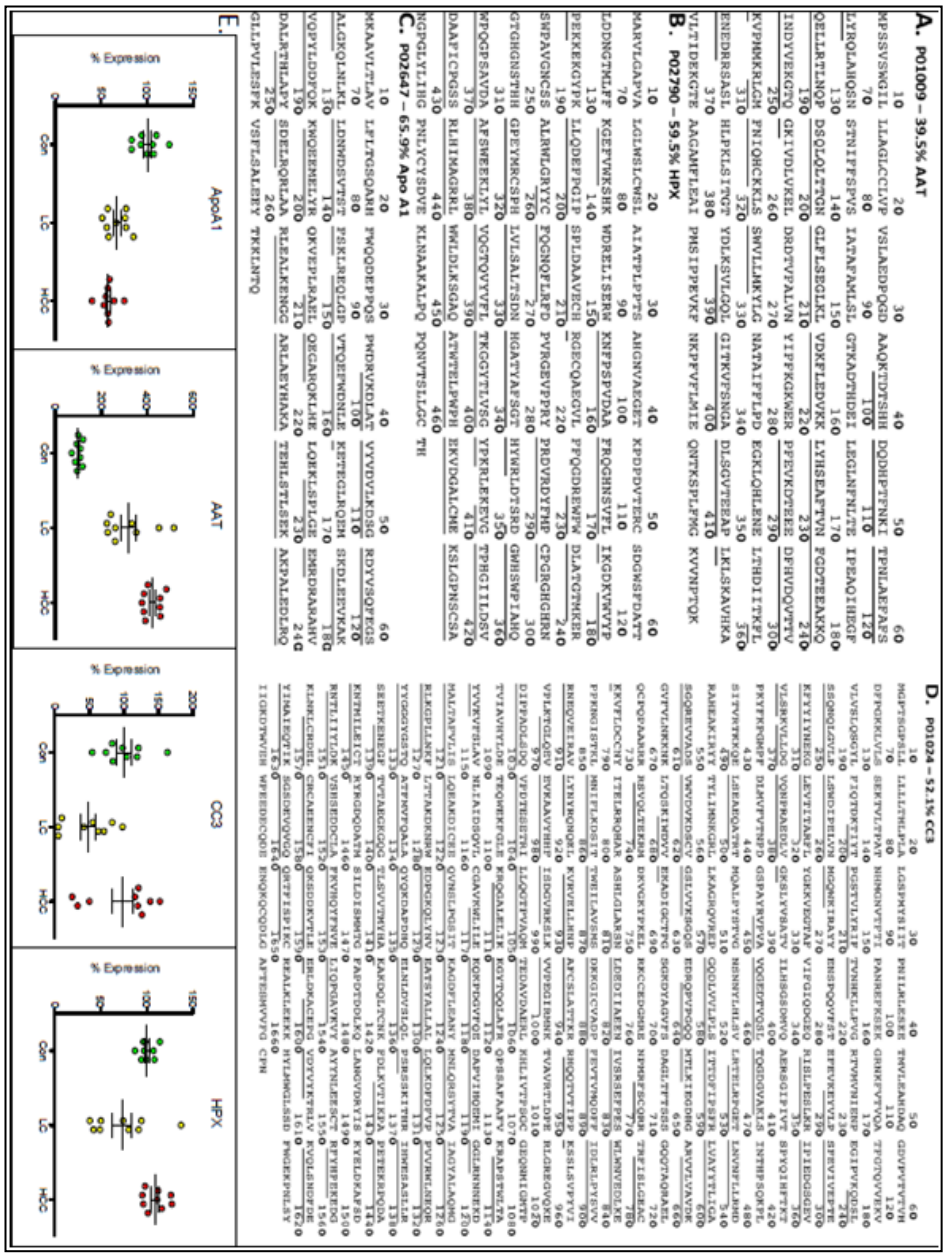


Figure 3.3: Identification of differentially expressed proteins in plasma from control, cirrhosis and HCC subjects by label-free quantitative MS. (**A-D**) Sequence coverage levels for differentially expressed proteins identified and selected for independent validation. Underlined sequences represent tryptic peptide fragments identified by MS^E. (**E**) Relative expression plots of robustly identified peptides uniquely attributed to either ApoA1, α1AT, CC3 or HPX and comparing their relative mass peak intensities across the three different disease groups.

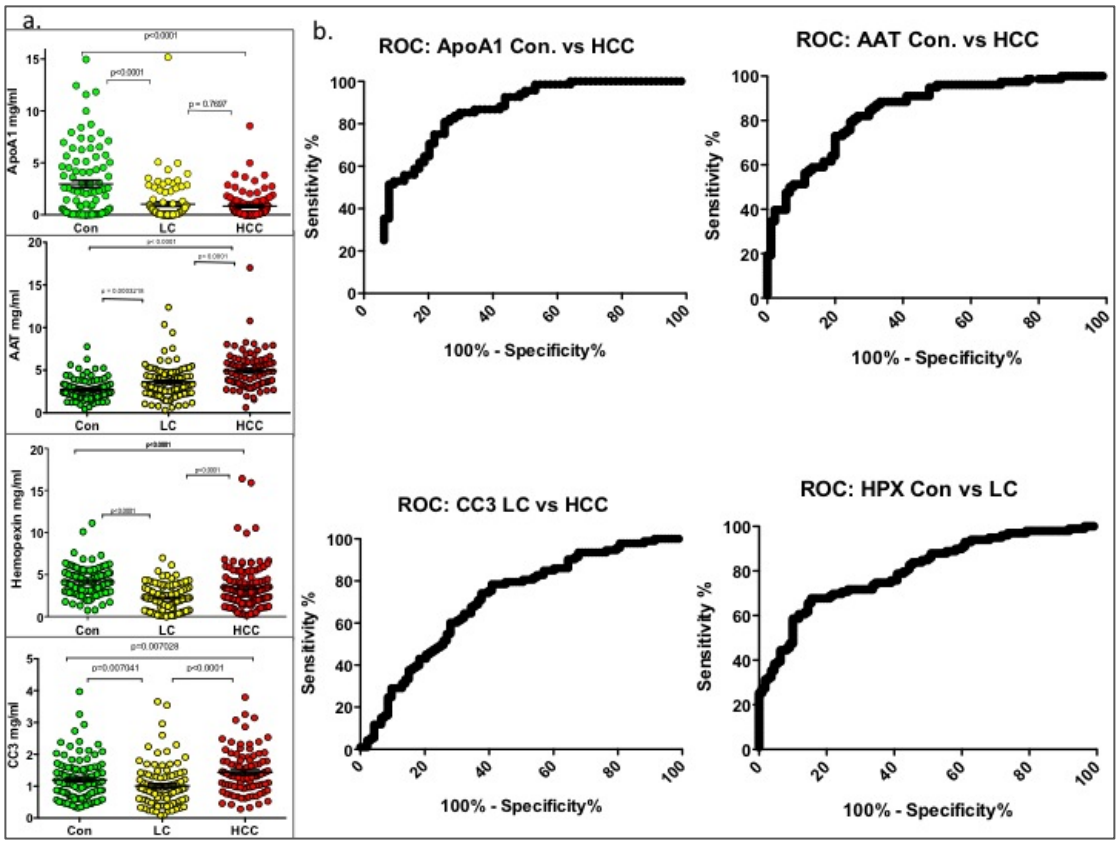


Figure 3.4: *Protein marker validation by ELISA*. (**a**) Dot plots of protein levels observed in plasma of individual subjects is shown for four proteins with mean and standard error of mean (SEM) following direct measurement, highlighting disease-specific changes in protein expression. (**b**) ROC curves of the stand-alone differentially expressed proteins shown in 3.4a as a measure of candidate marker discriminatory potential contrasted with AFP performance in same population.

Protein & Sample Numbers Tested	Con vs. HCC	Con vs. LC	LC vs. HCC
ApoA1 **Con: 64, LC: 65, HCC: 68**	AUC: 0.8341 (95% CI: 0.7642 to 0.9040) p< 0.0001	AUC: 0.7899 (95% CI: 0.7110 to 0.8688) p< 0.0001	AUC: 0.5595 (95% CI: 0.4613 to 0.6577) p=0.2365
α1AT **Con: 100, LC: 96, HCC: 87**	AUC: 0.839570 (95% CI: 0.7825 to 0.8966) p<0.0001	AUC: 0.6489 (CI: 0.5711 to 0.7266) p=0.0003218	AUC: 0.7006 (95% CI: 0.6253 to 0.7758) p<0.0001
CC3 **Con: 114, LC: 99, HCC: 93**	AUC: 0.6090 (95% CI: 0.5322 to 0.6859) p=0.007028	AUC: 0.6072 (95% CI: 0.5309 to 0.6834) p=0.007041	AUC: 0.7024 (95% CI: 0.6288 to 0.7760) p<0.0001
HPX **Con: 120, LC: 99, HCC: 119**	AUC: 0.6594 (95% CI: 0.5889 to 0.7299) p<0.0001	AUC: 0.8160 (95% CI: 0.7598 to 0.8722) p<0.0001	AUC: 0.6554 (95% CI: 0.5834 to 0.7274) p<0.0001

Table 3.3: Data of sample numbers tested per protein along with AUC's, confidence intervals and associated p-values.

3.3 Discussion

As with many solid tumours for which high performance non-invasive diagnostic tools are lacking, the identification of reliable predictive markers for HCC is of prime importance. The greatest burden of this disease and its major causative factors exist in the developing world where many of the sophisticated imaging tools used to compensate for the insufficient performance of AFP are not widely available. The liver, as the main site of plasma protein synthesis and metabolic activities such as detoxification and storage is a central organ in the body. Measurement of plasma proteins, most of which are processed by or derived from the liver, can as such provide deep and relevant insight into how specific changes in key proteins may be relevant in the development of hepatocarcinogenesis.

In the Gambia where the subjects under study in this project were drawn from, prior to the introduction of the HBV vaccination program, approximately 15% of the population were known to be chronic HBV carriers by the time they reached adulthood (Vall Mayans et al., 1990; Whittle et al., 1990) with horizontal transmission being the main route by which the virus is transmitted (Chang, 2007; Hsu et al., 1993). HCV carriage is generally low nationwide, with the highest rates seen in older persons in what is believed to be a cohort effect. The burden of HCV in our selection of GLCS subjects is higher than previously reported but this is likely due to the high mean age of HCC cases (55.3 years) in comparison to the populations other studies have been based on. Investigations comparing the presence of HCV antibodies in subjects from three different time periods show consistently higher infections rates in older age groups of 50+ (Kirk et al., 2006).

As stated in the introductory chapter, overall levels of population exposure to AFB1 are ubiquitous but the presence of aflatoxin-albumin adducts in as much as 95% of the population (Wild et al., 1992) suggests substantial levels of contact. Elevated HBV 'e' antigen rates in LC and HCC subjects as compared to the non liver ailment controls support the suggestion that active viral replication is a great risk factor for advanced liver disease (Yang et al., 2002). The mechanism through which this occurs is not well established but vast integration of the HBV with host DNA is suggestive of direct viral effects (Brechot et al., 1980; Hino et al., 1984; Matsubara and Tokino, 1990). Other sensitive assays exist for measuring direct AFB1 exposure (Everley et al., 2007; Wild et al., 1990); however the p53 mutation rates tabulated could be indicative of the overall disease stage which may be a

cause or effect of the observed genetic changes (Teramoto et al., 1994). The question of whether this mutation precedes the development of HCC or occurs as an after effect cannot be answered from this dataset. However, results from this subject group suggest that there are events which occur in the process of liver scarring and tumorigenesis of the hepatocytes which either lead to or result in an increased rate of mutation at codon 249, commonly resulting in an amino acid shift of arginine to serine (Aguilar et al., 1993). All of these factors contribute to HCC progression and render a systematic discovery of molecular hallmarks challenging.

To address the issue of heterogeneity which impacts all human studies, a large population of 339 subjects was selected and pooled as to average out individual differences and highlight more general overarching trends. Pooling does have the impact of neutralizing potentially interesting changes and observations, however it offers a workable platform for the application of a fractionation protocol enabling increased plasma proteome coverage, particularly on a large sample size. The pooling of samples by age within the distinct disease categories encouraged the extraction of proteins whose secretory and regulatory signatures are not age dependent. Candidates robust enough to remain highlighted through this analyses stand at some advantage, as the observation of uniform trends across the three age ranges suggest they will likely be usable at all levels of healthy or disease liver function. To address the issue of individual variety, selected biomarker candidates were then validated by ELISA in individual patient plasma to see if the changes suggested by MS analysis were maintained (figure 3.4a). Such

direct and individualised measurements compensated for some of the more unique trends which may have been lost by globalized sample pooling.

Direct measurement of Apo A1 by ELISA in individual subjects (figure 3.4a) confirms the changes in protein abundance predicted by MS and shows the protein to be severely down-regulated with impaired liver function as determined by deteriorating liver enzyme readings (table 3.1). This finding suggests that changes in its secretory pathway are altered during disease progression. Other studies (He et al., 2003; Steel et al., 2003) corroborate these findings with one group proposing the lowered expression of Apo A1 in subjects with confirmed HCC as a prognostic marker for portal vein metastases (Qiu et al., 2008). Further work needs to be carried out to confirm these results in independent cohorts as the ability to identify individuals at elevated risk for hepatic portal vein invasion could have significant clinical implications. α1AT by comparison shows a reverse trend with increasing circulatory plasma levels as liver function wanes. This protein has three reported glycosylation sites (Kolarich et al., 2006) with its core fucosylated form (Comunale et al., 2010) specifically shown to have altered modification patterns that are highly sensitive and specific to HCC and or LC (Wang et al., 2009b).

Direct ELISA measurements of CC3 and HPX (figure 3.4a) in individual subjects show a dip in expression with the development of LC and a restoration to healthy levels of protein expression with hepatotumorigenesis. This pattern may be indicative of a loss of liver functionality with LC due to a larger volume of the liver

being affected, whereas hepatocarcinogenesis on a liver without a background of extensive cirrhosis could have greater overall functionality and hence still be able to produce key proteins at customary levels. Such distinct signatures, particularly in the restoration of the expression of these two proteins to healthy levels could be exploited in discerning patients who have developed HCC over a background of LC. Success in this could have significant clinical impact. Our findings are consistent with other studies which demonstrate that CC3 levels or those of its processed units alter significantly between individuals with LC and HCC (Chang and Chuang, 1988; Lee et al., 2006). A study, which measured levels of fucosylated HPX also found significantly, elevated levels of its expression in a cohort of 229 serum samples from patients with chronic hepatitis, cirrhosis, and HCC (Morota et al., 2011), a finding that was confirmed independently (Comunale et al., 2009).

As described above, MS and ELISA based measurements of the levels of Apo A1, HPX, CC3 and α1AT suggest all four to have significant potential in discriminating between at least two of the three subject groups under study. The AUC's reported for each of these are for the most part better in performance than AFP (Gupta et al., 2003) when interpolated as a measure of sensitivity. Within the GLCS, AFP shows an AUC of 0.96 in discriminating between control and HCC subjects. However, this figure has to be placed into its proper context as all of the diagnoses made in this case-control study were based on multiple clinical and imaging based evaluations, hence boosting the overall accuracy of determination. A proper comparison of how AFP compares to the four candidate markers could

only be conducted if the test was solely utilised and not interpreted as part of a panel. These expression changes will importantly need to be validated in independent HCC-LC cohorts, ideally with uniform aetiology of liver disease to the Gambian population, in order to see if the trends identified in the GLCS subjects are sustained. It is worth noting however that these four proteins form only a subset of the total 26 which were shortlisted by label-free proteomics and following up as many of these candidates as possible through MS independent methods is essential in establishing their robustness as biomarkers which can be utilised in clinics globally. Thus, the next stage in their evolution would be to not only validate these outcomes in independent cohorts of similar and more distinct aetiology of liver disease, but also to use biochemical methods to home in on sustained, significant and unique changes in structure and modifications and propose these for development into low cost, non-invasive diagnostic tests.

3.4 Conclusions

This study is unique in its utilization of one of the larger cohorts from a population with high exposure to the main environmental & viral factors causative for HCC and LC and applying discovery proteomics methods to its pooled fractions. The results described show proteins whose condition specific changes have remained consistent from MS to ELISA based measurements. Exploitation of this could contribute to proposing a panel of markers which if sufficiently validated could be diagnostic or prognostic of advanced liver disease.

Chapter 4: Independent Validation of Putative Markers by ELISA and Biochemical Characterization in GLCS & Regional Subject Population.

4.0 Introduction

Protein biomarker discovery can be approached in several ways. Our previous chapters detail a targeted discovery scan of the sub-fractionated proteomes of control versus disease subject plasma pools. The ensuing validation steps focussed on six of the shortlisted proteins based on their differential abundance, ease of measurement and associated biology to see if the changes suggested by quantitative MS (figure 3.3 & table 3.2) were maintained. The results from this demonstrated that there were some consistent, noteworthy trends of differential protein expression in the three groups under study. In order to further elucidate and characterize these differences, biochemical approaches of sample analysis were adapted. These aimed to look at HPX, α1AT, CC3 and Apo A1 in greater structural detail using methods for glycoprotein enrichment, IEF and immunoblotting.

4.1 Need for better diagnostic tools for LC & HCC

The burden of HCC is greatest in the developing world, as such making it the primary area where access to high performing, affordable diagnostic tests useable in both urban and rural settings are primarily needed. The key difference in approach to HCC and CLD diagnosis and management between developed and

developing nations highlights the need for realistic standards and guidelines of diagnoses relevant to the differing capacities of high prevalence areas (Bialecki and Di Bisceglie, 2005; Torzilli et al., 2004).

Despite the introduction of the HBV vaccine into the EPI schemes of many nations; there still remains a significant amount of places where mass immunisation has been ineffectively introduced or not at all (figure 4.1).

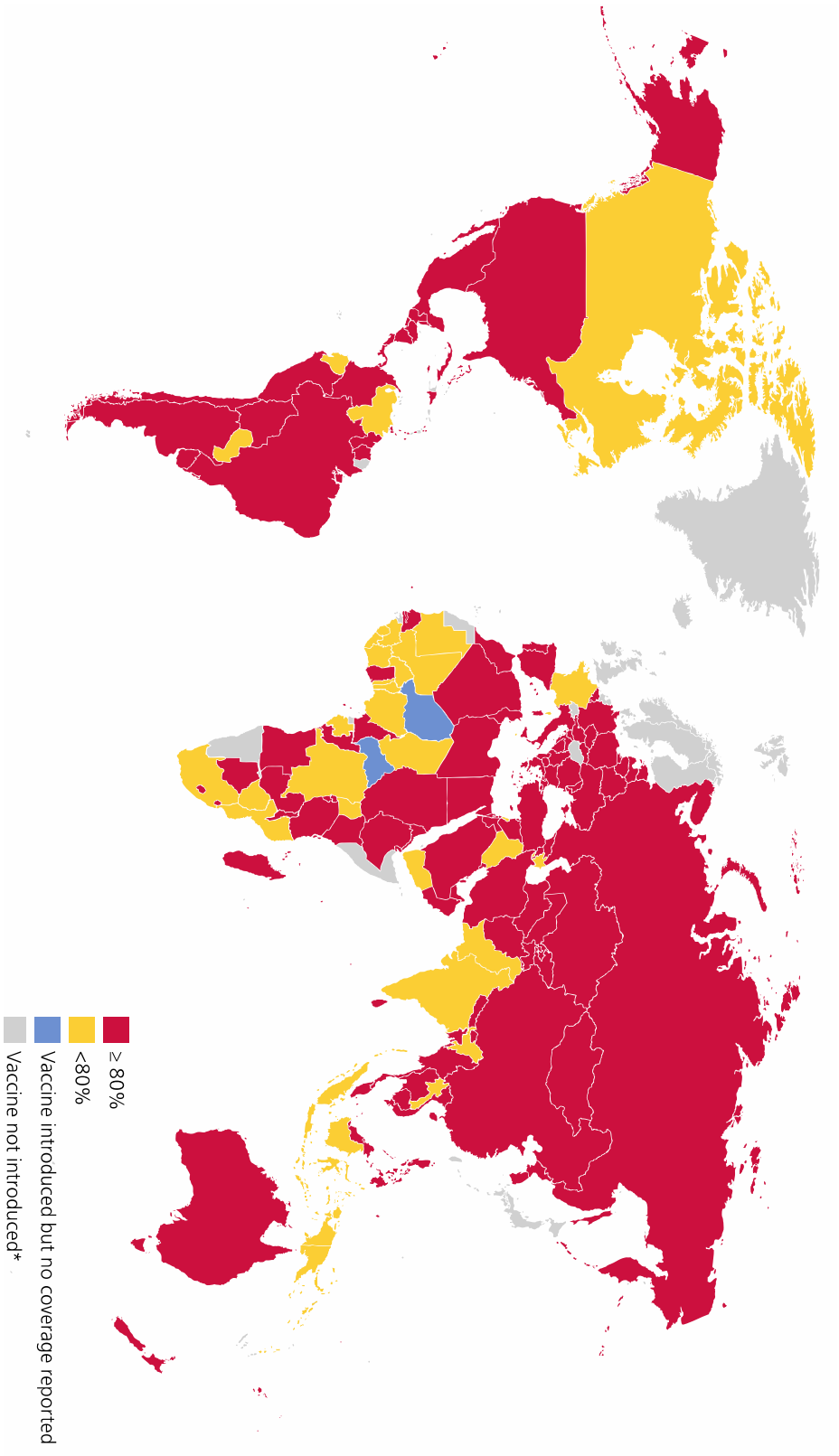


Figure 4.1: WHO/UNICEF estimated HBV vaccine program coverage amongst infants worldwide 1980-2008. Source: American Cancer Society, Global Cancer Facts & Figures 2nd Edition. Atlanta ACS; 2011

One of the expected outcomes of the mass immunisation of infants with the HBV vaccine is to reduce the incidence, morbidity and consequent mortality associated with the sequelae of liver associated conditions brought about by chronic HBV infection. Whether or not this target is achievable should become clearer in the coming decades.

Although significant positive impact is anticipated, mixed reports have thus far emanated from areas where vaccine implementation has reached its third decade (Chang et al., 2005). The biggest drawbacks to eradicating childhood HBV infection which carries a greater risk of chronicity, are limitations in preventing perinatal transmission from highly infectious mothers and some, though quite limited, reports of vaccine failure (Chang et al., 2005). Thus, despite the existence of the HBV vaccine, the potential elimination of the virus as a major cause of HCC is not a foregone conclusion; particularly with widespread reports of resistance to lamivudine monotherapy (Bartholomew et al., 1997; Zollner et al., 2001), the most widely approved first line treatment for chronic HBsAg carriage, also active against the human immunodeficiency virus (HIV). Other therapies against chronic carriage such as tenofovir, adefovir and entecavir exist with tenofovir showing significant promise in its long-term suppression of HBV viral load, activity against HIV and low rates of resistance (Cooke et al., 2010; van Bommel et al., 2010). The long-term ability of these anti-viral treatments to prevent the development of HCC is also controversial as discussed in this recent review (Thursz and Brown, 2011). These factors, coupled with the high costs of these newer therapies, lack of local approval and will to incorporate them into routine treatment programmes

(Thursz et al., 2010) mean that they are so far not widely available or recommended in the developing world (Holbrook, 2008; Leung, 2002).

Sporadic reports of low to medium levels of vaccine escape mutations and BTI in HBV endemic areas have also been documented (Carman et al., 1990; He et al., 2001; Karthigesu et al., 1994). Research suggests though that the risk of BTI changes according to HBV genotype with opportunistic infections becoming more likely later in life when anti-HBs levels wane (Mendy et al., 2008). These factors combined with existing, unresolved risk factors such as HCV infection, AF exposure, alcohol abuse as well as the emergence of newly recognized predisposing factors for liver cancer such as diabetes and obesity (Regimbeau et al., 2004; El-Serag and Rudolph, 2007) mean that many challenges lie between the current situation and any hopes of significantly reducing worldwide HCC mortality. This changing landscape of liver disease treatment and cause will require researchers to where possible anticipate prospective bottlenecks and concerns in order to offer timely and relevant solutions.

