University of Texas at El Paso

ScholarWorks@UTEP

Open Access Theses & Dissertations

2022-12-01

Autoantibody Against Tumor-Associated Antigens as Diagnostic Biomarkers in Hispanic Patients with Hepatocellular Carcinoma

Yangcheng Ma
University of Texas at El Paso

Follow this and additional works at: https://scholarworks.utep.edu/open_etd

Part of the Biology Commons, and the Oncology Commons

Recommended Citation
Ma, Yangcheng, "Autoantibody Against Tumor-Associated Antigens as Diagnostic Biomarkers in Hispanic Patients with Hepatocellular Carcinoma" (2022). *Open Access Theses & Dissertations*. 3702.
https://scholarworks.utep.edu/open_etd/3702

This is brought to you for free and open access by ScholarWorks@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.

AUTOANTIBODY AGAINST TUMOR-ASSOCIATED ANTIGENS AS DIAGNOSTIC BIOMARKERS IN HISPANIC PATIENTS WITH HEPATOCELLULAR CARCINOMA

YANGCHENG MA

Doctoral Program in Biosciences

APPROVED:

____
Jianying Zhang, M.D., Ph.D., Chair

____
Renato J. Aguilera, Ph.D., Co-chair

____
Giulio Francia, Ph.D.

____
Taslim A. Al-Hilal, Ph.D.

____
Jianjun Sun, Ph.D. (Advocate)

____
Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

Copyright ©

by

Yangcheng Ma

2022

Dedication

This work is dedicated to my family.

It is their endless love for me that made me reach my goal.

AUTOANTIBODY AGAINST TUMOR-ASSOCIATED ANTIGENS AS DIAGNOSTIC BIOMARKERS IN HISPANIC PATIENTS WITH HEPATOCELLULAR CARCINOMA

by

YANGCHENG MA, M.D.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Biological Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

December 2022

Acknowledgments

I would like first to send my sincere and profound thanks to my Ph.D. mentor, Dr. Jianying Zhang, who has given me enormous trust and support throughout the project. All my accomplishments would not have been completed without his encouragement and strict training. I really appreciate the help from my committee members Dr. Renato Aguilera, Dr. Taslim A. Al-Hilal, Dr. Giulio Francia, and Dr. Jianjun Sun. They provided me with excellent advisement and support for my project.

I would also like to thank all lab members and collaborators (Dr. Xiao Wang, Dr. Xiaojun Zhang, Dr. Mengtao Xing, Ms. Cuipeng Qiu, Ms. Nohemi Lopez-Valdez, Ms. Zenaida Fuentes, and Dr. Bofei Wang) in Dr. Zhang's lab for their help and support. The project cannot be finished without their help. I want to give special appreciation to my parents and my best friend, Rui Dong. They have been there for me through the highs and lows of my doctoral career.

Finally, I would like to thank the BBRC grant for providing me with financial support. I want to thank all the staff of the BBRC core facilities, who have given me technical support.

Abstract

Liver cancer is one of the most prevalent malignancies, in which hepatocellular carcinoma (HCC) takes up more than 80% of liver cancer cases. Early diagnosis and treatment are practical approaches to impede HCC progression. Identifying and classifying autoantibodies against tumor-associated antigens (TAAs) have long been interested in cancer diagnosis, especially for HCC. Circulating anti-TAA autoantibodies amplify the antigen expression and are more stable in blood, guaranteeing anti-TAA autoantibodies as valid diagnostic biomarkers in clinical application. Nonetheless, very few studies had focused on the Hispanic HCC group, which might retain distinct etiological risk factors compared to other ethnicities. The high incidence rate of metabolic disorders correlated liver diseases and increased consumption of alcohol in the Hispanic population have favored the development of HCC.

In this study, we identified 9 potential TAA targets using the immunoproteomic approach serological proteome analysis (SERPA) and 10 potential TAA targets by the differentially expressed driver genes from Hispanic patients. 9 TAA targets were detected by SERPA, including HSPA5, HSP60, TPI, P4HB, ACTG1, HSP70, ENO1, TPM3, CALR. Using the IntoGen driver gene database together with Oncomine and TCGA cancer genome databases, 10 driver genes were investigated as potential targets with significantly different mRNA expression between HCC and normal liver tissues, which are CDKN2A, SETDB1, DHX9, SMARCA4, GNAS, NF2, DNMT3A, NRAS, GMPS, and MAPK1. Among all the identified targets, anti- DNMT3A, p16, Hsp60, and HSPA5 autoantibodies were defined as diagnostic biomarkers with a significantly higher frequency in the Hispanic HCC group compared to the NHS group. The area under the receiver operating characteristic curve (AUC) value of the single TAA varies from 0.7505 to 0.8885, showing the great diagnostic value of these biomarkers. After combining all the TAAs, the

sensitivity of the anti-TAA autoantibody panel increased to 75%, while the highest sensitivity of the single TAA was 45.8%. Next, to explore the exclusiveness of the 4 autoantibodies for Hispanic HCC diagnosis, we established a multi-ethnicity study with a separate Asian sera cohort. It is shown that anti-p16 and anti-HSPA5 autoantibody titers or frequencies are significantly higher than the normal sera group, suggesting DNMT3A and HSP60 are exclusive diagnostic biomarkers for Hispanic HCC.

The present research shows the 4 protein targets TAAs including DNMT3A, p16, HSPA5, and HSP60 as diagnostic biomarkers that elicit remarkable autoimmune responses in sera from Hispanic HCC patients. Also, DNMT3A and HSP60 could be potential biomarkers exclusive to Hispanic HCC diagnosis, while HSPA5 and HSP60 might be effective biomarkers in both Hispanic and Asian HCC populations.

Keywords: Hepatocellular carcinoma (HCC); Tumor-associated antigen (TAA); Serological proteome analysis (SERPA); Hispanic; Immunodiagnosis.

Table of Contents

Table of Contents viii

List of Tables xi

List of Figures xii

Chapter 1: Introduction 1

1.1 Hepatocellular carcinoma 1

1.2. Hispanic HCC 5

1.3. Tumor-associated antigen (TAA) and anti-TAA autoantibodies 7

1.4. TAA and anti-TAA autoantibodies in HCC 9

1.5. Driver gene and TAA 10

1.6. Identification of TAAs 11

1.7. TAA and anti-TAA autoantibody panel 13

Chapter 2: Hypothesis & Specific aims 14

Hypothesis 14

Specific aims 15

Chapter 3: Material and methods 16

Chapter 4: 9 potential TAA targets were identified by SERPA in Hispanic HCC 23

Overview 23

Rationale 24

Experimental design 24

Results 26

Chapter 5: 10 potential TAA targets were identified from HCC driver genes 37

Overview 37

Rationale 37

Experimental design 38

Results 39

Chapter 6: Autoantibody prevalence of the potential TAA targets and establishment of anti-TAA autoantibody panel 41