4.2 Independent ELISA Validation and Biochemical Characterization

One such area where advances are urgently needed is in LC & HCC diagnostics – which form the key focus of this thesis. In contrast to previous chapters which highlighted the consistent differential expression of select proteins and their calculated diagnostic potential; here the prime focus will be on i) validating these findings in an independent subject group and ii) homing in on observed changes

at the structural level in order to see if there are any disease associated signatures which may be exploitable for the development of screening or diagnostic tests. One of the key protein modifications associated with liver disease are changes within glycoprotein subclasses (Blomme et al., 2009; Turner, 1992). N-linked glycosylation, as opposed to O-glycosylation which occurs at the carboxyl groups of Ser or Thr is the predominant class of glycoproteins expressed in human plasma. Much effort has been directed toward enriching for this sub-class of proteins and studying its alterations in liver disease. As the liver is a primary site of synthesis for many glycosylated proteins, it is expected that within the context of CLD, changes in the trends of PTMs of these proteins are likely to occur. Such discernible changes could serve as a basis for the design of more targeted and specific diagnostic tests for the differential diagnosis of progressive liver conditions.

Specific lectin based methods have been utilised in the enrichment of glycosylated proteins from the plasma of the initial discovery GLCS subjects as well as a pilot validation cohort from Jos, Nigeria. The depleted, enriched, desalted and concentrated samples were applied to 1 and 2D separation protocols followed by western blotting and the ensuing results and trends presented and analysed. These were applied to explore Apo A1, α1AT, CC3 and HPX in greater structural detail.

Independent ELISA based validation of the overall expression trends observed for the shortlisted putative markers was conducted on the sub-regional subjects and the two sets of results compared and contrasted.

4.1 Materials & Methods

Acetone was purchased from Fisher Scientific; the Enhanced Chemiluminescence (ECL+) western blot development kits were purchased from GE Healthcare, Amersham; Neuraminidase was purchased from Calbiochem, EMD Chemicals, USA; NP-40 was purchased from New England BioLabs® Inc.; the Pierce BCA Protein Assay Kit was purchased from Thermo SCIENTIFIC; Protein A Sepharose was purchased from Invitrogen®; Ultra Pure ProtoGEL™ was purchased from national diagnostics, USA; the Polyvinylidene (PVDF) Immobilon-P® membrane was purchased from Millipore, USA; ROTI® Block was purchased from Carl ROTH, Germany. The Wicks, pH 3-10 Ampholyte Buffer and Mineral oil were purchased from Agilent Technologies, USA; DTT and NaCl were purchased from Alfa Aesar®. The IPG ph 3-10 ReadyStrips, Criterion™ XT Precast Gels and Criterion™ Empty Cassettes were purchased from BIO-RAD. Agarose, Ammonium Persulfate (APS), BPB, $CaCl_2$, Glycerol, HEPES, Iodoacetamide, Methyl α-D-mannpyranoside, 3-(N-morpholino)propanesulfonic acid (MOPS), SDS, Tetramethylethylenediamine (Temed), Trizma base and Urea were purchased from SIGMA®. BME, $C_2H_3NaO_2$, 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), Convavallin A (Con A) from *Canavalia enstformis* (Jack Bean), EtOH, Glycine, HCL, Kodak Scientific Imaging film, MeOH, $MgCl_2$, $MnCl_2$, MOPS, Na-Deoxycholate (Na-

DOC), Tricholoroacetic Acid (TCA) and Tween®20 were purchased from SIGMA-ALDRICH®. Primary antibodies for Apo A1, α1AT, CC3 and HPX were purchased from abcam® (Cambridge, UK); HRP conjugated anti-Rabbit and anti-Goat secondary antibodies were purchased from Dako (Denmark).

4.1.1 Study Groups

Two distinct plasma sample sets were used in these investigations, one recruited from the Gambia; the primary source of samples for this study and the second from Nigeria; used as a regional validation cohort. Briefly, the GLCS samples were collected as part of a five-year liver cancer case-control study with subjects recruited from three major healthcare providers in the Gambia; RVTH, Bansang Hospital and the MRC Unit, Fajara.

The Nigerian samples were sourced from regional collaborators in a study overseen by colleagues at Imperial University, London.

4.1.2 Subject Selection Method & Diagnosis

4.1.2.1 Gambia Liver Cancer Study

In the Gambia, eligible and willing individuals were consented into the GLCS. EDTA plasma was collected from subjects at the three named centres and working aliquots stored at -20°C with the remaining plasma divided and frozen at -80°C. The samples from Bansang Hospital were separated immediately following bleeding and frozen at -80°C on site pending transport on dry ice to the MRC Serology Lab at Fajara. A combination of biochemical assays for AFP and liver

enzymes as well as clinical examinations, US and liver biopsy were used to classify subjects as controls, LC or HCC (figure 2.1).

Scientific and Ethical approvals (MRC SCC No. 1154) were granted for single aliquots of 0.5-1ml to be removed from the frozen archives for utilization in this study. The GLCS subjects were not tested for any viral infections aside from HBV and HCV.

4.1.2.2 Jos University Teaching Hospital Case-Control Subjects, Nigeria

The ethics committee at Jos University Teaching Hospital (JUTH) in Plateau state, Nigeria approved the study proposal; sample collection was conducted and completed in 2008. Patients visiting outpatient clinics with suspected diagnoses including HCC, LC or liver metastases were identified and asked for informed consent. Once given, recruited subjects were bled into plasma EDTA tubes. Adult volunteers with no history of liver disease and of Nigerian origin were enrolled as healthy controls. In total, plasma aliquots for 55 subjects were donated by the Taylor-Robinson group at Imperial for validation experiments with these classified as either Nigerian healthy controls (NN), Nigerian Asymptomatic Carriers (NASC), Nigerian cirrhotics (NCirr) or Nigerian HCCs (NHCC). The categorizations were based on tests for HBV, HCV and HIV virus status, AFP, biochemical liver indices, US, or Computed Tomography (CT), endoscopy and or liver biopsy.

Subjects who tested negative for the hepatitis viruses as well as HIV and had results not indicative of liver disease from biochemical assays for AFP, creatinine, urea, ALT, alkaline phosphatase (ALP) and total bilirubin (Bili T) were taken as healthy controls. Asymptomatic carrier subjects were identified as those positive only for HBV with no evidence of cirrhosis following liver biopsy (conducted on 80% of NASC group). Those without biopsy had US done which showed no evidence of early stage LC. The NCirr group within this population were all HBV positive with evidence of cirrhosis seen during US as portal hypertension. Only one subject had their LC diagnosis confirmed by liver biopsy. The HCC group had a >95% HBV positive rate with the clinical diagnosis of HCC based on AFP, US, CT and for one subject, liver biopsy.

The HCCs were assessed using the Okuda staging system (Okuda et al., 1985; Pons et al., 2005) which measures parameters related to liver function, presence of ascites and area of liver affected by tumour. All the HCC's in the JUTH subjects were identified as Okuda stage II or higher suggesting these tumours to at least be moderately advanced.

4.1.3 IgG Depletion

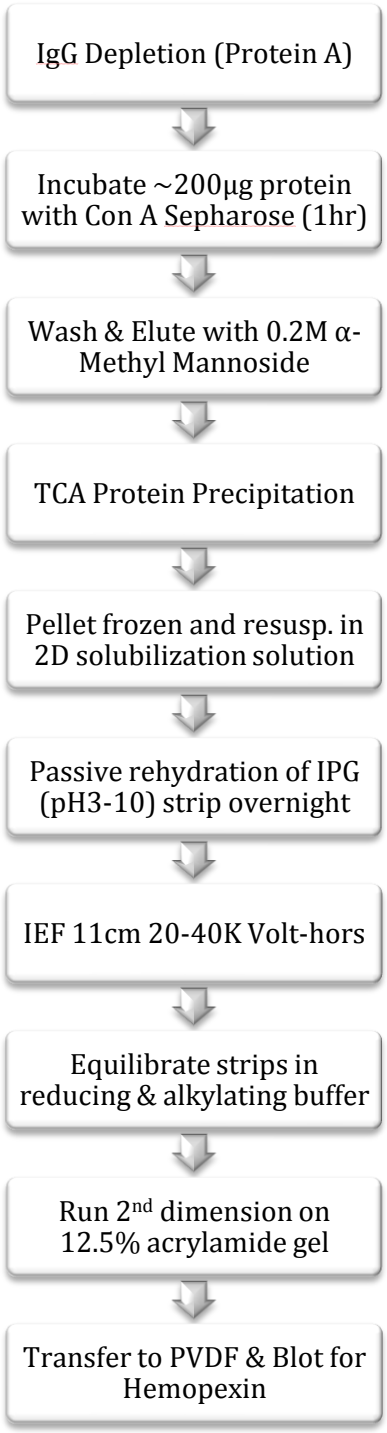


Figure 4.2: Protocol of glycoprotein enrichment. Summary of workflow detailing major steps undertaken from plasma sample depletion, glycoprotein enrichment to first and second dimensional separations and HPX detection by western blot.

All samples for 1 and 2Dimensional separation experiments underwent the same preparatory steps before loading on to either Immobilized pH Gradient (IPG) strips for first dimensional separation or pre-cast multi-well SDS-PAGE gels. Samples were firstly subjected to IgG depletion as previously described (section 3.1.4) with the IgG depleted supernatant transferred to a fresh eppendorf tube. The Protein A resin with bound IgG was given a further wash with a 100mM NaCl, 10mM HEPES – pH 7.4 solution and the supernatant collected and added to the IgG depleted plasma fraction.

4.1.4 Glycoprotein Enrichment with Concanavalin A

At the optimization phase of this assay, two techniques of glycoprotein enrichment were compared to determine which one gave higher protein yields. The two methods compared were column-assisted enrichment versus direct sample incubation with Con A sepharose for 1 hour at $4^{o}C$. It was found that the latter gave greater than four-fold yield and was thus decided upon for subsequent experiments (table 4.1).

The IgG depleted samples were taken forth for glycoprotein enrichment using Con A sepharose. The protocol used was optimised from a publication by Alonzi et al. (Alonzi et al., 2008). Prior to sample incubation, the Con A lectin underwent a series of wash and equilibration steps detailed as follows. One hundred microliters of well mixed Con A sepharose was twice washed with millipore H_2O followed by a single wash in a 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM $MnCl_2$ buffer and

lastly a two cycle wash in 50 mM Tris/HCl pH 7.2. The beads were spun down at 5000G for 30 seconds in between each step with all excess buffer removed prior to sample addition. With the Con A sepharose fully equilibrated and washed, 200µl of IgG depleted plasma was added and co-incubated at 4°C for 1 hour with gentle agitation. Post incubation, the enriched glycoproteins remain tightly bound to the lectin allowing for the glycoprotein-depleted supernatant to be removed. The beads were then washed once with 50 mM Tris/HCl pH 7.2 and millipore H_2O with 30 second spin-downs in between each step to remove excess buffer. To release bound glycoprotein structures, hot Methyl α-D-mannopyranoside at 70°C was added to the Con A lectin for elution and sample collection.

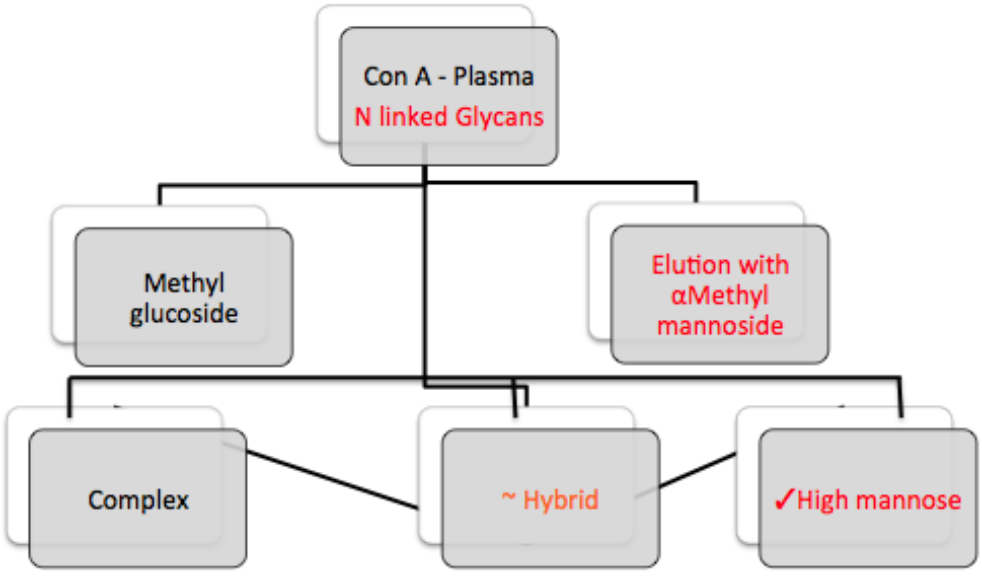


Figure 4.3: Chart showing the multiple enrichment steps applied to subject plasma during glycoprotein enrichment experiments. Plasma glycoproteins are naturally rich in N-Glycan structures with α-methyl mannoside known to preferentially elute high mannose and some hybrid sugar structures (Cummings and Kornfeld, 1982; Liscum et al., 1983).

Units	Whole Plasma	Eluate-Beads	Eluate-Column	IgG Depleted
µg/ml	41127	2413	508	6100
mg/ml	41.127	0.241	0.051	6.100

Table 4.1: Results from Bicinchoninic Acid (BCA) protein quantitation assays for a representative plasma sample used to compare protein yields using alternate glycoprotein enrichment protocols. Overall protein amounts were also determined for whole and IgG depleted plasma.

4.1.5 Bicinchoninic Acid Protein Quantitation

This was done on representative sample pools in order to determine the protein concentrations in:

- Whole plasma
- IgG depleted plasma
- IgG depleted and Con A enriched plasma by column
- IgG depleted and Con A enriched plasma by direct sepharose incubation

The assay was conducted according to manufacturer's instructions (Pierce® BCA Protein Assay Kit, Thermo Scientific). In brief, the key steps were the preparation of a series of Bovine Serum Albumin (BSA) standards plus a blank and a working reagent made by mixing BCA reagents A and B in the ratio 50:1. All standards and unknown samples were pipetted into a 96 well microplate in duplicate followed by the addition of the well-mixed working reagent. The plate was then incubated at 37^{o}C for 30 mins, allowed to cool and the absorbance measured at 562nm on a Safire^{2} Tecan (Invitrogen) plate reader.

4.1.6 Protein Precipitation – TCA/DOC

In order to prepare the samples for resuspension in appropriate solubilisation buffers, the enriched glycoproteins were precipitated using TCA, Na-DOC and acetone in an adapted protocol (Bensadoun and Weinstein, 1976). Briefly, the glycoprotein enriched plasma samples eluted with Methyl α-D-mannopyranoside were treated with 2% w/v Na-DOC, vortexed and left to stand for 15 minutes at room temperature. To this mixture was added 24% w/v TCA followed by a vortex and spin down by centrifugation for 30mins at 3000rpm at 4°C. The supernatant from this step was carefully aspirated ensuring no disturbance of the protein pellet at the base of the tube. The pellet was then washed by centrifugation for 15mins with cold acetone (-20°C) in order to remove excess TCA. The washed pellet was frozen at -80°C pending resuspension in 1 or 2DE sample buffer.

4.1.7 Sample Analysis by SDS PAGE & Immunoblotting

Samples analysed by this method had the glycoprotein enriched, IgG depleted protein pellets resuspended in 3X reducing sample buffer (30% v/v Glycerol, 15% v/v BME, 0.9g w/v SDS, 37.5% v/v Upper Tris (4X stock: 500mM Trizma base, 2g w/v SDS to 500ml, pH 6.8) with trace BPB and Millipore H_20 to 10 ml). Ten microliter sample aliquots were loaded into each gel well and the samples run on precast 26 well 4-12% Bis-Tris, 15µl 1.0mm Criterion™ XT gradient gels, at 200V for 1hour.

The gels were then removed from their holding cassettes and transferred overnight at 15 volts (V) onto a PVDF membrane with a 0.45µm pore size

(Immobilon-P®, Millipore) at 4°C in 1X Blotting Transfer Buffer (BTB) (10X stock: 100mM Tris Base, 1M glycine, Millipore H_2O to 1L) with 20% v/v MeOH. On day two, the membranes were removed and immediately immersed into 1X ROTI™ blocking solution for 1 hour. This was followed by a brief wash in 1X Tris Buffered Saline with Tween®20 (TBST) (10X stock: 200mM Tris base, 1.4M NaCl, 0.1% v/v Tween®20, pH 7.6) and incubation with the primary antibody diluted in 1X ROTI™ block for 30 mins at the following ratios:

- Goat Polyclonal to Apo A1 (Abcam, ab7613) – 1° Ab: 1:1500 2° Ab: 1:4000
- Goat Polyclonal to α1AT (Abcam, ab2113) – 1° Ab: 1:4000; 2° Ab: 1:15,000
- Rabbit Polyclonal to HPX (Abcam, ab48135) – 1°: 1:2000; 2° Ab: 1:5000
- Rabbit Polyclonal to CC3 (Abcam, ab48611) – 1° 1:1000; 2° 1:2500

Following incubation with the primary antibody, the membranes were washed three times in 1X TBS-T and incubated with the secondary antibodies as follows:

- HRP Rabbit anti Goat secondary; for Apo A1 1:4000; α1AT α1AT 1:15,000 (Dako, P0160)
- HRP Goat anti Rabbit secondary; for HPX 1:5000; CC3 1:2500 (Dako, P0448)

The membranes were then washed a final three times and developed by enhanced chemiluminescence (ECL) using ECL+ kits as per manufacturer's instructions. The bands were visualised on Kodak scientific imaging films using the Compact X4 Xograph Imaging system (Xograph Healthcare Ltd.).

4.1.8 Sample Analysis by 2DE & Immunoblotting

Well defined sample pools from the JUTH and GLCS subject groups were obtained following IgG depletion, glycoprotein enrichment and TCA precipitation and had their resultant protein pellets resuspended in sample solubilisation solution (8M Urea, 50mM DTT, 4% w/v CHAPS, 0.2% v/v pH 3-10 carrier ampholytes and trace BPB). The mixtures were sonicated and vortexed in order to ensure maximum solubilisation. Approximately 70µg of protein in 185µl of solubilisation solution was pipetted into a clean rehydration tray and an 11cm pH 3-10 linear/non-linear IPG strip gently placed over the liquid to allow for passive gel rehydration. After an hour, the channel was covered in mineral oil and allowed to incorporate the solubilised proteins overnight, at room temperature. This step was essential for optimal sample entry into the IPG gel.

On day two, the fully hydrated strips were assembled on a focussing tray according to manufacturer's instructions and transferred to a Bio-Rad PROTEAN® IEF cell or Agilent OFF-GEL fractionator for isoelectric pI based separation in linear ramping mode for a minimum of 20,000 Volt-Hours (V-hr). Dampened wicks were placed at the anode and cathode ends of each of the IPG strips to absorb excess salts. Upon the completion of this step, the gel strips with embedded sample were immersed successively for 10 minutes each into reducing and alkylating base buffers (6M Urea, 2% v/v SDS, 0.05M Tris-HCL pH 8.8 gel buffer, 20% Glycerol with either 2% w/v DTT or 2.5% w/v Iodoacetamide). Each strip was then loaded to run on a 12.5% Bis-Tris in-house SDS PAGE gel (41% v/v Acrylamide, 24.9% v/v Upper Tris, 0.5% v/v APS, 0.05% v/v Temed with Millipore H_2O to 7.0385ml) at 200V for 1 hour

for second dimension separation by molecular weight, ensuring no air bubbles were trapped between the gel front and strip. To secure the IPG strips, 1% w/v molten agarose with trace amounts of bromophenol blue to aid tracking was used to seal the gel front. The subsequent steps including overnight transfer onto PVDF membrane and immunoblotting for HPX were performed as described above (section 4.1.7).

4.1.9 Neuraminidase Digestion

Immunoprecipitated protein pellets were resusupended in 0.1% w/v SDS, briefly sonicated and heated for 5 mins at 95^{o}C. The tubes were allowed to cool, spun down and the enclosed sample separated into two equal aliquots; one as control and the other to digest. To each tube was added, with intermittent vortexing; 10% NP-40, 0.5M sodium acetate ($C_2H_3NaO_2$); 1µl 1IU/ml neuraminidase enzyme substituted with water in control; and millipore H_2O to a final volume of 12µl. The reaction mixtures were then incubated at 37^{o}C overnight and subsequently stopped on day 2 by freezing at -80^{o}C prior to addition of appropriate buffer.

4.1.10 ELISA Assays

These were performed following the same methods described in Chapter 3 (sections 3.1.8 & 3.1.9) on the Jos University Teaching Hospital Cohort for the proteins α1AT diluted 1:50,000; Apolipoprotein A1 1:500 (new lot), HPX 1:30,000 and CC3 1:400. All extrapolations and analyses were conducted using the Graph Pad Prism (version 5) statistical package.

4.2 Results

4.2.1 Chronic HBV infection is strongly associated with chronic liver disease development in JUTH subjects.

To an even greater degree than that observed in the GLCS, chronic HBV infection was seen to be strongly associated with the development of LC and HCC in the subjects from Nigeria. Other risk factors for HCC and LC development noted from both subject populations (tables 3.1 & 4.2) are being of male sex, increased age, elevated ALT and AFP levels. A low-lying trend of HCV infection is also observable in the NHCC group although it is slightly above the estimated prevalence for West Africa (2004). No assessment of aflatoxin exposure was conducted in this subject population.

Variable	NCon (n = 10)	NASC (n = 18)	NCirr (n = 6)	NHCC (n = 21)
Biopsy (%)	0	77.7	16.7	4.8
Clinical Exam (No.)	10	18	6	21
Ultrasound (%)	100	-	83.3	100
AFP > 100ng/ml (%)	0	0	16.7	80.9
Mean ALT (IU/L)	22.8	29.6	42	73.3
ALT > 56 IU/L (No.)	0	1	2	8
Male (%)	50	72.2	83.3	61.9
Female (%)	50	27.8	16.7	38.1
Mean Age (years)	41	38.1	39.2	47.7
HbsAg Positive (%)	0	100	100	90.5
HCV Positive (%)	0	0	0	4.8

Table 4.2: Key clinical parameters of the Jos University Hospital Cohort.

4.2.2 Previously reported protein expression trends validated in independent West African subjects.

ELISA assays which measured the expression levels of α1AT, Apo A1, CC3 and HPX in the JUTH subjects, showed trends consistent with those seen in the GLCS (figures 3.4, 4.4 & 4.5). In the latter population, α1AT expression was seen to discriminate best between controls and HCC subjects based on AUC calculations. A similar trend was seen in the Nigerian subjects which showed α1AT to give an AUC of 0.7857 (table 4.3) when comparing overall expression between control and HCC subjects. This was marginally lower than that for NCirr vs. NHCC discrimination with this slight discrepancy in trend likely due to the small number of cirrhotic patients compared to HCC's within the JUTH subject group. The marked decline in ApoA1 plasma secretion previously observed was once again seen, resulting in a peak AUC of 0.7667 between the NN and NHCC groups. Overall, the same trend witnessed in the Gambian subjects was mirrored in those from the JUTH with a marked down-regulation of Apo A1 as liver disease progressed towards hepatocarcinogenesis. The consistency between the two populations was also maintained for CC3 and HPX expression with the highest discriminatory potential for the former seen in the development of HCC from a background of LC. In the Nigerian subjects, this was seen with a peak AUC of 0.8571. With HPX, consistent with previous results, levels of expression were seen to drop in patients with LC and "bounce back" in those with HCC. The calculated AUC for the development of LC in the Nigerian subjects suggests excellent discriminatory potential for this protein with a statistically significant (p.001146) AUC value of 1.000.

Protein & Sample Numbers Tested	NN vs. NHCC	NN vs. NCirr	NCirr vs. NHCC
ApoA1 NN: 10, NASC: 18, NCirr: 6 HCC: 21	AUC: 0.7667 (95% CI: 0.6020 too.9313) p=0.01800	AUC: 0.6167 (95% CI: 0.3154 to 0.9179) p=0.4477	AUC: 0.6587 (95% CI: 0.3987 to 0.9187) p=0.2435
AAT NN: 10, NASC: 18, NCirr: 6 HCC: 21	AUC: 0.7857 (95% CI: 0.6039 to 0.9676) p=0.01126	AUC: 0.5833 (CI: 0.2647 to 0.9019) p=0.5876	AUC: 0.8016 (95% CI: 0.6360 to 0.9672) p=0.02672
CC3 NN: 10, NASC: 18, NCirr: 6 HCC: 21	AUC: 0.6238 (95% CI: 0.4200 to 0.8277) p=0.2720	AUC: 0.8167 (95% CI: 0.5921 to 1.042) p=0.03937	AUC: 0.8571 (95% CI: 0.6985 to 1.016) p=0.008705
HPX NN: 10, NASC: 18, NCirr: 6 HCC: 21	AUC: 0.8048 (95% CI: 0.6455 to 0.9640) p=0.006864	AUC: 1.000 (95% CI: 1.000 to 1.000) p=0.001146	AUC: 0.7937 (95% CI: 0.5594 to 1.028) p=0.03098

Table 4.3: Summary of sample numbers tested per protein along with AUC's, confidence intervals and associated p-values for JUTH subjects.

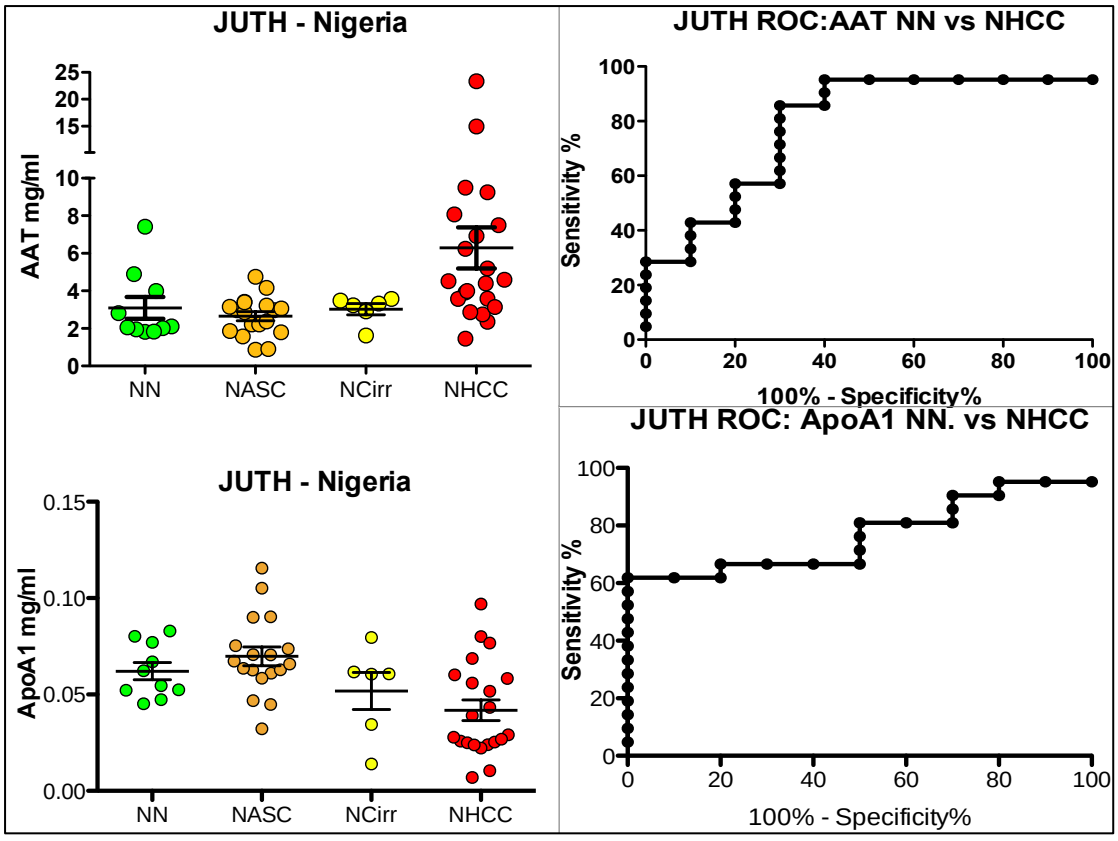


Figure 4.4: Figure showing expression trends and most discriminatory ROC curves for α1AT & Apo A1 in the Nigerian Subjects.

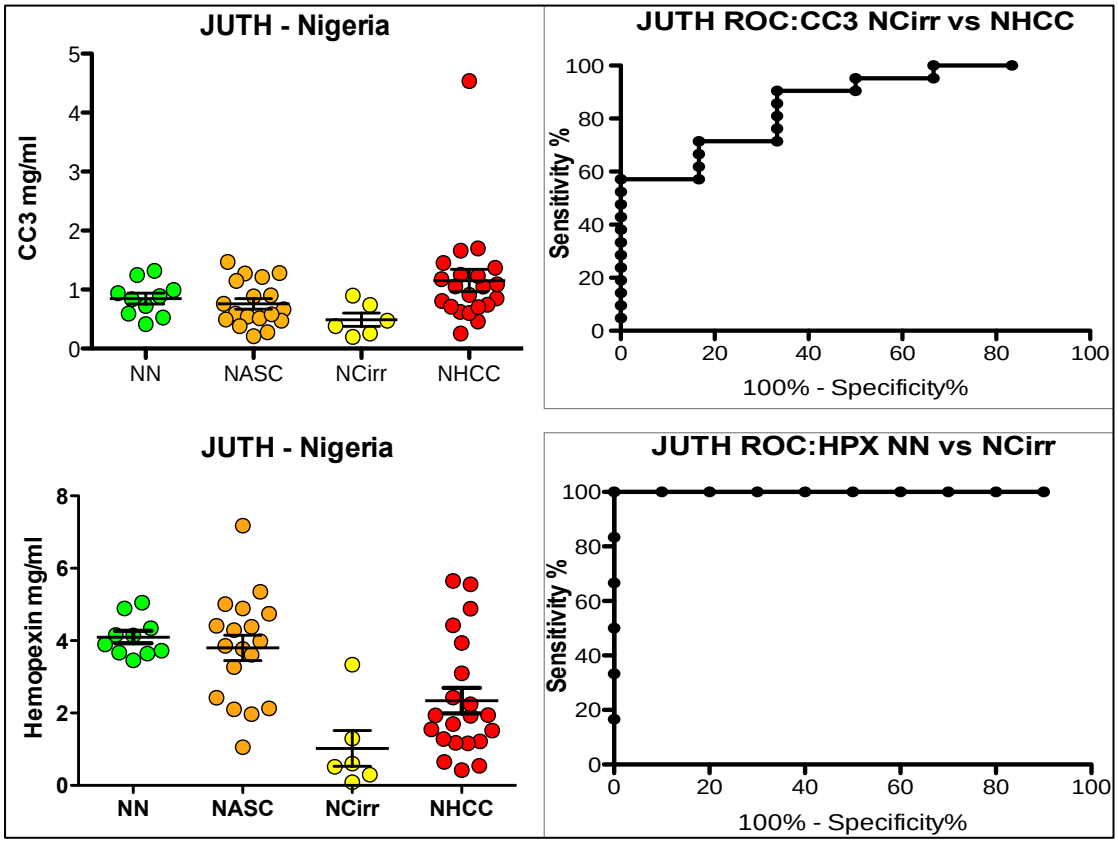


Figure 4.5: Figure showing expression trends and most discriminatory ROC curves for CC3 & HPX A1 in the Nigerian Subjects.

4.2.3 Glycoprotein enriched plasma from individual subjects indicate differences in glycosylation patterns for α1AT, HPX, CC3 & Apo A1; including some HBV related trends.

Immunoglobulin G depleted, glycoprotein enriched plasma from GLCS subjects negative for p53 mutations with mixed HBV status (table 4.4) were probed by polyclonal antibodies to the four putative biomarkers under validation. For CC3, two processed bands were seen at approximately 30 and 50 kDa with disease-correlated differences observed. In the control region, a lower intensity expression of the 30 kDa band in HBV negative subjects as compared to those

expressing the HBV surface antigen was seen. This is in contrast to the marked down-regulation of both the 30 and 50 kDa bands in all HCC patients regardless of virus status. Complement component 3 is a highly processed protein, with at least three known glycosylation sites (Bunkenborg et al., 2004; de Bruijn and Fey, 1985; Liu et al., 2005) thus making it difficult using 1D separation to identify its specific peptide chains, it does appear however that the lower molecular weight band expressed in these subjects correlates with the 302 amino acid length Complement C3d fragment. With only one known glycosylation site, no significant mass shift from the expected molecular weight of 28 kDa was seen for Apo A1. The most significant changes observed were in its levels of expression between HBV positive and negative subjects within the HCC group. Apo A1 expression was seen to be lowered in HCC patients with chronic HBV infection and higher in all but one of those with HBV negative HCC. Greater expression of glycoprotein-enriched Apo A1 in subjects diagnosed with LC compared to controls and those with HCC was also seen. α1AT expression from these experiments is seen to be comparable in control and LC subjects with once again possible virus induced differences in its expression observable in the HCC subject group. A distinct up-regulation in the expression of the glycosylated form of this protein was observed in virus negative individuals within the eight HCC patient plasma. A uniform upward shift in molecular weight from its expected 52 kDa was also observed for α1AT independent of the presence or absence of liver disease. HPX, a highly glycosylated protein with 6 known glycosylation sites (Takahashi et al., 1984) and an unmodified molecular weight of 52 kDa was seen to be differentially

glycosylated between the control and CLD groups with the unmodified protein predominantly expressed in liver disease free subjects.

01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
HBV+	HBV+	HBV+	HBV+	HBV-	HBV-	HBV-	HBV-	HBV+	HBV+	HBV+	HBV+	HBV-	HBV-	HBV-	HBV-	HBV+	HBV+	HBV+	HBV+	HBV-	HBV-	HBV-	HBV-
-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-

Key: -/- = Negative for p53 mutation and HCV
Controls LC HCC

Table 4.4: Outline of gel layout and individual sample profiles used in immunoblot experiments for Apo A1, α1AT, CC3 & HPX, ladder lanes not included.

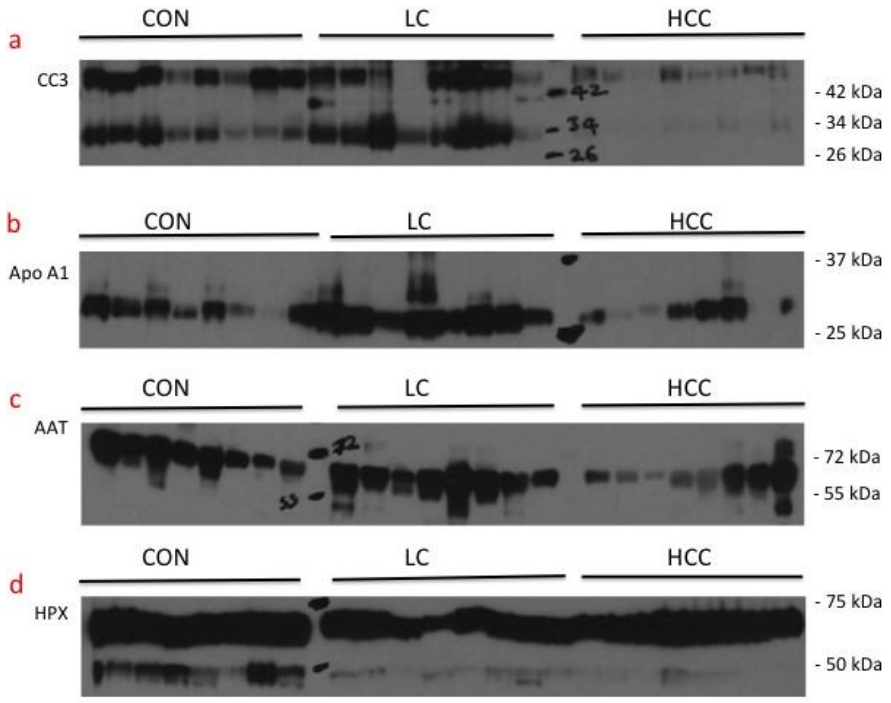


Figure 4.6: Immunoblot results for GLCS plasma samples probed by antibodies to CC3 (a), Apo A1 (b), α1AT (c) & HPX (d) in p53- and HBV+/- subjects.

4.2.4 2D pooled plasma profiles from GLCS and JUTH subjects indicate differences in enriched N-glycosylated high mannose structures exploitable as highly specific discriminatory tests for liver disease.

For the GLCS subjects where there was a significantly greater total sample number (339), sample pools based on the inclusion or exclusion of HBV, HCV and p53 mutation status as suspected primary aetiological agents for LC or HCC were made. Results from 2DE immunoblot analyses of these pools show successful enrichment of high mannose, N-linked glycosylated structures for HPX using Con A sepharose and α methyl mannoside (figure 4.3). Within this glycoprotein subclass in virus and p53 mutation free subjects, there appears a distinct pattern of HPX migration toward the acidic end of the pH gradient strips. This trend for the GLCS subjects is unique to liver disease, mutation and virus free controls. A similar but less marked migration of HPX toward the acidic end of the linear IPG strip is also seen in the JUTH healthy control pool (compare figures 4.7a & 4.8a). Within the LC GLCS subjects, the presence of low molecular weight HPX appears to correlate with HBV & HCV positivity and is absent from all HCC 2DE films. Other notable distinctions include the overarching trend for glycoprotein enriched HPX in both the Nigerian and Gambian HCC subjects to resolve to the mid pH gradient with little evidence of the presence of low molecular weight spots.

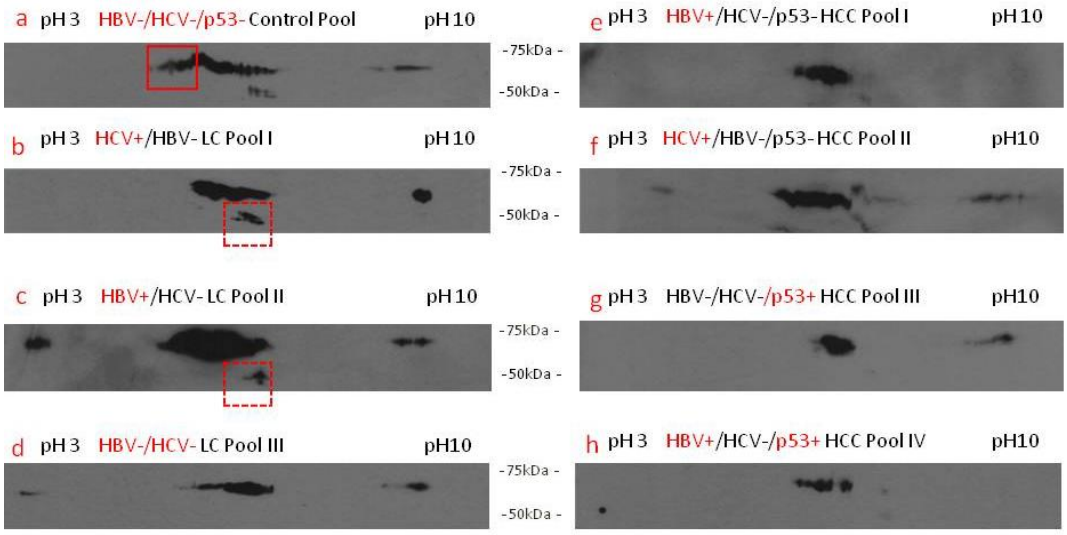


Figure 4.7: 2 Dimensional Immunoblots of GLCS samples pooled according to key aetiological profiles, separated on pH 3-10 Linear IPG strips and probed for the HPX protein. Panels a-h are profiled as follows: a) HBV/HCV/p53 negative control pool; b-d: LC pools representing HCV+, HBV+ and double negative subjects; e-h: HCC pools representing HBV+, HCV+, p53+ and dual HBV & p53 positive subject pools, respectively.

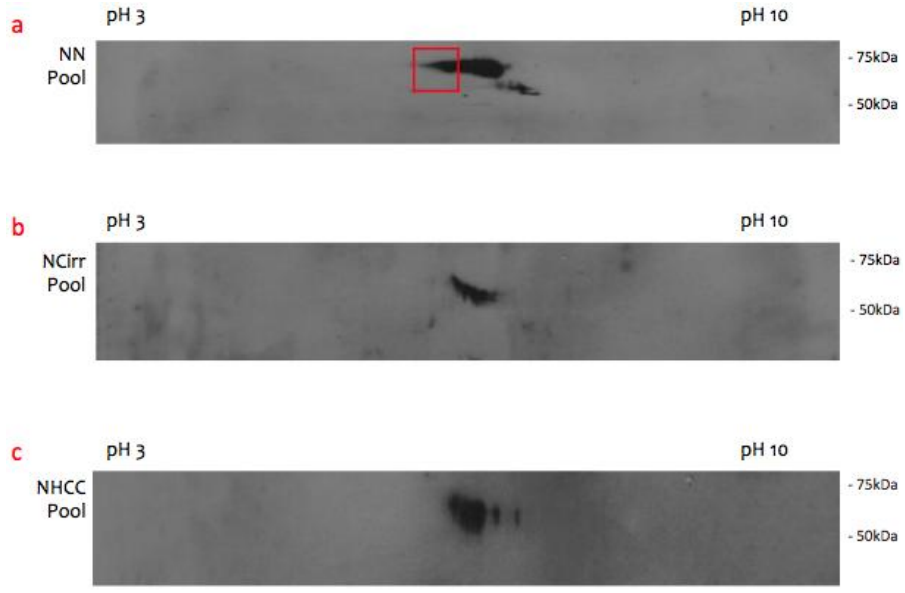


Figure 4.8: 2 Dimensional Immunoblots of JUTH samples pooled according to clinical diagnoses, separated on pH 3-10 Linear IPG strips and probed for the HPX protein. Panels a-c represent the three condition specific plasma pools of healthy, cirrhotic and HCC subjects – respectively.

4.2.5 Non-linear separation of Hemopexin across immobilized pH gradient highlights absence of low molecular weight spots which appear with neuraminidase treatment.

Immunoblotting experiments conducted on HCC, LC and sialidase treated and untreated control pools of 2DE separated samples showed some notable and potentially exploitable results. Repeat runs on non-linear IPG strips were conducted as they offer greater resolution of the mid pH ranges resulting in more detailed protein spot analysis within the crucial pH 5 – 7 region. In the GLCS subjects, it was seen that the absence of lower molecular weight HPX bands in the HCC pools was maintained across the two focussing conditions (figures 4.7 e-h & 4.9 a-d) despite the appearance of other discrepant spots. In the JUTH plasma pools, the isoelectric focussing of proteins on the non-linear IPG helped to better resolve the sequence of 50 – 60 kDa spots which reduce with declining liver function (figure 4.10 a-c). Treatment with a sialic acid cleaving enzyme in both populations appeared to produce similar effects; resulting in the appearance of reduced molecular weight, well resolved spots at the acidic end of the major observable HPX structure (figures 4.9 f & 4.10 d). Enzyme negative treatments were made for the respective control pools but the films developed suggested sample loss during experimentation (figures AF 1.4 a-c). However, comparisons between the neuraminidase treated samples and the natively expressed glycoprotein enriched plasma suggest the presence of a sub population of sialic acid rich HPX sugars in both the JUTH & GLCS liver disease free control groups.

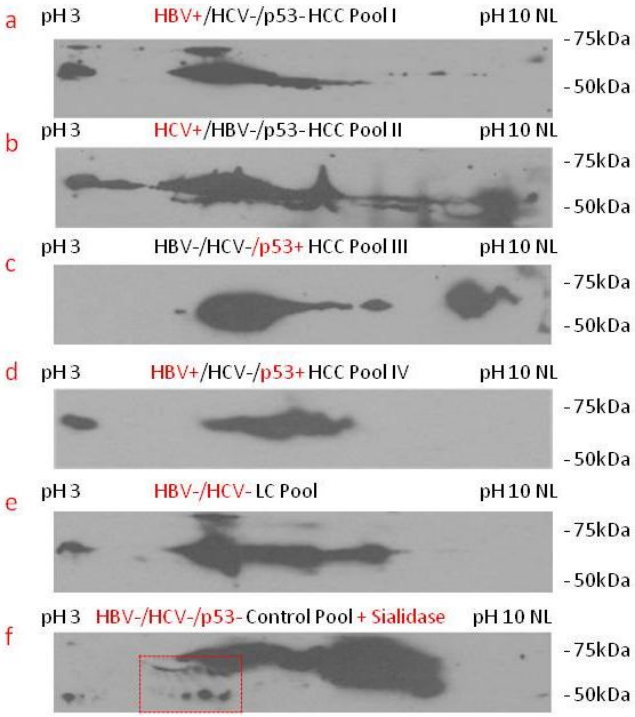


Figure 4.9: 2 Dimensional Immunoblots comparing HPX probed GLCS sample pools, separated on pH 3-10 Non-Liner IPG strips with a sialidase treated control. Panels a-d represent plasma pools from HCC subjects with the respective profiles of HBV+, HCV+, p53+ and dual HBV & p53 positivity; e represents an LC pool with dual negativity for HCV and HBV and f a sialydase treated control pool negative for HBV and HCV infections as well as a p53 mutation.

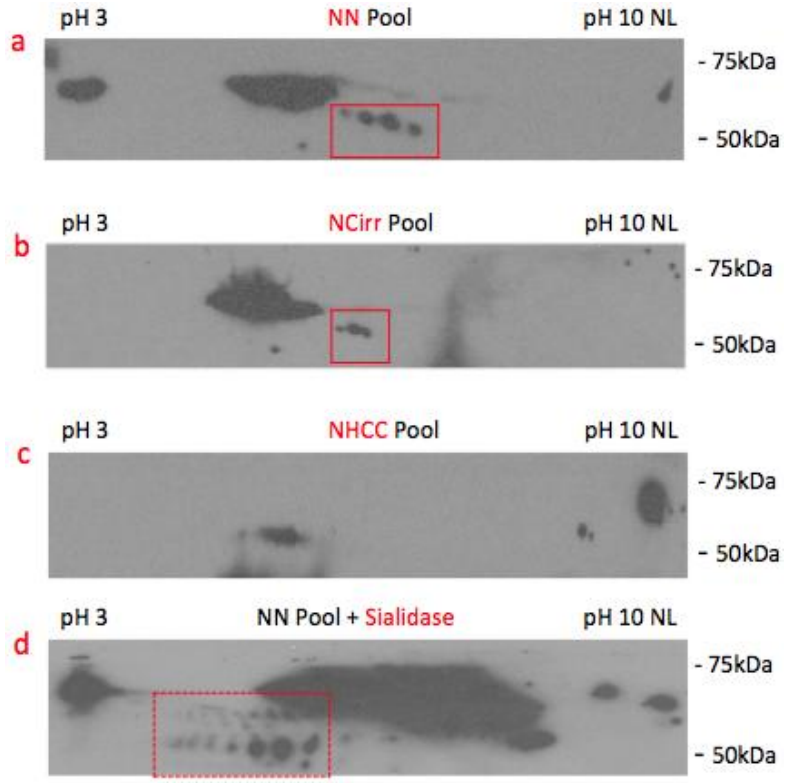


Figure 4.10: 2 Dimensional Immunoblots comparing HPX probed JUTH sample pools, separated on pH 3-10 Non-Liner IPG strips with a sialidase treated control. . Panels a-c represent the three condition specific plasma pools of healthy, cirrhotic and HCC subjects – respectively. Panel d represents the control pool subjects treated with a sialidase enzyme.

4.3 Discussion

Biomarker discovery and identification can be approached from several angles with a key element being the ability to replicate trends and findings using independent methods and subjects with well-characterized disease. One of the key objectives from the inception of this study was to independently validate any putative protein biomarkers suggested by the discovery GLCS subjects in an independent study group. The ability to replicate such trends and signatures is key to increasing confidence in the potential discovery of robust markers. An essential feature required for any proposed marker from this study was for it to

be robust enough to perform and be measurable in the conditions of the developing world. Both sample sets utilised in this thesis were collected with standard steering committee approved protocols not specialized for mass spectrometry based proteomic analyses. Yet, the results obtained have been reliable and with selective filtering generated a viable shortlist for validation. It is thus envisaged that all the putative markers from this exercise have some degree of "in-built" robustness, which could confer an advantage in field-based investigations in the developing world. In this chapter, investigations looking at the molecular structures of a subset of these proteins were conducted on α1AT, CC3, HPX & ApoA1. Based on encouraging results seen with quantitative ELISA and qualitative western blot data as well as its relatively uncomplicated expression profile on 1D PAGE – HPX was taken forth for additional investigations by 2DE characterization in two independent subject groups.

Quantitative measurements of all four shortlisted proteins conducted by ELISA in the JUTH subjects show expression trends largely identical to those seen in the Gambian group. These experiments independently confirm that the secretion of ApoA1 into plasma is increasingly inhibited during LC and HCC development; α1AT levels are marginally increased with LC followed by a marked up-regulation with progress to HCC. CC3 & HPX show alternative trends to the afore mentioned markers suggesting them to possess signatures more exploitable for the diagnosis of LC or HCC from a cirrhotic background than that from a healthy non-cirrhotic liver. It is worth noting though that the distinction of the first two groups would be of greatest clinical relevance as up to 80 – 90% of HCC's have

been reported to develop from a background of LC (Constantin et al., 2010) making its development a primary risk factor for HCC onset. The differences in discriminatory power shown by these proteins suggest their involvement in alternate pathways; an implication that could prove particularly useful in their utilization as part of a panel of multiplexed markers with calculated cut-offs for diagnoses based on achievable sensitivity and specificity.

LC initiation, a key consequence of CLD is characterised by reduced hepatocyte proliferation indicating a defection in the regenerative capacity of the liver possibly due to telomere shortening. This process has been associated with LC in studies that found telomeres in the hepatocytes of cirrhotic livers to be significantly shorter than those of non-cirrhotic liver sections (Delhaye et al., 1996). In mouse models, telomere dysfunction has been shown to increase the growth of hepatic neoplasms although this was also coupled to a decline in malignant transformations, suggesting that these mechanisms individually are not the sole causes of tumour development. Such alterations may provide an enhanced environment which in the presence of excessive secretion of cytokines, growth factors and products that trigger oxidative stress (Bataller and Brenner, 2005; El-Serag and Rudolph, 2007) work synergistically with known aetiological agents to allow for un-arrested tumour progression.

Results obtained from the 1D immunoblotting of glycoprotein enriched plasma from the Control, LC & HCC groups of the GLCS suggest the presence of disease stage related effects on glycosylated forms of CC3, Apo A1 and HPX and potential

virus related effects highlighted by differences in comparative abundance of glyco-enriched ApoA1 and α1AT. To begin to resolve the sources of the modifications at the root of these changes would require extensive analysis beyond the scope of this project. However, initial experiments were conducted by 2D western blots on plasma pooled according to the major aetiological factors for LC and HCC compared to controls negative for HBV, HCV & p53 mutations. Whereas previous chapters in this report have focussed on quantitative measurements of proteins suggested by MS as differentially expressed; HPX was taken forth out of the four proteins based on its relatively simple structure and the identification of a consistent and clear difference in glycosylation between the controls and all other sub-groups (figure 4.6 d).

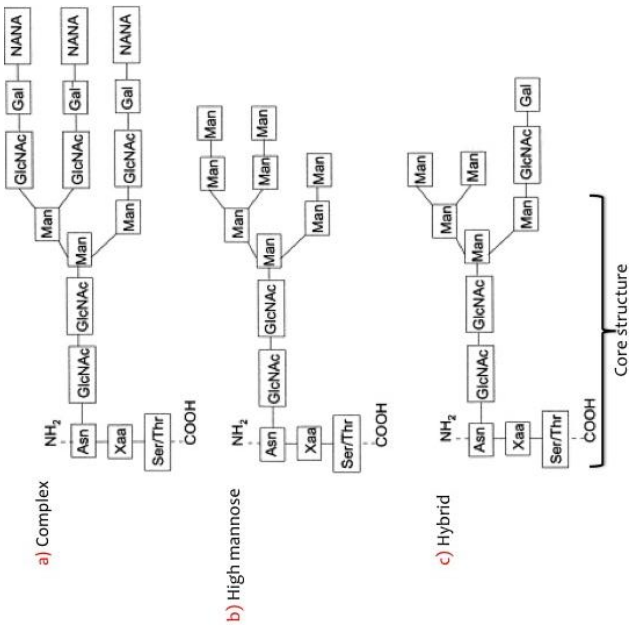


Figure 4.12: Summary of the major classes of N-glycans, adapted from (Charlwood et al., 2001). The core structures across the three are uniform with the differences resulting from the linkages to the branched $(Man)_3$ linked core.

Of the six known amino acid modification sites for HPX (Frantikova et al., 1984; Takahashi et al., 1984) five are N-linked at amino acid positions 64, 187, 240, 246 and 453 making them the optimal subclass to enrich for (figure 4.3) when searching for disease specific glycosylation modifications. Blood plasma serves as an ideal medium for such study due to its ready accessibility and large pool of dynamically expressed modified and unmodified proteins. In the context of carcinogenesis, there exist well-known associations between branched structures such as high-mannose glycoproteins (figure 4.12) and biological functions such as cell adhesion and tumour invasiveness and metastasis as well as physiological alterations to cell growth, migration and differentiation. The addition of sugar moieties to proteins as mediated by glycosyltransferases such as N-acetylglucosaminyltransferase III (GnT-III), GnT V and α-1,6-fucosyltransferase at the primary sites of the endoplasmic reticulum (ER), golgi apparatus, cytosol and nucleus (figure 4.13) could thus hold the key to distinct changes which occur in the transformation from a healthy liver to cirrhosis or carcinogenesis. This rationale was applied to the glycoprotein enriched plasma experiments reported in this chapter with more extensive separation methods utilised for HPX.

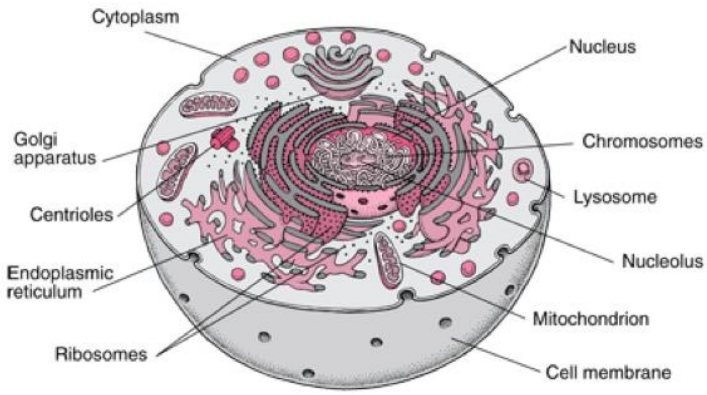


Figure 4.13: Generic image of human cell, source: http://www.agedefyingbody.com/Border.html

The consistent and distinct pattern of HPX migration on a linear pH 3 to 10 gradient in liver disease free controls as compared to the two disease groups suggests intrinsic changes in its expression or processing which are altered in the course of disease onset and or progression. Another distinct change observed with linear two-dimensional separation is the general absence of the lower molecular weight HPX fragment from all HCC pools. In the LC pools the highlighted fragment (figure 4.7 b & c) resolves to the mid pH range at approximately 52kDa in the HCV & HBV virus positive pools, but is absent from the virus-free subjects. This trend is not replicated in the JUTH LC samples possibly because of the small sample size of the group and its associated data not allowing for the control and mutual exclusion of the primary aetiological factors; HBV, HCV and p53 mutation. The absence of this lower molecular weight specie in the JUTH & GLCS HCC pools however remain consistent making it a worthy candidate of further study as a potential discriminatory marker for the detection of early hepatocarcinogenesis.

The 2D probing of HPX was repeated on non-linear IPG strips within the same pH range in order to see if it helped better resolve the concentration of spots in the low to mid pH range. Due to sample limitations, only six of the eight pools previously compared were ran. The patterns seen were quite complex but noteworthy expression profiles include the consistent absence of lower molecular weight HPX from the HCC pools as well as virus free LC GLCS subjects. This is consistent with the resolved trends observed in the linear IPG's described above. In figure 4.10, better-resolved spots at the 52kDa region once again support the

possibility of this lower molecular weight specie being specifically discriminatory for HCC. To attempt to characterise possible molecular changes responsible for the distinct spot patterns toward the pH 3 ends in the control groups, sialidase and phosphatase enzymatic treatments were piloted on liver disease free plasma pools from the GLCS and JUTH subjects. Sialidase treatment in both populations resulted in distinct spot formations toward the acidic end of the immobilized pH gradient strips. This distinctive effect from sialic acid cleaving enzymes suggests the presence of sialic acid positive HPX linked sugars which may be more pronounced in liver disease free individuals hence resulting in a more pronounced extension of spots toward the acidic end. Sialidase free experiments were run in conjunction with these treatments, however due to suspected changes in solubility as a result of added chemicals or sample loss during preparation, the developed films contained no clear protein spots (figure AF 1.5 b-c). The phosphatase treatment (figure 1.6 a-b) showed no distinct changes between enzyme positive and negative samples – except for the disappearance of some HPX species at the pH 3 end of the gel gradient. This notable but minimal change ruled out phosphorylation as a major contributor to the unique distribution of HPX within the control groups.

The data presented and discussed above represents attempts to independently validate and further characterize Apo A1, CC3, HPX and A1AT using independent subject groups and methods. The processes involved in biomarker discovery are multi-layered and multi-faceted requiring investigators to commit to well-established methods of analysis to chance at identifying exploitable signatures

applicable to specific disease contexts. Here we see the confirmed validation in an independent albeit smaller population of the major trends observed from the GLCS discovery case-control subject group. Using one and two-dimensional immunoblot based assays, there also begins the suggestion of unique HPX profiles linked to control and HCC subjects which with further characterization and development may present far more specific markers than what may be achievable by simply observing the changes in abundance of whole proteins.

4.4 Conclusion

The attempt made in these investigations to advance the differentially expressed proteins from the previous chapter and validate their expression profiles in a regional subject group with expected similar aetiology of liver disease is an important step in establishing the differential expression and robustness of these markers. The biochemical characterization analyses presented highlights the complexity of each proposed candidate and how their glycosylation profiles change with disease progression. Overall, this work begins to validate proposed biomarkers utilizing two distinct approaches; direct measurement of markers carries the advantage of allowing for multiplexing whereas focusing on changes in key molecular structures though difficult could offer unmatched specificity. Further investigations would be warranted to potentially exploit observed changes for development into useable kit based assays.

Chapter 5: Statistical Multiplexing of Putative Markers and Current Clinical Tests for Liver Disease

5.0 Introduction

One of the major approaches to the identification of high performance biomarkers, apart from utilizing specialised, high-definition '-omics discovery methods for probing human tissue and bio-fluids is by taking candidate markers and using statistics to evaluate improvements in diagnostic ability from various test combinations. These methods have the potential to reduce noise and complementarily improve upon weaknesses of single marker tests on sensitivity or specificity. With numerous data mining approaches already being applied to this objective (Li et al., 2004), no single technique has come to the forefront as optimal in combining selections of putative biomarkers. What they all aim to achieve, however, is the identification of unique combinations of tests which optimally discriminate progressive disease states and vastly improve upon diagnostic ability. There are many factors that limit current abilities to carry out such analyses, of which are the vast differences in number between proposed biomarker candidates and the availability of well characterised sample cohorts. To avoid the problem of multiple measurements of proposed markers on precious and limited human samples; procedures must be established to curate and select for which markers and comparisons are to be prioritized and based on the performances of these, take them forth for further validation.

This chapter focuses on evaluating the measurements of α1AT, Apo A1, CC3 and HPX compared with a routinely used marker of hepatocyte damage; serum ALT, to examine whether any diagnostic advantage is conferred by using either of these proteins singly or in tandem with this test. Idealistically, AFP would have been the best marker to assess these four proteins against but that would require the recruitment of a novel subject group whose HCC and LC diagnoses were made independent of serum AFP expression. This issue will be further expanded on in the ensuing discussion.

The ALT protein is primarily expressed in the liver, with lower levels seen in the kidneys, heart, pancreas and muscle. It is normally present in small amounts in the blood, but chronic liver injury and inflammation causing hepatocyte damage releases this intracellular enzyme in large amounts causing its circulatory levels to increase. It is thus a sensitive and quite specific clinical indicator, able to highlight the presence of continuous liver damage regardless of aetiology. ALT has also been suggested as a useful marker in the assessment of HCC recurrence in patients who have had HCV associated tumour resections (Tarao et al., 2000). Amongst long-term HBV carriers, high ALT expression is reported to be associated with elevated risk of morbidity and poor development of liver health (Kim et al., 2008; Yuen et al., 2005). The results displayed below from multiple comparisons of single and combined markers demonstrate the variable effects of statistical combinations at designated cut-offs.

5.1 Methods

5.1.1 Subject Selection

The full GLCS subject population was collated with subjects who had a full panel of individual measurements for α1AT, Apo A1, CC3, HPX along with ALT selected for inclusion into direct statistical multiplexing of these markers. The same collation was performed for the JUTH subjects but due to the small subject numbers, no statistical combinations were attempted or viable. Following data assembly, the combined ROC curve generation and putative marker multiplexing was conducted by a collaborative statistician based in Lyon, France.

5.1.2 AUC generation for selecting multiplexing candidates

Data analyses, ROC curve comparisons and statistics were conducted by Dr Suzi Camey using the SAS statistical program (SAS 9.2, SAS Institute Inc., North Carolina, USA). For each putative biomarker, ROC curves were plotted and AUC values calculated along with their associated 95% CI and SEM using the PROC LOGISTIC function and %PLOTROC macro sourced from SAS Institute website (support.sas.com/kb/25/018). Comparisons of AUC values were conducted using the %ROC macro downloaded from support.sas.com/kb/25/017.

5.1.3 Biomarker combinations at specified cut-offs

For each of the seven protein combinations that demonstrated greater discriminatory power than ALT, optimal cut-offs were calculated based on the maximum likelihood ratio with highest achievable balance in sensitivity and

specificity. These indices were calculated for ALT, α1AT, Apo A1, CC3 and HPX for the Con, LC and HCC categories using the Graph Pad Prism statistical package. To further elucidate these results, the cut-offs were applied to the three subject categories, to see what levels of sensitivity and specificity were attainable. Binary combinations based on the "AND" and "OR" methodologies were analysed and resultant sensitivity and specificity indices tabulated (Tables 5.8 a-c).

Due to the small sample numbers, no stratifications were made for ALT levels.

5.2 Results

5.2.1 Analysis of test combinations suggest HPX, CC3 and Apo A1 as superior to ALT in diagnosing LC.

Comparisons of various combinations of the four validated candidate biomarkers and ALT highlight that HPX, CC3 and Apo A1 are individually all superior to ALT in discriminating control and LC cases. HPX in this subject population has been shown to have an AUC of approximately 0.8000 (tables 3.3 & 5.1), a value that is significantly higher than that for ALT when used singly to distinguish the same two groups; 0.6507. A negligible boost in diagnostic accuracy from single use of HPX is seen when these two markers are amalgamated (figure 5.1) although the difference between using HPX and HPX & ALT combined does not reach significance. However, combining ALT and a complex of HPX & ALT offers similar discriminatory power (AUC 0.7949) and carries a significant improvement compared to using ALT alone.

The other two proteins suggested to perform better than ALT in detecting LC are CC3 and Apo A1, with respective AUC's of 0.7039 and 0.7638 (tables 5.2 & 5.3). In both cases, some discriminatory benefit is indicated by complexing the proteins with ALT alone although the degree of changes seen are not significant. When both proteins are complexed with ALT, the AUC's obtained improve significantly from that seen with single ALT measurements.

Marker	AUC (Con vs. LC)	SEM	95% CI
ALT IU/ml	0.6507	0.0532	0.5464 - 0.7550
HPX mg/ml	0.7949	0.0429	0.7108 - 0.8789
HPX_ALT	0.8014	0.0428	0.7176 - 0.8852
Comparison	**p-value of difference**		
ALT & HPX	0.0242		
ALT & HPX_ALT	0.0095		
HPX & HPX_ALT	0.5871		

Table 5.1: Summary of statistical indices for HPX versus ALT and ALT_HPX in Con vs. LC GLCS subjects.

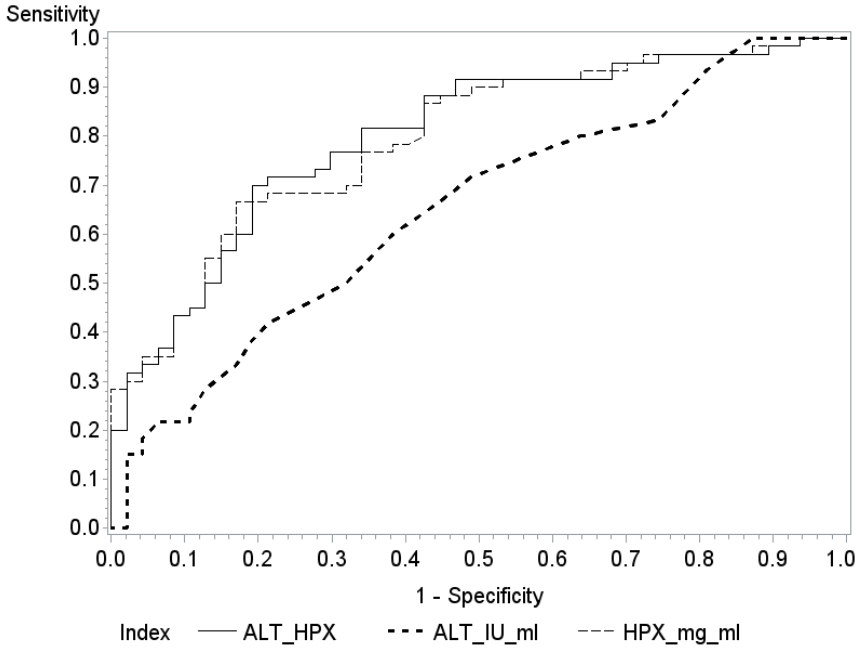


Figure 5.1: Combined ROC curves for HPX and ALT in Con vs. LC subjects highlighting which marker combinations are superior in diagnostic ability.

Marker	AUC (Con vs. LC)	SEM	95% CI
ALT IU/ml	0.6507	0.0532	0.5464 - 0.7550
CC3 mg/ml	0.7039	0.0504	0.6052 - 0.8026
ALT_CC3	0.7535	0.0480	0.6594 - 0.8477
Comparison	**Pr > ChiSq.**		
ALT & CC3	0.4446		
ALT & ALT_CC3	0.0434		
CC3 & ALT_CC3	0.0687		

Table 5.2: Summary of statistical indices for CC3 versus ALT and ALT_CC3 in Con vs. LC GLCS subjects.

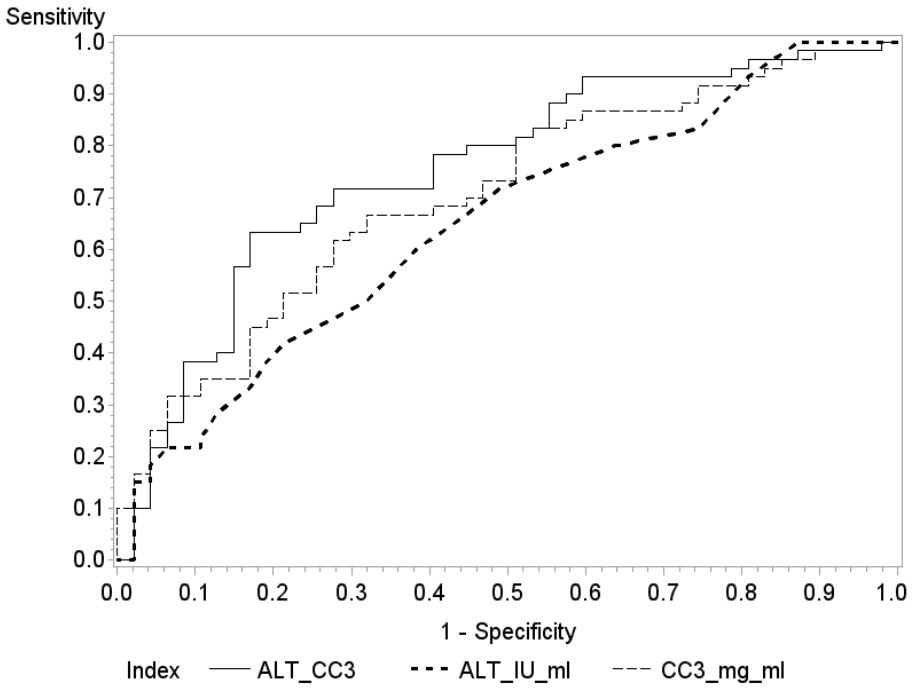


Figure 5.2: Combined ROC curves for CC3 and ALT in Con vs. LC subjects, highlighting which test combinations are superior in diagnostic ability.

Marker	AUC (Con vs. LC)	SEM	95% CI
ALT IU/ml	0.6507	0.0532	0.5464 - 0.7550
ApoA1 mg/ml	0.7638	0.0480	0.6687 - 0.8569
ALT_ApoA1	0.7922	0.0464	0.7012 - 0.8832
Comparison	**p-value of difference**		
ALT & ApoA1	0.1017		
ALT & ALT_ApoA1	0.0121		
ApoA1 & ALT_ApoA1	0.1541		

Table 5.3: Summary of statistical indices for ApoA1 versus ALT and ALT_ ApoA1 in Con vs. LC GLCS subjects.

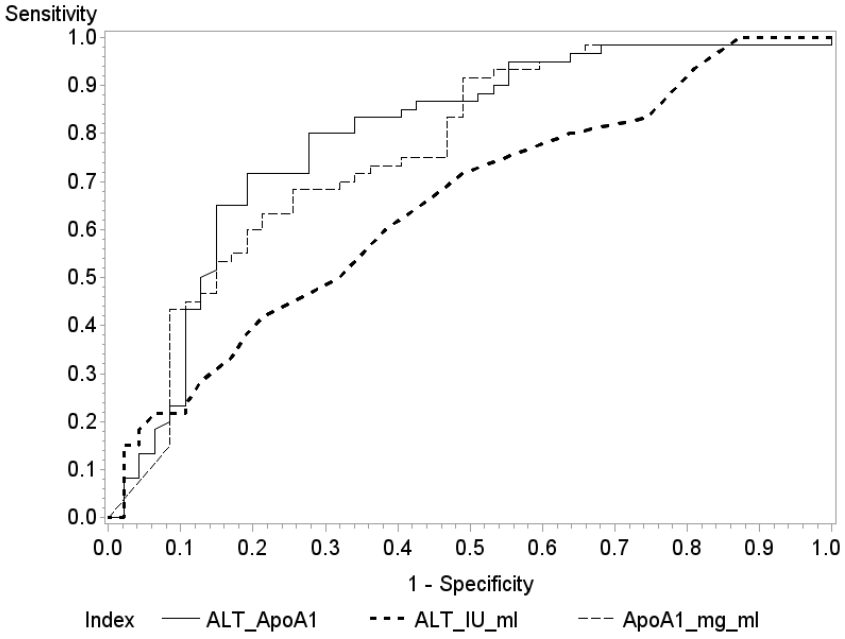


Figure 5.3: Combined ROC curves for ApoA1 and ALT, in Con vs. LC subjects, highlighting which test combinations are superior in diagnostic ability.

5.2.2 Analysis of test combinations suggests ALT combined with α1AT, Apo A1 or HPX boost discriminatory potential in HCC patients.

ALT in the GLCS population shows high performance (AUC 0.8372) in discriminating the control and HCC groups. Despite this, some benefit is still observed when it is complexed with α1AT or Apo A1 with upwards shifts in AUC to 0.8723 and 0.8899 respectively. The latter combination of ALT & Apo A1 results in noteworthy statistics suggesting that the increase in diagnostic ability conferred by using the two markers in a panel is significant; p = 0.0183 (table 5.5).

As has been the trend in all the analyses of this study, the two groups most difficult to differentiate are those classed as LC and HCC. From the multiple comparisons performed only two of the markers demonstrated an improvement

from the discriminatory values seen with ALT alone. α1AT shows enhanced discriminatory potential above that of ALT with an AUC of 0.7171; marginally higher than that calculated for the routinely indicator of liver cell damage or inflammation (table 5.6). HPX on its own has poorer performance than ALT in separating out these two groups (table 5.7) but when complexed together, the two proteins start to show a significantly superior ability to differentiate between subjects diagnosed with LC and HCC.

Marker	**AUC (Con vs. HCC)**	**SEM**	**95% CI**
ALT IU/ml	0.8372	0.0437	0.7516 - 0.9228
α1AT mg/ml	0.8197	0.0466	0.7284 - 0.9109
ALT_ α1AT	0.8723	0.0412	0.7915 - 0.9513
Comparison	**p-value of difference**		
ALT & α1AT	0.7777		
ALT & ALT_α1AT	0.4288		
α1AT & ALT_α1AT	0.0765		

Table 5.4: Summary of statistical indices for α1AT versus ALT and ALT_α1AT in Con vs. HCC GLCS subjects.

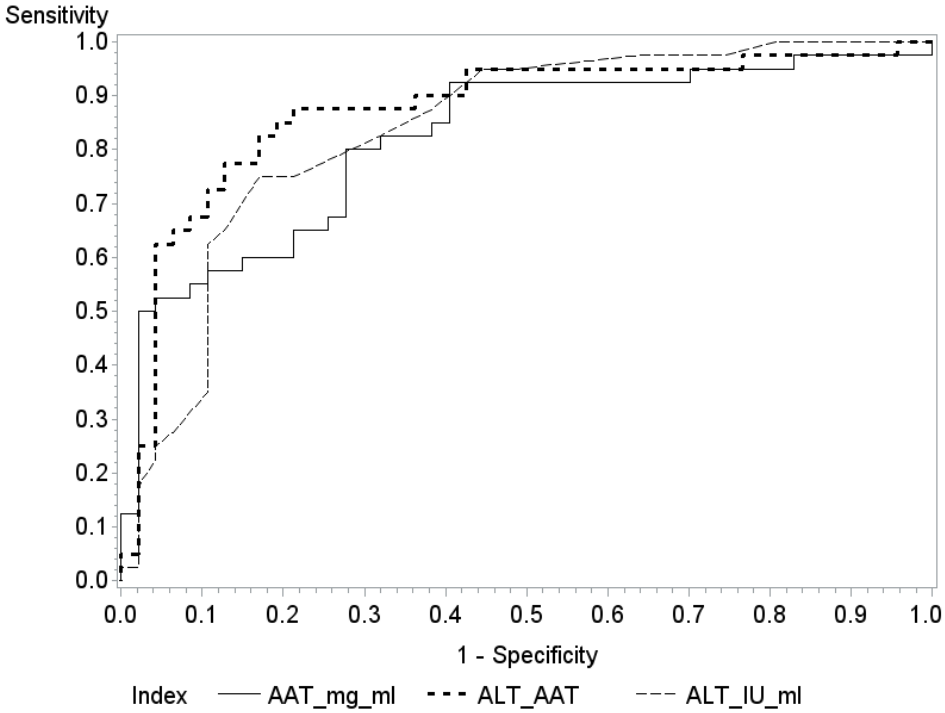


Figure 5.4: Combined ROC curves for α1AT and ALT in Con vs. HCC subjects, highlighting which test combinations are superior in diagnostic ability.

Marker	AUC (Con vs. HCC)	SEM	95% CI
ALT IU/ml	0.8372	0.0437	0.7516 - 0.9228
ApoA1 mg/ml	0.8255	0.0451	0.7372 - 0.9139
ALT_ApoA1	0.8899	0.0357	0.8199 - 0.9599
Comparison	**p-value of difference**		
ALT & ApoA1	0.8309		
ALT & ALT_ApoA1	0.1379		
ApoA1 & ALT_ApoA1	0.0183		

Table 5.5: Summary of statistical indices for ApoA1 versus ALT and ALT_ApoA1 in Con vs. HCC GLCS subjects.

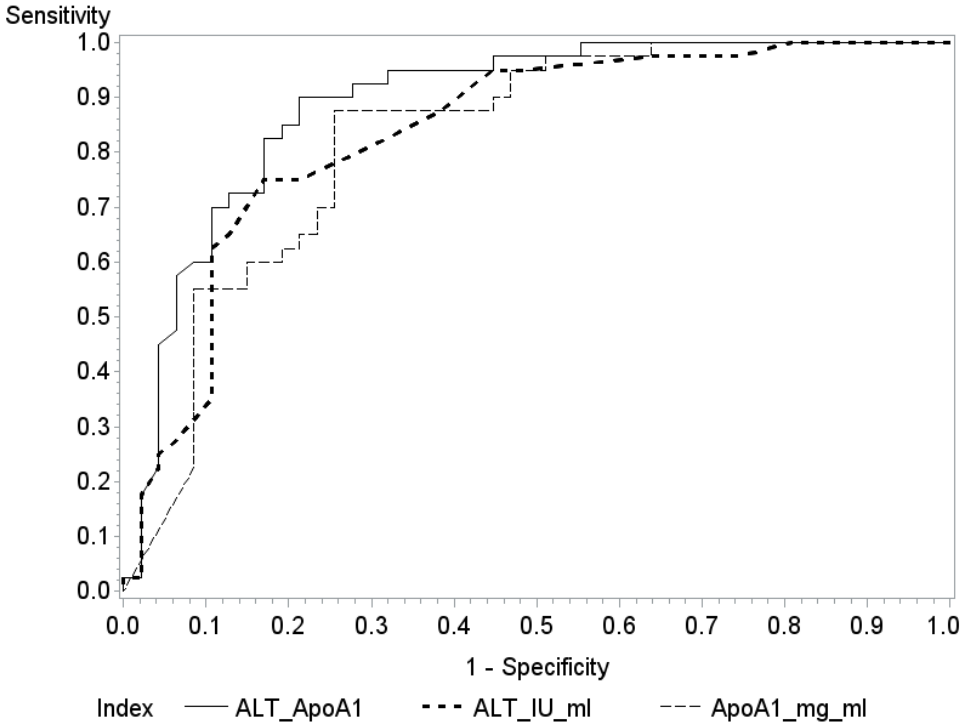


Figure 5.5: Combined ROC curves for ApoA1 and ALT in Con vs. HCC subjects, highlighting which test combinations are superior in diagnostic ability.

Marker	AUC (LC vs. HCC)	SEM	95% CI
ALT IU/ml	0.7079	0.0524	0.6053 - 0.8105
α1AT mg/ml	0.7171	0.0525	0.6141 - 0.8201
α1AT_ALT	0.7663	0.0488	0.6707 - 0.8618
Comparison	**p-value of difference**		
ALT & α1AT	0.8918		
ALT & ALT_α1AT	0.2066		
α1AT & ALT_ α1AT	0.0899		

Table 5.6: Summary of statistical indices for α1AT versus ALT and ALT_α1AT in LC vs. HCC GLCS subjects.

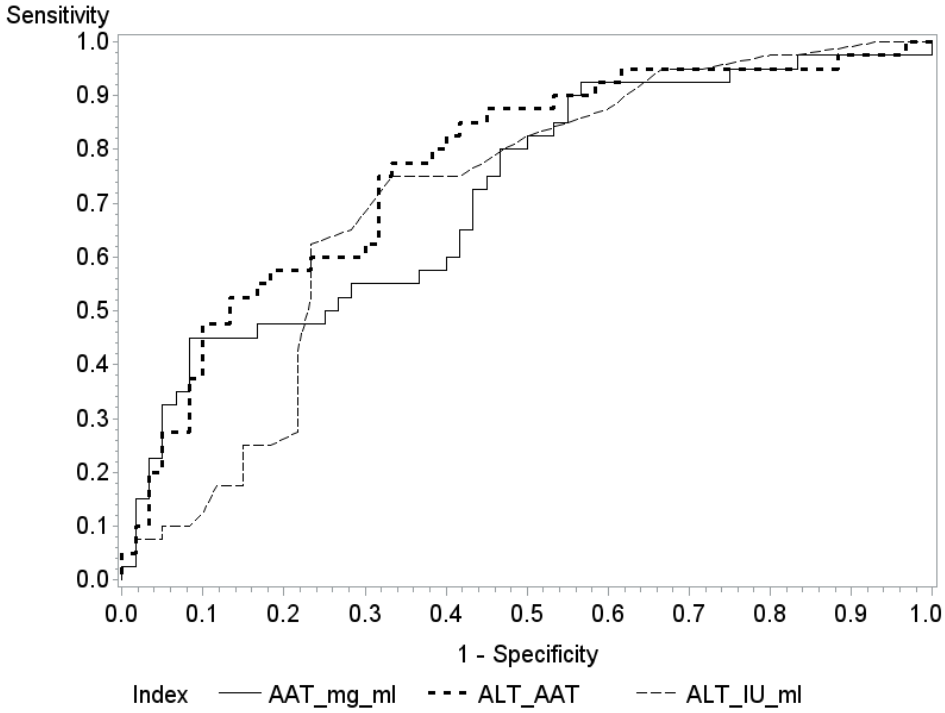


Figure 5.6: Combined ROC curves for CC3 and ALT in LC vs. HCC subjects, highlighting which test combinations are superior in diagnostic ability.

Marker	AUC (LC vs. HCC)	SEM	95% CI
ALT IU/ml	0.7079	0.0524	0.6053 - 0.8105
HPX mg/ml	0.6250	0.0567	0.5138 - 0.7362
HPX_ALT	0.7650	0.0475	0.6720 - 0.8580
Comparison	**p-value of difference**		
ALT & HPX	0.3583		
ALT & ALT_HPX	0.2614		
HPX & ALT_ HPX	0.0100		

Table 5.7: Summary of statistical indices for HPX versus ALT and ALT_ HPX in LC vs. HCC GLCS subjects.

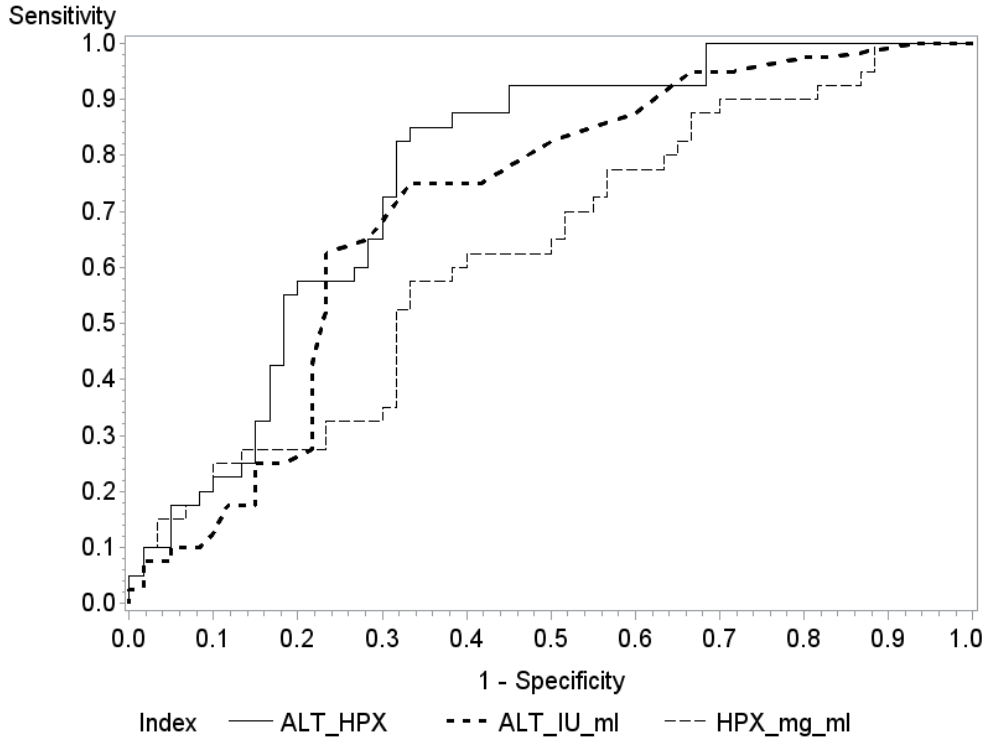


Figure 5.7: Combined ROC curves for HPX and ALT in LC vs. HCC subjects, highlighting which test combinations are superior in diagnostic ability.

5.2.3 Application of specified cut-offs to binary protein combinations significantly boosts diagnostic ability

The application of defined cut-offs to protein combinations with superior AUCs to single ALT use primarily highlighted that the OR criteria for test design may be superior in offering balanced sensitivity and specificity values for diagnosis. These results also suggest that higher individual sensitivity or specificity values are attainable with the "AND" criterion.

ALT combined with CC3 in discriminating CON and LC subjects; HPX for LC and HCC and ApoA1 in CON and HCC discrimination showed upward shifts in AUC with significant p values, though the significance level for the first comparison was

borderline (p<0.0687 table 5.2). When specified cut-offs were used, accurate subject assignments between the CON and HCC groups with the following criteria – "ALT of >10.50 IU/ml OR ApoA1 >0.7673 mg/ml" had respective sensitivity and specificity values of 82.5% and 74.47. HPX & ALT combined with the same OR criterion at cut-offs of HPX >2.701 mg/ml and ALT >13.50 IU/ml in the LC and HCC categories had high sensitivity, but poorer specificity; 97.50% and 51.67% respectively.

In the use of ALT and CC3 together for discerning the CON and LC groups, the OR command again showed a greater balance in achievable sensitivity and specificity; 85% & 57.45%, respectively than using the cut-offs of ALT >9.5 IU/ml and CC3 <0.9579 mg/ml with the AND command. The latter produced results which skewed the balance towards poor sensitivity (18.33%) but high specificity (93.62).

Con (n=47) vs. HCC (n=40)	Sens %	95% CI	p	Spec %	95% CI	p
ALT >10.50 IU/ml	75.00	0.5880 - 0.8731	0.0080	80.85	0.6674 - 0.9085	<0.0001
AAT >3.873 mg/ml	60.00	0.4333 - 0.7514	0.1341	82.98	0.6919 - 0.9235	<0.0001
ALT_AAT (>10.50 IU/ml AND >3.873 mg/ml)	47.50	0.3151 - 0.6387	0.4373	95.74	0.8546 - 0.9948	<0.0001
ALT_AAT (>10.50 IU/ml OR >3.873 mg/ml)	87.50	0.7320 - 0.9581	<0.0001	68.09	0.5288 - 0.8091	0.0100
ApoA1 <0.7673 mg/ml	60.00	0.4333 - 0.7514	0.1341	85.11	0.7169 - 0.9380	<0.0001
ALT_ApoA1 (>10.50 IU/ml AND <0.7673 mg/ml)	52.50	0.3613 - 0.6849	0.3759	91.49	0.8351 - 0.9947	<0.0001
ALT_ApoA1 (>10.50 IU/ml OR <0.7673 mg/ml)	82.50	0.6722 - 0.9428	<0.0001	74.47	0.5965 - 0.8606	<0.0001

Table 5.8a: Statistical summary from the combination of proteins using select cut-offs as a measure of diagnostic potential demonstrated by achievable sensitivity and specificity, for the Con and HCC groups.

Con (n=47) vs. LC (n=60)	Sens %	95% CI	p	Spec %	95% CI	p
ALT >9.50 IU/ml	41.67	0.2907 - 0.5512	0.1225	78.72	0.6434 - 0.8930	<0.0001
ApoA1 <0.8079 mg/ml	53.33	0.4000 - 0.6633	0.3028	82.98	0.6919 - 0.9235	<0.0001
ALT_ApoA1 (>9.50 IU/ml AND <0.8079 mg/ml)	20.00	0.1078 - 0.3233	<0.0001	91.49	0.7962 - 0.9763	<0.0001
ALT_ApoA1 (>9.50 IU/ml OR <0.8079 mg/ml)	75.00	0.6214 - 0.8528	<0.0001	70.21	0.5511 - 0.8266	0.0040
CC3 <0.9579 mg/ml	61.67	0.4821 - 0.7393	0.0462	72.34	0.5736 - 0.8438	0.0011
ALT_CC3 (>9.50 IU/ml AND <0.9579 mg/ml)	18.33	0.0952 - 0.3044	<0.0001	93.62	0.8246 - 0.9866	<0.0001
ALT_CC3 (>9.50 IU/ml OR <0.9579 mg/ml)	85.00	0.7343 - 0.9290	<0.0001	57.45	0.4218 - 0.7174	0.1536
HPX <2.5570 mg/ml	60.00	0.4654 - 0.7244	0.0607	85.11	0.7169 - 0.9380	<0.0001
ALT_HPX (>9.50 IU/ml AND <2.5570 mg/ml)	28.33	0.1745 - 0.4144	0.0004	97.87	0.8871 - 0.9995	<0.0001
ALT_HPX (>9.50 IU/ml OR <2.5570 mg/ml)	73.33	0.6034 - 0.8393	0.0002	65.96	0.5069 - 0.7914	0.0143

Table 5.8b: Statistical summary from the combination of proteins using select cut-offs as a measure of diagnostic potential demonstrated by achievable sensitivity and specificity, for the Con and LC groups.

$LC^{(n=60)}$ vs. $HCC^{(n=40)}$	Sens %	95% CI	p	Spec %	95% CI	p
ALT >13.50 IU/ml	62.50	0.4580 - 0.7727	0.0569	76.67	0.6395 - 0.8662	<0.0001
HPX >2.701 mg/ml	57.50	0.4089 - 0.7296	0.2148	66.67	0.5331 - 0.7831	0.0049
ALT_HPX (>13.50 IU/ml AND >2.701 mg/ml)	22.50	0.1084 - 0.3845	0.0003	91.67	0.8161 - 0.9724	<0.0001
ALT_HPX (>13.50 IU/ml OR >2.701 mg/ml)	97.50	0.8684 - 0.9994	<0.0001	51.67	0.3839 - 0.6477	0.4487
AAT >4.587 mg/ml	50.00	0.3380 - 0.6620	0.5000	75.00	0.6214 - 0.8528	<0.0001
ALT_AAT (>13.50 IU/ml AND >4.587 mg/ml)	32.50	0.1799 - 0.4701	0.0134	90.00	0.7949 - 0.9625	<0.0001
ALT_AAT (>13.50 IU/ml OR >4.587 mg/ml)	80.00	0.6435 - 0.9095	<0.0001	61.67	0.4821 - 0.7393	0.0354

Table 5.8c: Statistical summary from the combination of proteins using select cut-offs as a measure of diagnostic potential demonstrated by achievable sensitivity and specificity for the LC and HCC groups.

5.3 Discussion

In any attempt to identify a viable, high performance biomarker within a subject population; shortlisted candidates will at some point have to be compared against contemporary tests in current clinical use. A major challenge faced in doing this is when these benchmark assays have been used intrinsically to classify the various subject groups under study. AFP formed part of the diagnostic profile of both the GLCS and JUTH subjects – thus when used in these population to approximate its diagnostic ability, the observed AUC is grossly exaggerated (AT 1.1 & AF 1.8). Consequently, when it came to the utilization of statistical methods to combine α1AT, Apo A1, CC3 and HPX with more established markers – an independent indicator of liver health not specific to HCC or LC diagnoses had to be used. This approach, although not ideal is still useful in offering insight as to how these four proteins perform in comparison to current clinical indicators of liver disease. Not much has been reported on the independent diagnostic ability of AFP, but as stated earlier in this thesis – it is approximated at 0.70, with at least one publication showing AUC values as low as <0.60 when used at a cut off of 100ng/ml (Giannini et al., 2012).

The results detailed above highlight the protein combinations which perform better than the sole use of ALT. The shifts seen with the application of calculated cut-offs depict the fine balance that has to be struck when combining multiple proteins into a single diagnostic test. No single approach to this has been universally proposed; rather the logic for how multiple tests are combined

continues to be uniquely tailored to the complimentary values of the markers under use. To begin to move these combination methods toward successful application in real life assays some synergy will need to be established between medical researchers who identify and propose candidates and kit developers with the expertise to design viable pilot assays useable in the extensive validation of novel test panels.

As '-omics based biomarker studies start to mature and move into advanced validation stages, new methods of subject recruitment must also be developed which allow for direct, unbiased comparisons between established clinical bench-marks and proposed biomarkers. The statistics and results depicted above highlight the value which access to such cohorts would have had in this project. In the absence of these specialized subject groups however, other validatory approaches such as comparisons against secondary clinical indices as done here as well as direct measurements of putative biomarkers in independent well characterized subject populations must be utilised. These exercises all form vital steps in the advancement of panels of proteins useable in the diagnosis and monitoring of HCC and its associated conditions.

5.4 Conclusions

The direct measurement of four of the twenty-six shortlisted proteins from the discovery phase of this study allowed for the valuable application of combination statistics to the GLCS population. The superiority over ALT of all of these proteins

in discriminating at least one of our three comparative groups suggests that they should be followed up for development as novel indicators of CLD in the clinics of HBV related HCC and LC in regions where these conditions are prevalent.

6.0 Discussion

This thesis presents an extensive study of protein expression in liver diseases as a potential tool of differential diagnosis. The most consistent and significant of our findings have been presented in the preceding chapters. Much insight can however be garnered from additional unreported outcomes observed during the course of these investigations. These additional results, limitations faced as well as other salient points will be discussed in this chapter.

Of the key proteins from table 3.2 that were considered for individual subject based validation; clusterin and haptoglobin were selected in addition to HPX, CC3, Apo A1 & α1AT because of their significant reported regulation factors, commercial kit availability as well as links to LC and HCC in literature. The initial strategy adapted when these two proteins were assayed was to have more targeted measurements of proteins in the disease vs. control group where they showed the most significant biomarker potential. For haptoglobin, label free discovery proteomics suggested it as a candidate for LC differentiation whereas clusterin was presented as a marker for both LC and HCC. To validate these suggestions, ELISA measurements were made comparing expression levels in Con vs. HCC and LC subjects for clusterin and haptoglobin, respectively. The results obtained displayed trends of expression consistent with those reported in table 3.2; however, the differences observed at the whole protein level were not of such significance to warrant continued analyses (figures AF 1.1 & 1.2), with poor reported AUC scores that did not reach statistical significance.

At the biochemical level, haptoglobin demonstrated evidence of disease related alterations within the lower molecular weight 142 amino acid length alpha chain region (figure AF 1.3). The beta chain of this protein has an expected molecular weight of approximately 31 kDa; however no bands corresponding to this size were detected in our assayed samples. Instead, a band slightly above the native alpha chain was identified and suspected to be a potentially modified form of alpha chain haptoglobin. The uniprot database lists four known glycosylation sites for haptoglobin. To confirm whether or not this modification is responsible for the observed mass shift, glycoprotein enrichment and digestion procedures such as those conducted for hemopexin would have to be applied. Resolving the identity and source of these alterations, could provide a worthy inroad, able to better describe the condition specific changes already observable in this protein.

With glycosylation suggested as one of the major PTM changes that occur in liver diseases and as such given particular focus in this thesis; significant efforts were exerted in the latter stages of this project to identify which sub-species within expressed glycoproteins may have resulted in the unique 2DE separation pattern observed between the control groups from both the JUTH and GLCS, and all other disease pools (figures 4.7 & 4.8). The first step taken in this endeavour was to confirm that the sub-population of proteins enriched by co-incubation with the Con A lectin were significantly glycosylated, resulting in the upward shift in molecular weight seen in all of the HPX immunoblots. This was confirmed as demonstrated in the appendix figure AF 1.4, which shows the effect of treatment with the PNGase F enzyme - a glycosidase which cleaves glycoproteins at the

innermost asparagine residue (figure 4.12) of high mannose, complex and hybrid structures, thus completely removing any added sugar groups.

Once this was certified, the next step taken to attempt to characterise the key molecular species being expressed within the enriched glycoprotein populations were treatments with sialic acid cleaving enzymes. Sialylation forms an important modification in oligosaccharide-protein complexes; sialic acids are the collective term used to represent the N- or O- substituted products of neuraminic acids. They are negatively charged sugar moieties widely distributed in animal tissues as terminal compounds attached to protein or lipid molecules. Their over-expression in cancer cells has been steadily documented, (Fuster and Esko, 2005; Hedlund et al., 2008; Narayanan, 1994) with particular focus placed on their role in the promotion of metastasis. The unique negative charges carried by glycoproteins expressing these moieties allows them to participate in a wide array of cellular functions (Schauer, 2000) such as glycoprotein conformation on cell membranes, transport of positively charged compounds and cell-cell repulsion which enhances the ability of poorly differentiated tumour cells to enter blood vessels. As a commonly expressed population within N-linked glycoproteins, these structures were targeted for enzyme assisted probing of their potential expression in glycoprotein enriched control samples blotted for HPX. Despite evidence of the existence of sialic acid species within this protein, a great drawback faced in these experiments was the apparent loss of sample seen in the enzyme un-treated control pools. Further investigations will need to be carried out to certify at which stage this occurs (figure AF 1.5 a-c). Results seen in the 1D immunoblots however

indicate that the sample loss occurred during the first phase of preparation, prior to second dimensional IEF. Possible avenues of sample loss include inadequate pellet resuspension, changes in protein solubility causing the material of interest to be precipitated and discarded or un-intentional sample removal.

Additional experimental problems were faced with the ELISA assays used for Apo A1 measurement, which demonstrated marked inconsistencies between kit lots. A distinct decline in sensitivity was detected following the order of a second batch of kits for determination of the expression levels of the protein in the JUTH population. A similar drop in kit sensitivity was not observed for any of the other three markers, thus ruling out instrumentation or sample based problems as the major cause. Investigations made in conjunction with the kit manufacturers in the USA to identify the source of the issue yielded no conclusive results. However, as no substantial changes were effected by altering the concentration of HRP labelled conjugate antibody, it was strongly suspected that the error likely occurred at the point of Apo A1 antibody coating of the 96-well polystyrene microtitre plates. Comparisons of results from the GLCS subjects all assayed with the original kit lot performing as expected at a suggested sample dilution of 1:20000 and the lowered sensitivity lot used for the JUTH samples at 1:500, demonstrated uniform overall trends of expression but with a 100 fold decline in measured Apo A1 levels (figures 3.4 a & 4.4). This consistency supports the notion that a systematic error was made at the manufacturing stage resulting in the lowered sensitivity output.

Aside from these major issues faced during laboratory based experiments, some key factors not accounted for or scrutinized in the two sample populations used in this study are the possible effects of HIV and occult HBV infection on the performance of proposed biomarker candidates. There is growing evidence suggesting that the impact of occult HBV infection could be a significant but underestimated contributor to HCC in HBV endemic areas. Conventional assays used to detect HBV viral loads have been insufficient in identifying this profile of patients, save for a new generation of more sensitive molecular assays. An expert meeting on occult HBV infection held in Italy in 2008 defined occult infection as:

"(The) Presence of HBV DNA in the liver (with detectable or undetectable HBV DNA in the serum) of individuals testing HBsAg negative by currently available assays." Source: (Raimondo et al., 2008)

Detectable levels of serum DNA in these cases are often very low (<200 IU/ml). Studies looking at the prevalence of occult HBV in patients co-infected with HIV found markedly higher infection rates in HIV positive individuals with 5 of 15 HIV positive subjects testing positive in one retrospective analysis (Mphahlele et al., 2006); and 4.8% of subjects in a large urban South African cohort also found to carry the occult infection (Firnhaber et al., 2009). In the Gambia, no studies have as yet been completed on the potential impact of occult HBV infection, but in a landmark publication from the GHIS, one of the points raised in the revised evaluation of its proposed duration was the lack of data on the un-quantified effects of occult HBV on the attributable risk of HCC in various age groups. Such

investigations should be prioritized in order to determine what role occult infection plays in non-HbsAg positive chronic HBV and non-HCV related HCC.

Another study in the Gambia looking at the seroprevalence of HBV and HCV infections in HIV-1 and HIV-2 infected subjects found 12.2% HBsAg positivity in HIV positive individuals and HCV seroprevelance at 0.9%; both figures were in line with that observed in the general population (Jobarteh et al., 2010). Subjects with low CD4 counts, indicative of poor HIV prognosis tended to have higher HBV viral loads suggesting positive virus proliferation in this disease phenotype. The impact of these contextual factors which form a significant backdrop in many of the populations where any proposed tests would be utilised cannot be ignored. Recent assessments by experts in the field of virology have begun to advocate for greater consensus on the treatment of HBV-HIV co-infection, particularly with the emergence of therapies active against both infections (Wiersma et al., 2011).

The major findings detailed in these chapters all carry the potential for further investigations helping to characterize how these proposed HCC and LC biomarkers may perform in non-African CLD subjects as well as a more thorough description of possible mechanisms at play in bringing about the described expression changes. Such investigations are warranted but will require extensive resources and expertise as well as collaborations between international laboratories in order to promote these candidates to a point where they can be strongly proposed for development into routine laboratory assays.

6.1 Study Limitations

A study of this size, involving various methods and subject sources will always carry some limitations. Some of the major ones identified are summarized below:

- An ideal subject population for a study such as this would have been one were all the LC and HCC cases were confirmed by liver biopsy, as this would not only adhere to the gold standard of diagnosis, but also allow for direct comparison of any proposed markers with AFP measurements taken but not integral to subject classification.
- The small size of the JUTH validation subject group meant that the statistical combination methods applied in chapter 5 could not be performed on it, thus making it impossible to independently validate the GLCS based trends.
- Lack of individual subject data with regard to viral status in the JUTH population did not allow for sub-stratification when creating plasma pools for 2DE.
- It was assumed using US based evidence in all of the non-biopsied HCC's that the identified tumours were primary and localized.
- Though efforts were made to identify one, no well-characterised Caucasian or Chinese subjects could be attained to validate the trends of expression observed in our proposed markers within a non-African population.
- Time and resource constraints meant that only a fraction of the proposed shortlisted biomarker candidates from label-free proteomics could be followed up for more detailed analysis as opposed to performing more targeted scans on all 26 and moving forth with candidates with the most significant levels of differential expression.

7.0 Conclusions

From the key objectives stated in the opening chapter, we can, following the extensive experiments and analyses carried out on the two described subject populations conclude that:

1. Disease specific expression signatures do exist in a number of whole proteins and peptide fragments; discernible by MS.
2. These differences have been quantified and demonstrated to be of statistical significance for Apo A1, α1AT, CC3 and HPX when contrasted against at least one binary control-disease or LC-HCC group comparison.
3. There is some evidence suggesting that the variable processing of fragment proteins in haptoglobin and glyco-enriched Apo A1, α1AT, CC3 and HPX may be linked to progressive disease stages as well as the presence or absence of the HBV.
4. Various statistical combinations of named proteins illustrate the achievement of greater discriminatory performance in determining progressive liver disease than the frequently used ALT liver function test.

References

Aguilar, F., Hussain, S.P. & Cerutti, P. (1993) Aflatoxin B1 induces the transversion of G-->T in codon 249 of the p53 tumor suppressor gene in human hepatocytes. *Proc Natl Acad Sci U S A*, **90**, 8586-8590.

Ahn, Y.H. et al. (2012) A lectin-coupled, targeted proteomic mass spectrometry (MRM MS) platform for identification of multiple liver cancer biomarkers in human plasma. *J Proteomics*,

Akobeng, A.K. (2007) Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr*, **96**, 644-647.

Alonzi, D.S. et al. (2008) Glucosylated free oligosaccharides are biomarkers of endoplasmic- reticulum alpha-glucosidase inhibition. *Biochem J*, **409**, 571-580.

Autrup, H. et al. (1987) Aflatoxin exposure measured by urinary excretion of aflatoxin B1-guanine adduct and hepatitis B virus infection in areas with different liver cancer incidence in Kenya. *Cancer Res*, **47**, 3430-3433.

Bah, E., Hall, A.J. & Inskip, H.M. (1990) The first 2 years of the Gambian National Cancer Registry. *Br J Cancer*, **62**, 647-650.

Bah, E. et al. (2001) Cancer in the Gambia: 1988-97. *Br J Cancer*, **84**, 1207-1214.

Balkwill, F. & Mantovani, A. (2001) Inflammation and cancer: back to Virchow? *Lancet*, **357**, 539-545.

Bantscheff, M. et al. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem*, **389**, 1017-1031.

Barin, F. et al. (1981) Cross-sectional and longitudinal epidemiology of hepatitis B in Senegal. *Prog Med Virol*, **27**, 148-162.

Bartholomew, M.M. et al. (1997) Hepatitis-B-virus resistance to lamivudine given for recurrent infection after orthotopic liver transplantation. *Lancet*, **349**, 20-22.

Bartosch, B. et al. (2009) Hepatitis C virus-induced hepatocarcinogenesis. *J Hepatol*, **51**, 810-820.

Bataller, R. & Brenner, D.A. (2005) Liver fibrosis. *J Clin Invest*, **115**, 209-218.

Batycka, M. et al. (2006) Ultra-fast tandem mass spectrometry scanning combined with monolithic column liquid chromatography increases throughput in proteomic analysis. *Rapid Commun Mass Spectrom*, **20**, 2074-2080.

Beasley, R.P. et al. (1981) Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. *Lancet*, **2**, 1129-1133.

Beasley, R.P. et al. (1982) Incidence of hepatitis B virus infections in preschool children in Taiwan. *J Infect Dis*, **146**, 198-204.

Beasley, R.P. et al. (1983) Efficacy of hepatitis B immune globulin for prevention of perinatal transmission of the hepatitis B virus carrier state: final report of a randomized double-blind, placebo-controlled trial. *Hepatology*, **3**, 135-141.

Beasley, R.P. et al. (1977) The e antigen and vertical transmission of hepatitis B surface antigen. *Am J Epidemiol*, **105**, 94-98.

Behne, T. & Copur, M.S. (2012) Biomarkers for hepatocellular carcinoma. *Int J Hepatol*, **2012**, 859076.

Bensadoun, A. & Weinstein, D. (1976) Assay of proteins in the presence of interfering materials. *Anal Biochem*, **70**, 241-250.

Benyon, R.C. & Iredale, J.P. (2000) Is liver fibrosis reversible? *Gut*, **46**, 443-446.

Bialecki, E.S. & Di Bisceglie, A.M. (2005) Diagnosis of hepatocellular carcinoma. *HPB (Oxford)*, **7**, 26-34.

Blomme, B. et al. (2009) Alteration of protein glycosylation in liver diseases. *J Hepatol*, **50**, 592-603.

Blonski, W., Kotlyar, D.S. & Forde, K.A. (2010) Non-viral causes of hepatocellular carcinoma. *World J Gastroenterol*, **16**, 3603-3615.

Bonis, P.A., Friedman, S.L. & Kaplan, M.M. (2001) Is liver fibrosis reversible? *N Engl J Med*, **344**, 452-454.

Bouchard, M.J. & Schneider, R.J. (2004) The enigmatic X gene of hepatitis B virus. *J Virol*, **78**, 12725-12734.

Brechot, C. et al. (1980) Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. *Nature*, **286**, 533-535.

Bressac, B. et al. (1991) Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature*, **350**, 429-431.

Bruix, J. & Sherman, M. (2011) Management of hepatocellular carcinoma: an update. *Hepatology*, **53**, 1020-1022.

Bunkenborg, J. et al. (2004) Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics*, **4**, 454-465.

Carman, W.F. et al. (1990) Vaccine-induced escape mutant of hepatitis B virus. *Lancet*, **336**, 325-329.

Chan, H.L. et al. (2004) Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. *Gut*, **53**, 1494-1498.

Chang, M.H. (2007) Hepatitis B virus infection. *Semin Fetal Neonatal Med*, **12**, 160-167.

Chang, M.H. et al. (2005) Prevention of hepatocellular carcinoma by universal vaccination against hepatitis B virus: the effect and problems. *Clin Cancer Res*, **11**, 7953-7957.

Chang, W.C. et al. (2007) Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited. *Anal Chim Acta*, **582**, 1-9.

Chang, W.Y. & Chuang, W.L. (1988) Complements as new diagnostic tools of hepatocellular carcinoma in cirrhotic patients. *Cancer*, **62**, 227-232.

Charlwood, J. et al. (2001) Analysis of N-linked oligosaccharides: progress towards the characterisation of glycoprotein-linked carbohydrates. *Biomol Eng*, **18**, 229-240.

Chelius, D. & Bondarenko, P.V. (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res*, **1**, 317-323.

Chemin, I. & Zoulim, F. (2009) Hepatitis B virus induced hepatocellular carcinoma. *Cancer Lett*, **286**, 52-59.

Chen, C.J. et al. (2006) Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA*, **295**, 65-73.

Chen, J.Y. et al. (1986) Detection of hepatitis B virus DNA in hepatocellular carcinoma. *Br J Exp Pathol*, **67**, 279-288.

Choo, Q.L. et al. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*, **244**, 359-362.

Colombo, M. (1999) Hepatitis C virus and hepatocellular carcinoma. *Semin Liver Dis*, **19**, 263-269.

Comunale, M.A. et al. (2006) Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. *J Proteome Res*, **5**, 308-315.

Comunale, M.A. et al. (2010) Linkage specific fucosylation of alpha-1-antitrypsin in liver cirrhosis and cancer patients: implications for a biomarker of hepatocellular carcinoma. *PLoS One*, **5**, e12419.

Comunale, M.A. et al. (2009) Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. *J Proteome Res*, **8**, 595-602.

Comunale, M.A. et al. (2011) Novel changes in glycosylation of serum Apo-J in patients with hepatocellular carcinoma. *Cancer Epidemiol Biomarkers Prev*, **20**, 1222-1229.

Constantin, C.V. et al. (2010) Cirrhosis and Chronic Viral Hepatitis as Risk Factors for Hepatocellular Carcinoma: Romanian Single-clinic Experience. *Maedica (Buchar)*, **5**, 265-270.

Cooke, G.S., Main, J. & Thursz, M.R. (2010) Treatment for hepatitis B. *BMJ*, **340**, b5429.

Coursaget, P. et al. (1987) Age- and sex-related study of hepatitis B virus chronic carrier state in infants from an endemic area (Senegal). *J Med Virol*, **22**, 1-5.

Cummings, R.D. & Kornfeld, S. (1982) Fractionation of asparagine-linked oligosaccharides by serial lectin-Agarose affinity chromatography. A rapid, sensitive, and specific technique. *J Biol Chem*, **257**, 11235-11240.

Daniele, B. et al. (2004) Alpha-fetoprotein and ultrasonography screening for hepatocellular carcinoma. *Gastroenterology*, **127**, S108-12.

de Bruijn, M.H. & Fey, G.H. (1985) Human complement component C3: cDNA coding sequence and derived primary structure. *Proc Natl Acad Sci U S A*, **82**, 708-712.

Delhaye, M. et al. (1996) Relationship between hepatocyte proliferative activity and liver functional reserve in human cirrhosis. *Hepatology*, **23**, 1003-1011.

Di Bisceglie, A.M. (1997) Hepatitis C and hepatocellular carcinoma. *Hepatology*, **26**, 34S-38S.

Dodd, G.D.r. et al. (1992) Detection of malignant tumors in end-stage cirrhotic livers: efficacy of sonography as a screening technique. *AJR Am J Roentgenol*, **159**, 727-733.

DL, E. & Groopman, J.D. (1994) *The toxicology of aflatoxins: Human health, vetinary and agricultural significance*. Academic Press, San Diego, CA.

Echejoh, G.O. et al. (2008) Hepatocellular carcinoma in Jos, Nigeria. *Niger J Med*, **17**, 210-213.

El-Kasti, M.M. et al. (2011) Urinary peptide profiling identifies a panel of putative biomarkers for diagnosing and staging endometriosis. *Fertility and sterility*.

El-Serag, H.B. & Rudolph, K.L. (2007) Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*, **132**, 2557-2576.

Everley, R.A. et al. (2007) Measurement of aflatoxin and aflatoxin metabolites in urine by liquid chromatography-tandem mass spectrometry. *J Anal Toxicol*, **31**, 150-156.

Farazi, P.A. & DePinho, R.A. (2006) Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer*, **6**, 674-687.

Fattovich, G. et al. (2004) Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology*, **127**, S35-50.

Ferlay, J. et al. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*, **127**, 2893-2917.

Firnhaber, C. et al. (2009) Occult hepatitis B virus infection in patients with isolated core antibody and HIV co-infection in an urban clinic in Johannesburg, South Africa. *Int J Infect Dis*, **13**, 488-492.

Forbes, S.J. et al. (2004) A significant proportion of myofibroblasts are of bone marrow origin in human liver fibrosis. *Gastroenterology*, **126**, 955-963.

Fortuin, M. et al. (1994) Breakthrough infections and identification of a viral variant in Gambian children immunized with hepatitis B vaccine. *J Infect Dis*, **169**, 1374-1376.

Frantikova, V. et al. (1984) Amino acid sequence of the N-terminal region of human hemopexin. *FEBS Lett*, **178**, 213-216.

Fried, M.W. et al. (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med*, **347**, 975-982.

Friedman, S.L. (2003) Liver fibrosis -- from bench to bedside. *J Hepatol*, **38 Suppl 1**, S38-53.

Friedman, S.L. et al. (1985) Hepatic lipocytes: the principal collagen-producing cells of normal rat liver. *Proc Natl Acad Sci U S A*, **82**, 8681-8685.

Fu, Q. et al. (2005) A robust, streamlined, and reproducible method for proteomic analysis of serum by delipidation, albumin and IgG depletion, and two-dimensional gel electrophoresis. *Proteomics*, **5**, 2656-2664.

Fuster, M.M. & Esko, J.D. (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer*, **5**, 526-542.

Ganem, D. & Prince, A.M. (2004) Hepatitis B virus infection--natural history and clinical consequences. *N Engl J Med*, **350**, 1118-1129.

Gangadharan, B. et al. (2011) New approaches for biomarker discovery: the search for liver fibrosis markers in hepatitis C patients. *J Proteome Res*, **10**, 2643-2650.

Gangadharan, B. et al. (2007) Novel serum biomarker candidates for liver fibrosis in hepatitis C patients. *Clin Chem*, **53**, 1792-1799.

Gangadharan, B. et al. (2012) Discovery of novel biomarker candidates for liver fibrosis in hepatitis C patients: a preliminary study. *PLoS One*, **7**, e39603.

Giannini, E.G. et al. (2012) Alpha-fetoprotein has no prognostic role in small hepatocellular carcinoma identified during surveillance in compensated cirrhosis. *Hepatology*,

Gines, P. et al. (2004) Management of cirrhosis and ascites. *N Engl J Med*, **350**, 1646-1654.

(2004) Global burden of disease (GBD) for hepatitis C. *J Clin Pharmacol*, **44**, 20-29.

Goding, J.W. (1978) Use of staphylococcal protein A as an immunological reagent. *J Immunol Methods*, **20**, 241-253.

Goldman, R. et al. (2007) Candidate markers for the detection of hepatocellular carcinoma in low-molecular weight fraction of serum. *Carcinogenesis*, **28**, 2149-2153.

Groopman, J.D. & Kensler, T.W. (2005) Role of metabolism and viruses in aflatoxin-induced liver cancer. *Toxicol Appl Pharmacol*, **206**, 131-137.

Gupta, S., Bent, S. & Kohlwes, J. (2003) Test characteristics of alpha-fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *Ann Intern Med*, **139**, 46-50.

Gygi, S.P. et al. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, **17**, 994-999.

Hammel, P. et al. (2001) Regression of liver fibrosis after biliary drainage in patients with chronic pancreatitis and stenosis of the common bile duct. *N Engl J Med*, **344**, 418-423.

Hann, H.W. et al. (2010) Analysis of GP73 in patients with HCC as a function of anti-cancer treatment. *Cancer Biomark*, **7**, 269-273.

He, C. et al. (2001) Prevalence of vaccine-induced escape mutants of hepatitis B virus in the adult population in China: a prospective study in 176 restaurant employees. *J Gastroenterol Hepatol*, **16**, 1373-1377.

He, Q.Y. et al. (2003) Serum biomarkers of hepatitis B virus infected liver inflammation: a proteomic study. *Proteomics*, **3**, 666-674.

Hedlund, M. et al. (2008) alpha 2-6-Linked sialic acids on N-glycans modulate carcinoma differentiation in vivo. Cancer Res, 68, 388-394.

Hildt, E. et al. (2002) The PreS2 activator MHBs(t) of hepatitis B virus activates c-raf-1/Erk2 signaling in transgenic mice. *EMBO J*, **21**, 525-535.

Hino, O. et al. (1984) Detection of hepatitis B virus DNA in hepatocellular carcinomas in Japan. *Hepatology*, **4**, 90-95.

Hjelm, H., Hjelm, K. & Sjoquist, J. (1972) Protein A from Staphylococcus aureus. Its isolation by affinity chromatography and its use as an immunosorbent for isolation of immunoglobulins. *FEBS Lett*, **28**, 73-76.

Holbrook, M.L. (2008) Cost-effective pharmacologic therapies to treat chronic hepatitis B: how far have we really come? *J Manag Care Pharm*, **14**, 65-69.

Hsu, I.C. et al. (1991) Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature*, **350**, 427-428.

Hsu, S.C. et al. (1993) Horizontal transmission of hepatitis B virus in children. *J Pediatr Gastroenterol Nutr*, **16**, 66-69.

Huang, S.N. & Chisari, F.V. (1995) Strong, sustained hepatocellular proliferation precedes hepatocarcinogenesis in hepatitis B surface antigen transgenic mice. *Hepatology*, **21**, 620-626.

Hung, C.H. et al. (2003) Correlation between ultrasonographic and pathologic diagnoses of hepatitis B and C virus-related cirrhosis. *J Gastroenterol*, **38**, 153-157.

Hyams, K.C. (1995) Risks of chronicity following acute hepatitis B virus infection: a review. *Clin Infect Dis*, **20**, 992-1000.

Jobarteh, M. et al. (2010) Seroprevalence of hepatitis B and C virus in HIV-1 and HIV-2 infected Gambians. *Virol J*, **7**, 230.

Kamoto, T. et al. (2002) Lectin-reactive alpha-fetoprotein (AFP-L3%) curability and prediction of clinical course after treatment of non-seminomatous germ cell tumors. *Jpn J Clin Oncol*, **32**, 472-476.

Kang, Y.K. et al. (2004) Overexpression of clusterin in human hepatocellular carcinoma. *Hum Pathol*, **35**, 1340-1346.

Karas, M., Gluckmann, M. & Schafer, J. (2000) Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom*, **35**, 1-12.

Karas, M. & Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, **60**, 2299-2301.

Karthigesu, V.D. et al. (1994) A novel hepatitis B virus variant in the sera of immunized children. *J Gen Virol*, **75**, 443-448.

Kew, M.C. (2010a) Epidemiology of chronic hepatitis B virus infection, hepatocellular carcinoma, and hepatitis B virus-induced hepatocellular carcinoma. *Pathol Biol (Paris)*, **58**, 273-277.

Kew, M.C. (2010b) Hepatocellular carcinoma in African Blacks: Recent progress in etiology and pathogenesis. *World J Hepatol*, **2**, 65-73.

Khien, V.V. et al. (2001) Clinical evaluation of lentil lectin-reactive alpha-fetoprotein-L3 in histology-proven hepatocellular carcinoma. *Int J Biol Markers*, **16**, 105-111.

Kidd-Ljunggren, K., Miyakawa, Y. & Kidd, A.H. (2002) Genetic variability in hepatitis B viruses. *J Gen Virol*, **83**, 1267-1280.

Kim, W.R. et al. (2008) Serum activity of alanine aminotransferase (ALT) as an indicator of health and disease. *Hepatology*, **47**, 1363-1370.

Kirk, G.D., Bah, E. & Montesano, R. (2006) Molecular epidemiology of human liver cancer: insights into etiology, pathogenesis and prevention from The Gambia, West Africa. *Carcinogenesis*, **27**, 2070-2082.

Kirk, G.D. et al. (2000) Ser-249 p53 mutations in plasma DNA of patients with hepatocellular carcinoma from The Gambia. *J Natl Cancer Inst*, **92**, 148-153.

Kirk, G.D. et al. (2004) The Gambia Liver Cancer Study: Infection with hepatitis B and C and the risk of hepatocellular carcinoma in West Africa. *Hepatology*, **39**, 211-219.

Kolarich, D. et al. (2006) Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics*, **6**, 3369-3380.

Kremsdorf, D. et al. (2006) Hepatitis B virus-related hepatocellular carcinoma: paradigms for viral-related human carcinogenesis. *Oncogene*, **25**, 3823-3833.

Lau, S.H. et al. (2006) Clusterin plays an important role in hepatocellular carcinoma metastasis. *Oncogene*, **25**, 1242-1250.

Lee, I.N. et al. (2006) Identification of complement C3a as a candidate biomarker in human chronic hepatitis C and HCV-related hepatocellular carcinoma using a proteomics approach. *Proteomics*, **6**, 2865-2873.

Leung, N. (2002) Treatment of chronic hepatitis B: case selection and duration of therapy. *J Gastroenterol Hepatol*, **17**, 409-414.

Li, L. et al. (2004) Data mining techniques for cancer detection using serum proteomic profiling. *Artif Intell Med*, **32**, 71-83.

Lin, D.Y. et al. (1993) Ultrasonographic changes of early liver cirrhosis in chronic hepatitis B: a longitudinal study. *J Clin Ultrasound*, **21**, 303-308.

Liscum, L. et al. (1983) 3-Hydroxy-3-methylglutaryl-CoA reductase: a transmembrane glycoprotein of the endoplasmic reticulum with N-linked "high-mannose" oligosaccharides. *Proc Natl Acad Sci U S A*, **80**, 7165-7169.

Liu, H., Sadygov, R.G. & Yates, J.R.r. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*, **76**, 4193-4201.

Liu, L. et al. (2006) Sorafenib blocks the RAF/MEK/ERK pathway, inhibits tumor angiogenesis, and induces tumor cell apoptosis in hepatocellular carcinoma model PLC/PRF/5. *Cancer Res*, **66**, 11851-11858.

Liu, S. et al. (2009) Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. *J Natl Cancer Inst*, **101**, 1066-1082.

Liu, T. et al. (2005) Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res*, **4**, 2070-2080.

Liu, T. et al. (2011) Proteomic profiling of hepatitis B virus-related hepatocellular carcinoma with magnetic bead-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Acta Biochim Biophys Sin (Shanghai)*, **43**, 542-550.

Livraghi, T. et al. (2008) Sustained complete response and complications rates after radiofrequency ablation of very early hepatocellular carcinoma in cirrhosis: Is resection still the treatment of choice? *Hepatology*, **47**, 82-89.

Lleonart, M.E. et al. (2005) Quantitative analysis of plasma TP53 249Ser-mutated DNA by electrospray ionization mass spectrometry. *Cancer Epidemiol Biomarkers Prev*, **14**, 2956-2962.

Llovet, J.M. & Bruix, J. (2003) Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival. *Hepatology*, **37**, 429-442.

Llovet, J.M. et al. (2008) Sorafenib in advanced hepatocellular carcinoma. *New England Journal of Medicine*, **359**, 378-390.

Llovet, J.M., Schwartz, M. & Mazzaferro, V. (2005) Resection and liver transplantation for hepatocellular carcinoma. *Semin Liver Dis*, **25**, 181-200.

Luk, J.M. & Liu, A.M. (2011) Proteomics of hepatocellular carcinoma in Chinese patients. *OMICS*, **15**, 261-266.

Magder, L.S. et al. (2005) Estimation of the risk of transmission of hepatitis C between spouses in Egypt based on seroprevalence data. *Int J Epidemiol*, **34**, 160-165.

Marrero, J.A. et al. (2005) GP73, a resident Golgi glycoprotein, is a novel serum marker for hepatocellular carcinoma. *J Hepatol*, **43**, 1007-1012.

Matsubara, K. & Tokino, T. (1990) Integration of hepatitis B virus DNA and its implications for hepatocarcinogenesis. *Mol Biol Med*, **7**, 243-260.

Mazzaferro, V. et al. (2009) Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. *Lancet Oncol*, **10**, 35-43.

Mboto, C.I. et al. (2005) Hepatocellular Carcinoma in The Gambia and the role of Hepatitis B and Hepatitis C. *Int Semin Surg Oncol*, **2**, 20.

Mehta, A. et al. (2012) Increased levels of tetra-antennary N-linked glycan but not core fucosylation are associated with hepatocellular carcinoma tissue. *Cancer Epidemiol Biomarkers Prev*, **21**, 925-933.

Mendy, M. et al. (2008) Envelope protein variability among HBV-Infected asymptomatic carriers and immunized children with breakthrough infections. *J Med Virol*, **80**, 1537-1546.

Mendy, M. et al. (2005) Hepatitis B surface antigenaemia and alpha-foetoprotein detection from dried blood spots: applications to field-based studies and to clinical care in hepatitis B virus endemic areas. *J Viral Hepat*, **12**, 642-647.

Mendy, M. & Walton, R. (2009) Molecular pathogenesis and early detection of hepatocellular carcinoma--perspectives from West Africa. *Cancer Lett*, **286**, 44-51.

Mendy, M.E. et al. (2010) Hepatitis B viral load and risk for liver cirrhosis and hepatocellular carcinoma in The Gambia, West Africa. *J Viral Hepat*, **17**, 115-122.

Monto, A. & Wright, T.L. (2001) The epidemiology and prevention of hepatocellular carcinoma. *Semin Oncol*, **28**, 441-449.

Moradpour, D., Penin, F. & Rice, C.M. (2007) Replication of hepatitis C virus. *Nat Rev Microbiol*, **5**, 453-463.

Morgan, T.R., Mandayam, S. & Jamal, M.M. (2004) Alcohol and hepatocellular carcinoma. *Gastroenterology*, **127**, S87-96.

Moriya, K. et al. (1998) The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat Med*, **4**, 1065-1067.

Moriya, K. et al. (1997) Hepatitis C virus core protein induces hepatic steatosis in transgenic mice. *J Gen Virol*, **78**, 1527-1531.

Morota, K. et al. (2011) A comparative evaluation of Golgi protein-73, fucosylated hemopexin, alpha-fetoprotein, and PIVKA-II in the serum of patients with chronic hepatitis, cirrhosis, and hepatocellular carcinoma. *Clin Chem Lab Med*, **49**, 711-718.

Motulsky, H. & Christopoulos, A. (2004) *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press, USA,

Motulsky, H.J. (2003) Prism 4 statistics guide. *Statistical analysis for laboratory and clinical researchers. San Diego: GraphPad Software*, 61-65.

Mphahlele, M.J. et al. (2006) High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. *J Clin Virol*, **35**, 14-20.

Nafee, A.M. et al. (2012) Clinical significance of serum clusterin as a biomarker for evaluating diagnosis and metastasis potential of viral-related hepatocellular carcinoma. *Clin Biochem*,

Namieno, T. et al. (1995) Age-related, different clinicopathologic features of hepatocellular carcinoma patients. *Ann Surg*, **221**, 308-314.

Narayanan, S. (1994) Sialic acid as a tumor marker. Ann Clin Lab Sci, 24, 376-384.

Ndububa, D.A. et al. (2001) Primary hepatocellular carcinoma in Ile-Ife, Nigeria: a prospective study of 154 cases. *Niger J Med*, **10**, 59-63.

Neilson, K.A., Mariani, M. & Haynes, P.A. (2011) Quantitative proteomic analysis of cold-responsive proteins in rice. *Proteomics*, **11**, 1696-1706.

Nguyen, M.H. et al. (2002) Racial differences in effectiveness of alpha-fetoprotein for diagnosis of hepatocellular carcinoma in hepatitis C virus cirrhosis. *Hepatology*, **36**, 410-417.

Niu, D., Feng, H. & Chen, W.N. (2010) Proteomic analysis of HBV-associated HCC: insights on mechanisms of disease onset and biomarker discovery. *J Proteomics*, **73**, 1283-1290.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*, **250**, 4007-4021.

Okuda, H. (2007) Hepatocellular carcinoma development in cirrhosis. *Best Pract Res Clin Gastroenterol*, **21**, 161-173.

Okuda, K. (2002) Natural history of hepatocellular carcinoma including fibrolamellar and hepato-cholangiocarcinoma variants. *J Gastroenterol Hepatol*, **17**, 401-405.

Okuda, K. et al. (1982) Hepatocellular carcinoma arising in noncirrhotic and highly cirrhotic livers: a comparative study of histopathology and frequency of hepatitis B markers. *Cancer*, **49**, 450-455.

Okuda, K. et al. (1985) Natural history of hepatocellular carcinoma and prognosis in relation to treatment. Study of 850 patients. *Cancer*, **56**, 918-928.

Olubuyide, I.O. (1992) The natural history of primary liver cell carcinoma: a study of 89 untreated adult Nigerians. *Cent Afr J Med*, **38**, 25-30.

Ong, S.E., Kratchmarova, I. & Mann, M. (2003) Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res*, **2**, 173-181.

Ozturk, M. (1991) p53 mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet*, **338**, 1356-1359.

Park, N.H., Song, I.H. & Chung, Y.H. (2007) Molecular Pathogenesis of Hepatitis-B-virus-associated Hepatocellular Carcinoma. *Gut Liver*, **1**, 101-117.

Parkin, D.M. (2001) Global cancer statistics in the year 2000. *Lancet Oncol*, **2**, 533-543.

Parkin, D.M. (2006) The global health burden of infection-associated cancers in the year 2002. *Int J Cancer*, **118**, 3030-3044.

Parkin, D.M. et al. (2005) Global cancer statistics, 2002. *CA Cancer J Clin*, **55**, 74-108.

Parkin, D.M. et al. (2002) Cancer incidence in five continents. Volume VIII. *IARC Sci Publ*, 1-781.

Peers, F. et al. (1987) Aflatoxin exposure, hepatitis B virus infection and liver cancer in Swaziland. *Int J Cancer*, **39**, 545-553.

Peers, F.G., Gilman, G.A. & Linsell, C.A. (1976) Dietary aflatoxins and human liver cancer. A study in Swaziland. *Int J Cancer*, **17**, 167-176.

Peterson, M.S. & Baron, R.L. (2001) Radiologic diagnosis of hepatocellular carcinoma. *Clin Liver Dis*, **5**, 123-144.

Plentz, R.R. et al. (2004) Hepatocellular telomere shortening correlates with chromosomal instability and the development of human hepatoma. *Hepatology*, **40**, 80-86.

Plentz, R.R. et al. (2007) Telomere shortening and inactivation of cell cycle checkpoints characterize human hepatocarcinogenesis. *Hepatology*, **45**, 968-976.

Podwojski, K. et al. (2010) Peek a peak: a glance at statistics for quantitative label-free proteomics. *Expert Rev Proteomics*, **7**, 249-261.

Pollicino, T. et al. (2004) Hepatitis B virus maintains its pro-oncogenic properties in the case of occult HBV infection. *Gastroenterology*, **126**, 102-110.

Pons, F., Varela, M. & Llovet, J.M. (2005) Staging systems in hepatocellular carcinoma. *HPB (Oxford)*, **7**, 35-41.

Pontisso, P. et al. (2004) Overexpression of squamous cell carcinoma antigen variants in hepatocellular carcinoma. *Br J Cancer*, **90**, 833-837.

Poon, T.C. et al. (2003) Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. *Clin Chem*, **49**, 752-760.

Pusch, W. et al. (2003) Mass spectrometry-based clinical proteomics. *Pharmacogenomics*, **4**, 463-476.

Qian, G.S. et al. (1994) A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol Biomarkers Prev*, **3**, 3-10.

Qiao, B. et al. (2012a) Detection and identification of peroxiredoxin 3 as a biomarker in hepatocellular carcinoma by a proteomic approach. *Int J Mol Med*, **29**, 832-840.

Qiao, B. et al. (2012b) Detection and identification of peroxiredoxin 3 as a biomarker in hepatocellular carcinoma by a proteomic approach. *Int J Mol Med*, **29**, 832-840.

Qiu, J.G. et al. (2008) Screening and detection of portal vein tumor thrombi-associated serum low molecular weight protein biomarkers in human hepatocellular carcinoma. *J Cancer Res Clin Oncol*, **134**, 299-305.

Raimondo, G. et al. (2008) Statements from the Taormina expert meeting on occult hepatitis B virus infection. *J Hepatol*, **49**, 652-657.

Ramadori, G. & Saile, B. (2004) Portal tract fibrogenesis in the liver. *Lab Invest*, **84**, 153-159.

Regimbeau, J.M. et al. (2004) Obesity and diabetes as a risk factor for hepatocellular carcinoma. *Liver Transpl*, **10**, S69-73.

Ressom, H.W. et al. (2012) Utilization of metabolomics to identify serum biomarkers for hepatocellular carcinoma in patients with liver cirrhosis. *Anal Chim Acta*, **743C**, 90-100.

Ross, R.K. et al. (1992) Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet*, **339**, 943-946.

Ryder, R.W. (1982) A proposal for a long term controlled trial of hepatitis B virus vaccine in The Gambia to prevent hepatitis and liver cancer. *Gambia Medical and Health Department*,

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

Ryder, S.D. et al. (1996) Chemoembolisation with lipiodol and doxorubicin: applicability in British patients with hepatocellular carcinoma. *Gut*, **38**, 125-128.

Safadi, R. & Friedman, S.L. (2002) Hepatic fibrosis--role of hepatic stellate cell activation. *MedGenMed*, **4**, 27.

Schauer, R. (2000) Achievements and challenges of sialic acid research. *Glycoconj J*, **17**, 485-499.

Schmucker, D.L. (2005) Age-related changes in liver structure and function: Implications for disease ? *Exp Gerontol*, **40**, 650-659.

Schuppan, D. et al. (2003) Hepatitis C and liver fibrosis. *Cell Death Differ*, **10 Suppl 1**, S59-67.

Schwartz, D.C. & Cantor, C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, **37**, 67-75.

Shank, R.C. et al. (1972) Dietary aflatoxins and human liver cancer. IV. Incidence of primary liver cancer in two municipal populations of Thailand. *Food Cosmet Toxicol*, **10**, 171-179.

Shen, Q. et al. (2012) Serum DKK1 as a protein biomarker for the diagnosis of hepatocellular carcinoma: a large-scale, multicentre study. *Lancet Oncol*, **13**, 817-826.

Sherman, M. (2001) Alphafetoprotein: an obituary. *J Hepatol*, **34**, 603-605.

Sideso, E. et al. (2010) Assessing the quality and reproducibility of a proteomic platform for clinical stroke biomarker discovery. *Translational Stroke Research*, 1-11

Sighoko, D. et al. (2011) Increase in female liver cancer in the Gambia, West Africa: evidence from 19 years of population-based cancer registration (1988-2006). *PLoS One*, **6**, e18415.

Silva, J.C. et al. (2006) Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. *Mol Cell Proteomics*, **5**, 589-607.

Soresi, M. et al. (2003) Usefulness of alpha-fetoprotein in the diagnosis of hepatocellular carcinoma. *Anticancer Res*, **23**, 1747-1753.

Spangenberg, H.C., Thimme, R. & Blum, H.E. (2006) Serum markers of hepatocellular carcinoma. *Semin Liver Dis*, **26**, 385-390.

Staib, F. et al. (2003) TP53 and liver carcinogenesis. *Hum Mutat*, **21**, 201-216.

Staufer, K. et al. (2012) High toxicity of sorafenib for recurrent hepatocellular carcinoma after liver transplantation. *Transpl Int*,

Steel, L.F. et al. (2003) A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. *Proteomics*, **3**, 601-609.

Stefaniuk, P., Cianciara, J. & Wiercinska-Drapalo, A. (2010) Present and future possibilities for early diagnosis of hepatocellular carcinoma. *World journal of gastroenterology: WJG*, **16**, 418.

Stevens, C.E. et al. (1975) Vertical transmission of hepatitis B antigen in Taiwan. *N Engl J Med*, **292**, 771-774.

Stuver, S.O. (1998) Towards global control of liver cancer? *Semin Cancer Biol*, **8**, 299-306.

Sun, C.A. et al. (2003) Incidence and cofactors of hepatitis C virus-related hepatocellular carcinoma: a prospective study of 12,008 men in Taiwan. *Am J Epidemiol*, **157**, 674-682.

Szymanska, K. et al. (2004) Ser-249TP53 mutation in tumour and plasma DNA of hepatocellular carcinoma patients from a high incidence area in the Gambia, West Africa. *Int J Cancer*, **110**, 374-379.

Takahashi, N., Takahashi, Y. & Putnam, F.W. (1984) Structure of human hemopexin: O-glycosyl and N-glycosyl sites and unusual clustering of tryptophan residues. *Proc Natl Acad Sci U S A*, **81**, 2021-2025.

Taketa, K. et al. (1993) A collaborative study for the evaluation of lectin-reactive alpha-fetoproteins in early detection of hepatocellular carcinoma. *Cancer Res*, **53**, 5419-5423.

Tarao, K. et al. (2000) Close association between high serum ALT and more rapid recurrence of hepatocellular carcinoma in hepatectomized patients with HCV-associated liver cirrhosis and hepatocellular carcinoma. *Intervirology*, **43**, 20-26.

Teramoto, T. et al. (1994) p53 gene abnormalities are closely related to hepatoviral infections and occur at a late stage of hepatocarcinogenesis. *Cancer Res*, **54**, 231-235.

Group, T.G.H.S. (1987) The Gambia Hepatitis Intervention Study. *Cancer Res*, **47**, 5782-5787.

Thursz, M. & Brown, A. (2011) Can antiviral therapy of chronic hepatitis B prevent the development of hepatocellular carcinoma? *Gut*, **60**, 1025-1026.

Thursz, M., Cooke, G.S. & Hall, A.J. (2010) Hepatitis B treatment in resource poor settings: time for action. *Trop Med Int Health*, **15**, 2-4.

Torzilli, G., Belghiti, J. & Makuuchi, M. (2004) Differences and similarities in the approach to hepatocellular carcinoma between Eastern and Western institutions. *Liver Transpl*, **10**, S1-2.

Trotter, J.F. (2002) Adult-to-adult Right Hepatic Lobe Living Donor Liver Transplantation. *Curr Treat Options Gastroenterol*, **5**, 491-501.

Trudgian, D.C. et al. (2011) Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline. *Proteomics*, **11**, 2790-2797.

Turner, G.A. (1992) N-glycosylation of serum proteins in disease and its investigation using lectins. *Clin Chim Acta*, **208**, 149-171.

Vall Mayans, M. et al. (1990) Risk factors for transmission of hepatitis B virus to Gambian children. *Lancet*, **336**, 1107-1109.

van Bommel, F. et al. (2010) Long-term efficacy of tenofovir monotherapy for hepatitis B virus-monoinfected patients after failure of nucleoside/nucleotide analogues. *Hepatology*, **51**, 73-80.

van Gijssel, H.E. et al. (1997) p53 protein expression by hepatocarcinogens in the rat liver and its potential role in mitoinhibition of normal hepatocytes as a mechanism of hepatic tumour promotion. *Carcinogenesis*, **18**, 1027-1033.

Van Rensburg, S.J. et al. (1985) Hepatocellular carcinoma and dietary aflatoxin in Mozambique and Transkei. *Br J Cancer*, **51**, 713-726.

Vandelli, C. et al. (2004) Lack of evidence of sexual transmission of hepatitis C among monogamous couples: results of a 10-year prospective follow-up study. *Am J Gastroenterol*, **99**, 855-859.

Vertes, A., Irinyi, G. & Gijbels, R. (1993) Hydrodynamic model of matrix-assisted laser desorption mass spectrometry. *Analytical Chemistry*, **65**, 2389-2393.

Viviani, S. et al. (2008) 20 years into the Gambia Hepatitis Intervention Study: assessment of initial hypotheses and prospects for evaluation of protective effectiveness against liver cancer. *Cancer Epidemiol Biomarkers Prev*, **17**, 3216-3223.

Viviani, S. et al. (1999) Hepatitis B vaccination in infancy in The Gambia: protection against carriage at 9 years of age. *Vaccine*, **17**, 2946-2950.

Wang, A.G. et al. (2009a) Non-structural 5A protein of hepatitis C virus induces a range of liver pathology in transgenic mice. *J Pathol*, **219**, 253-262.

Wang, M. et al. (2009b) Novel fucosylated biomarkers for the early detection of hepatocellular carcinoma. *Cancer Epidemiol Biomarkers Prev*, **18**, 1914-1921.

Ward, R.P. & Kugelmas, M. (2005) Using pegylated interferon and ribavirin to treat patients with chronic hepatitis C. *Am Fam Physician*, **72**, 655-662.

Washburn, M.P., Wolters, D. & Yates, J.R.r. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol, 19, 242-247.

Wessel, D. & Flugge, U.I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem*, **138**, 141-143.

Whiteaker, J.R. et al. (2007) Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal Biochem*, **362**, 44-54.

Whittle, H. et al. (1990) The pattern of childhood hepatitis B infection in two Gambian villages. *J Infect Dis*, **161**, 1112-1115.

Whittle, H.C. et al. (1983) Hepatitis B virus infection in two Gambian villages. *Lancet*, **1**, 1203-1206.

Whittle, H.C. et al. (1991) Vaccination against hepatitis B and protection against chronic viral carriage in The Gambia. *Lancet*, **337**, 747-750.

Wiersma, S.T. et al. (2011) Treatment of chronic hepatitis B virus infection in resource-constrained settings: expert panel consensus. *Liver Int*, **31**, 755-761.

Wiese, S. et al. (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, **7**, 340-350.

Wild, C.P. et al. (1990) Aflatoxin-albumin adducts in human sera from different regions of the world. *Carcinogenesis*, **11**, 2271-2274.

Wild, C.P. et al. (1992) Field studies of aflatoxin exposure, metabolism and induction of genetic alterations in relation to HBV infection and hepatocellular carcinoma in The Gambia and Thailand. *Toxicol Lett*, **64-65 Spec No**, 455-461.

Wogan, G.N. (1992) Aflatoxin carcinogenesis: interspecies potency differences and relevance for human risk assessment. *Prog Clin Biol Res*, **374**, 123-137.

Organization, W.H. (1983) Prevention of Liver Cancer. *Technical Report Series*, **691**,

Organization, W.H. (2007) Hepatitis B Fact Sheet WHO/204, Revised October 2000. *Geneva: WHO*,

Xu, D. et al. (2008) Novel MMP-9 substrates in cancer cells revealed by a label-free quantitative proteomics approach. *Mol Cell Proteomics*, **7**, 2215-2228.

Yamashiki, N. et al. (1999) Usefulness of Lens culinaris agglutinin A-reactive fraction of alpha-fetoprotein (AFP-L3) as a marker of distant metastasis from hepatocellular carcinoma. *Oncol Rep*, **6**, 1229-1232.

Yang, H.I. et al. (2002) Hepatitis B e antigen and the risk of hepatocellular carcinoma. *N Engl J Med*, **347**, 168-174.

Yang, H.I. et al. (2008) Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma. *J Natl Cancer Inst*, **100**, 1134-1143.

Yao, J.L. (1996) Perinatal transmission of hepatitis B virus infection and vaccination in China. *Gut*, **38 Suppl 2**, S37-8.

Yeh, F.S. et al. (1985) A serological case-control study of primary hepatocellular carcinoma in Guangxi, China. *Cancer Res*, **45**, 872-873.

Yeh, F.S. et al. (1989) Hepatitis B virus, aflatoxins, and hepatocellular carcinoma in southern Guangxi, China. *Cancer Res*, **49**, 2506-2509.

You, S.L., Yang, H.I. & Chen, C.J. (2004) Seropositivity of hepatitis B e antigen and hepatocellular carcinoma. *Ann Med*, **36**, 215-224.

Yu, L.H., Li, N. & Cheng, S.Q. (2011) The Role of Antiviral Therapy for HBV-Related Hepatocellular Carcinoma. *Int J Hepatol*, **2011**, 416459.

Yuen, M.F. et al. (2005) Prognostic determinants for chronic hepatitis B in Asians: therapeutic implications. *Gut*, **54**, 1610-1614.

Zhou, L. et al. (2012) The significance of serum AFP cut-off values, 20 and 400 ng/mL in curatively resected patients with hepatocellular carcinoma and cirrhosis might be of difference. *Hepatogastroenterology*, **59**, 840-843.

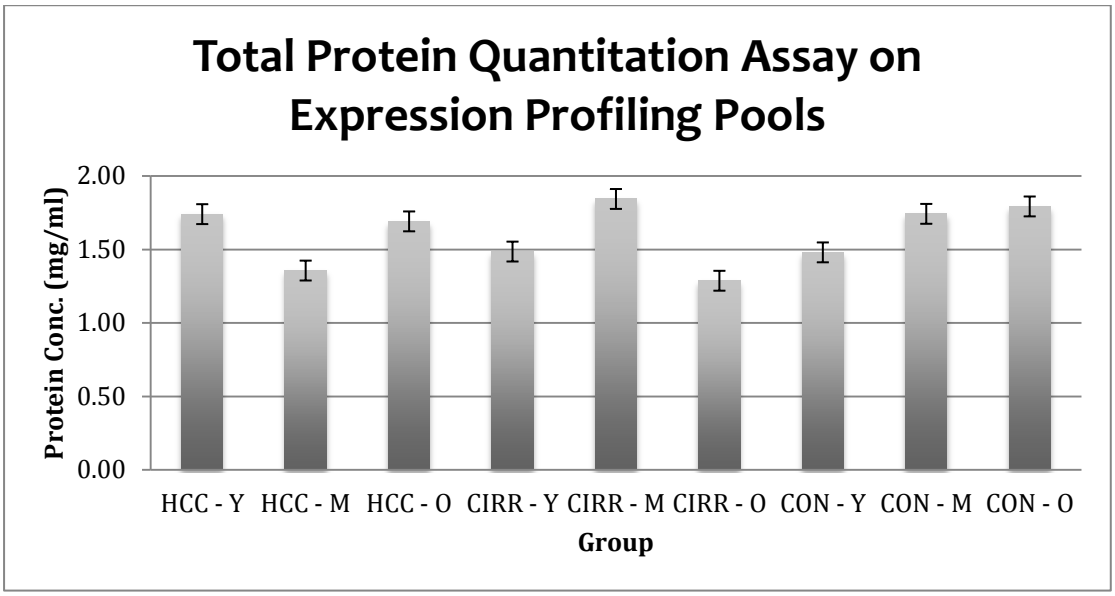
Zhu, W.L. et al. (2009) Abnormal expression of fibrinogen gamma (FGG) and plasma level of fibrinogen in patients with hepatocellular carcinoma. *Anticancer Res*, **29**, 2531-2534.

Zhu, Z.J. et al. (2011) [Adult-to-adult living donor liver transplantation with 94 cases report]. *Zhonghua Wai Ke Za Zhi*, **49**, 1100-1104.

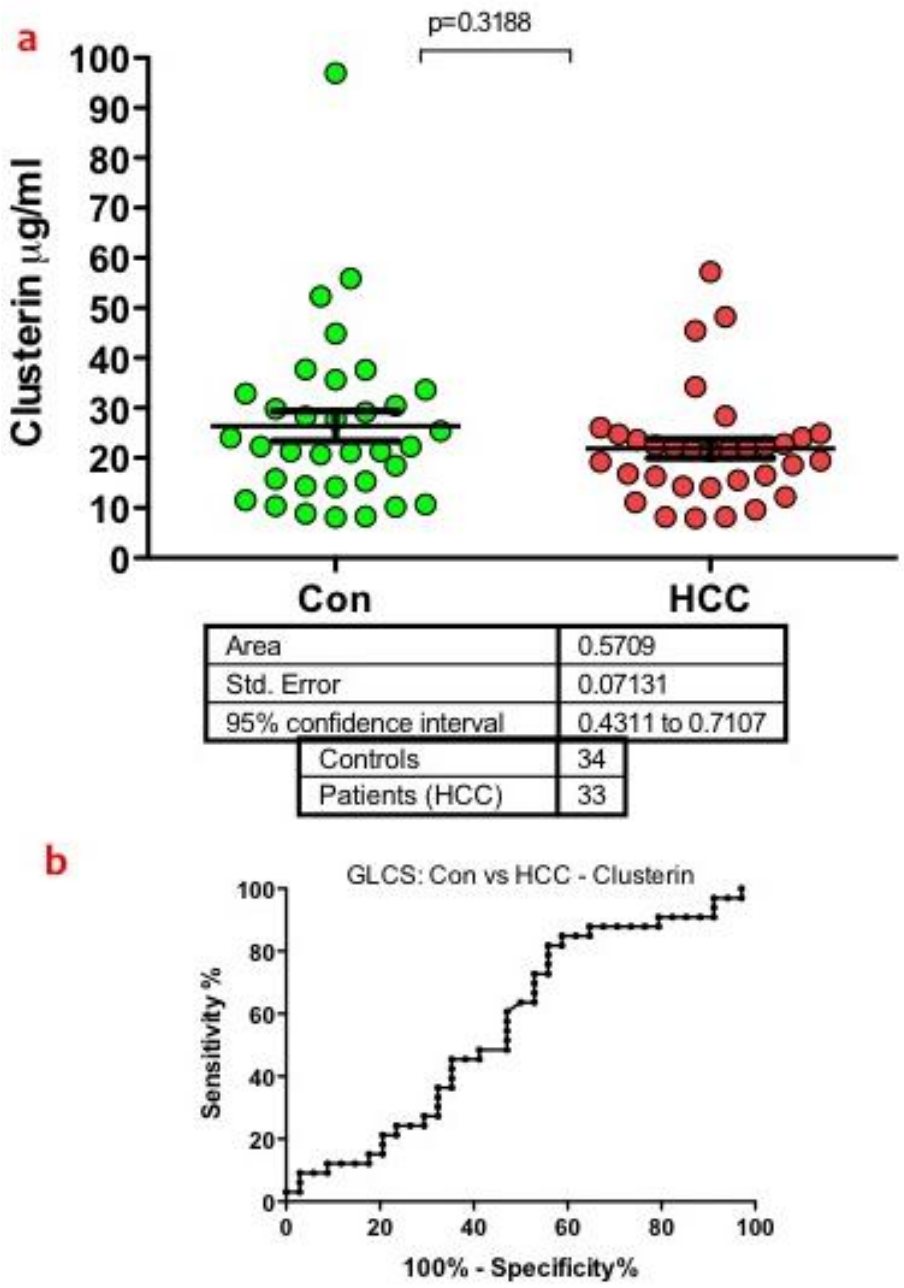
Zollner, B. et al. (2001) 20-fold increase in risk of lamivudine resistance in hepatitis B virus subtype adw. *Lancet*, **357**, 934-935.

Zweig, M.H. & Campbell, G. (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*, **39**, 561-577.

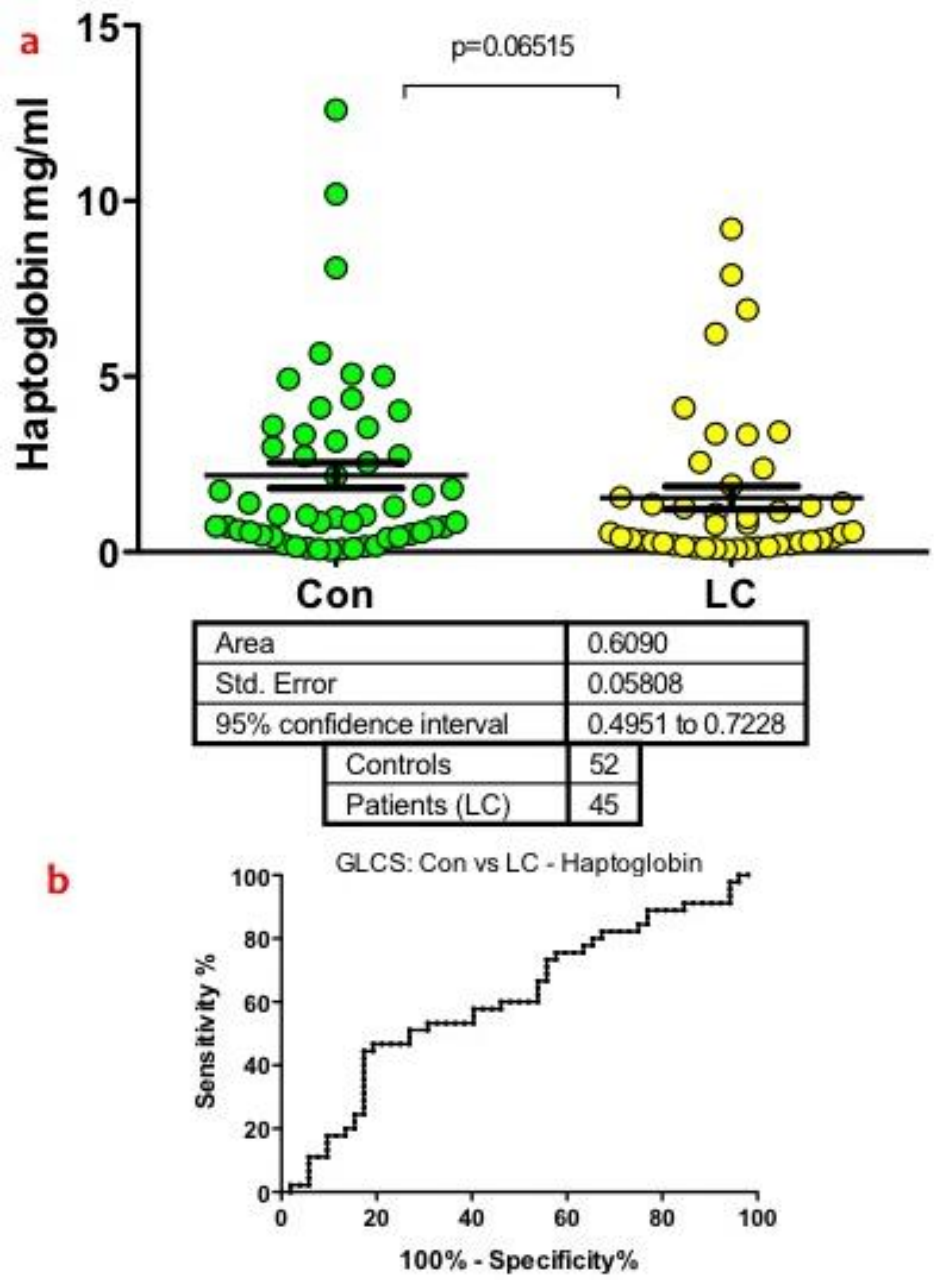
Appendix



Appendix Figure 1.0 (AF 1.0): Graph of total protein quantity in 9 plasma pools showing total proteins levels in the same order of magnitude.



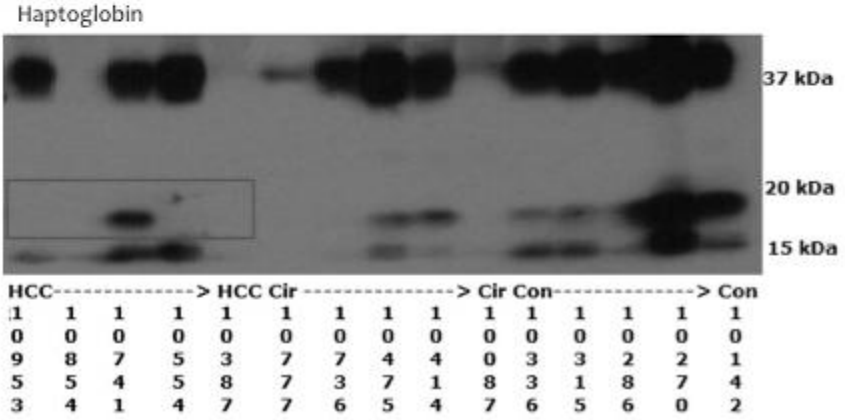
AF 1.1: (a) Dot plot showing levels of clusterin expression between Con and HCC subjects with associated AUC and (b) ROC curve.



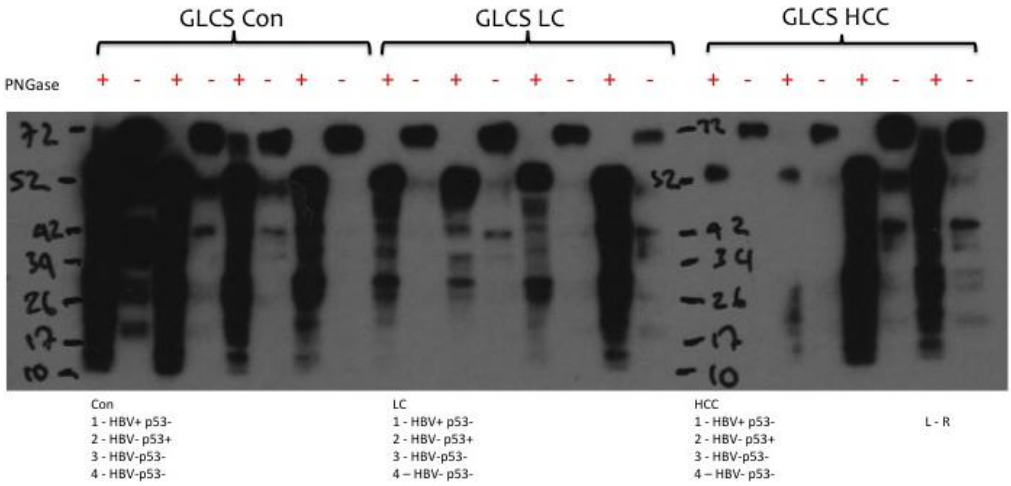
AF 1.2: (a) Dot plot showing levels of haptoglobin expression between Con and LC subjects with associated AUC and (b) ROC curve.

Control

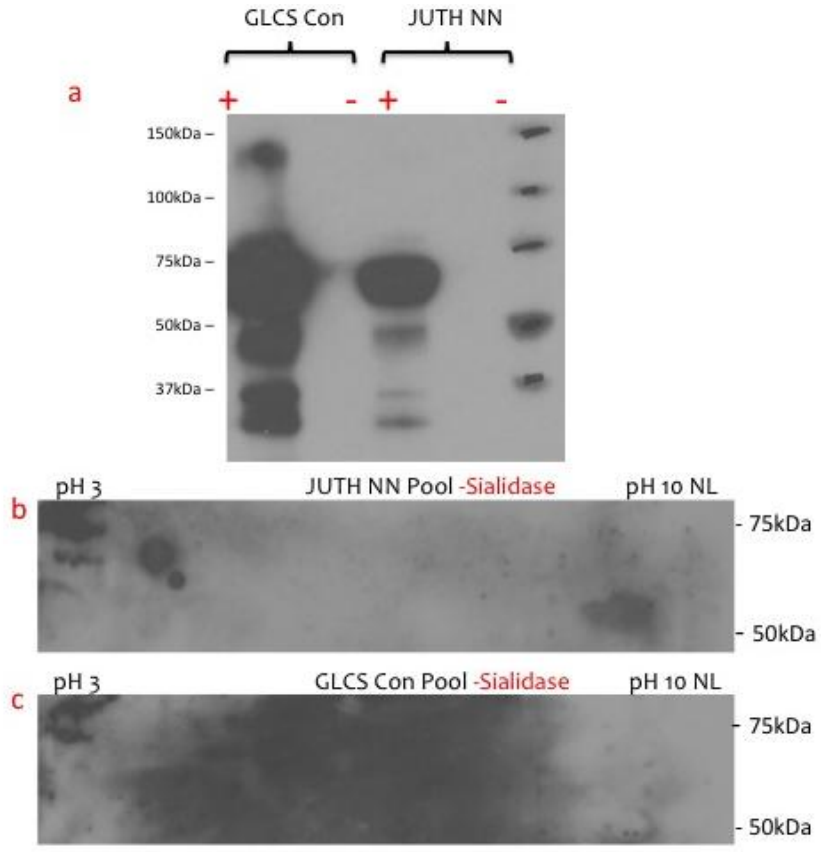
sampleid	AFP	HCVSTAT	HBV SAG	PATHDXB	PATHDXA	HBEAG
10142	4	Neg	Neg			
10270	4	Neg	Neg			
10286	4	Neg	Neg			
10315	4	Neg	Neg			
10336	4	Neg	Neg			
Cirrhotic						
sampleid	AFP	HCVSTAT	SAG	PATHDXB	PATHDXA	HBEAG
10087	32	Neg	Pos			positive
10414	66	Neg	Pos			positive
10475	3500	Neg	Pos	Hepatic Cirrhosis		negative
10736	4	Neg	Neg	Chronic hepatitis with fibrosis		
10777	4	Neg	Neg	chronic Lobular hepatitis		
HCC						
sampleid	AFP	HCVSTAT	HBV SAG	PATHDXB	PATHDXA	HBEAG
10387	388	Neg	Pos	HCC	HCC	negative
10554	215	indeterminate	Neg	HCC	HCC	
10741	5000	Neg	Pos	Hepatocellular carcinon		negative
10854	5000	Neg	Pos	Hepatocellular carcinoma		
10953	4	Neg	Neg	Hepatocarcinoma (poorly diff)		



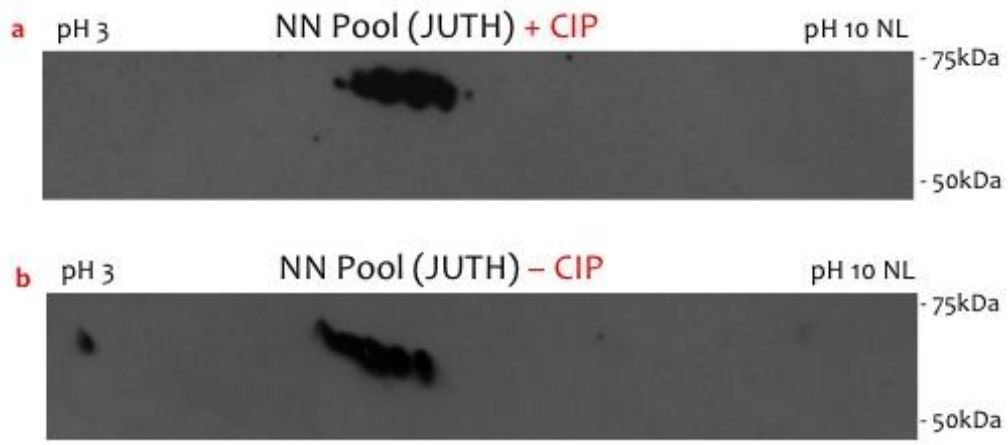
AF 1.3: Immunoblot for Haptoglobin on five subjects each with well-characterized control and liver disease profiles as summarized in the displayed table. The results show most disease related alterations to occur in the alpha chain region of the protein.



AF 1.4: Immunoblot for HPX showing consecutive treatments with and without PNGase enzyme in 4 CON; 4 LC & 4 HCC subjects. Subjects were specifically selected to represent mixed HBV and p53 mutation states.



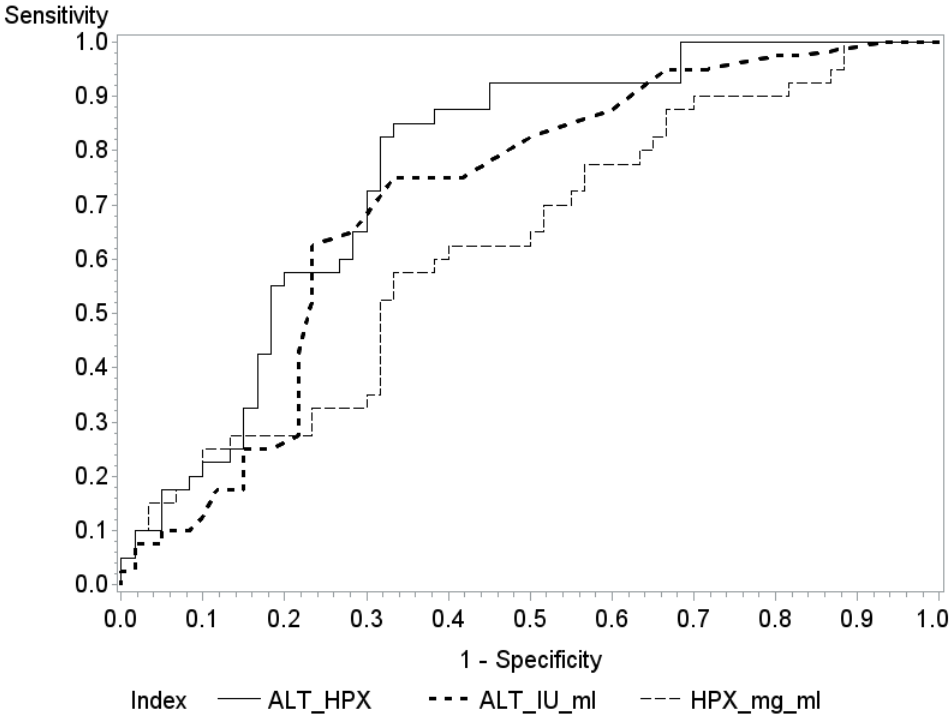
AF 1.5: (a) Immunoblot for HPX showing consecutive treatments with and without neuraminidase enzyme in the respective control pools from the GLCS and JUTH populations. This figure shows that sample loss in the neuraminidase negative treatments (b & c) occurred prior to the application of second dimension separation steps.



AF 1.6: 2DE separation of JUTH NN plasma pools treated with and without a Calf Intestinal Phosphatase (CIP) enzyme to see if it resulted in major changes in protein migration patterns on a pH 3-10 NL IPG strip.

Marker	AUC (LC vs. HCC)	SEM	95% CI
ALT ng/ml	0.7079	0.0524	0.6053 - 0.8105
HPX mg/ml	0.6250	0.0567	0.5138 - 0.7362
AFP_HPX	0.7850	0.0475	0.6720 - 0.8580
Comparison	**p-value of difference**		
ALT & HPX	0.3583		
ALT & ALT_HPX	0.2614		
HPX & ALT_HPX	0.0100		

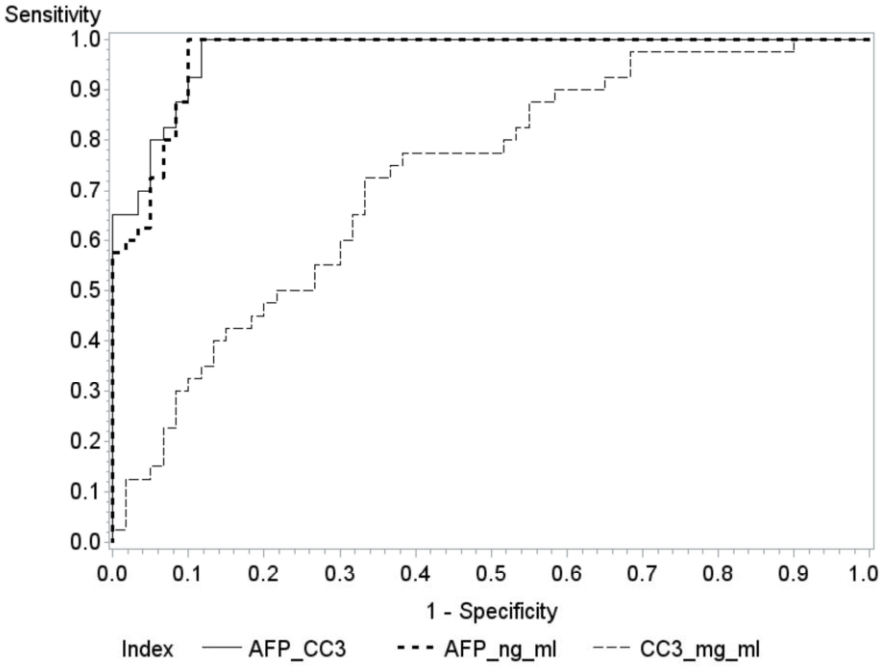
Appendix Table 1.0 (AT 1.0): Summary of statistical indices for HPX versus ALT and ALT_HPX in LC vs. HCC GLCS subjects.



AF 1.7: Combined ROC curves for HPX and ALT in LC vs. HCC subjects, highlighting which test combinations are superior in diagnostic ability.

Marker	AUC (LC vs. HCC)	SEM	95% CI
AFP ng/ml	0.9700	0.0140	0.9426 - 0.9974
CC3 mg/ml	0.7258	0.0507	0.6265 - 0.8252
AFP_CC3	0.9738	0.0122	0.9498 - 0.9977
Comparison	**p-value of difference**		
AFP & CC3	<0.0001		
AFP & AFP_CC3	0.3491		
CC3 & AFP_CC3	<0.0001		

AT 1.1: Summary of statistical indices for CC3 versus AFP and AFP_CC3 in Con vs. HCC GLCS subjects.



AF 1.8: Combined ROC curves for CC3 and ALT highlighting which test combinations are superior in diagnostic ability.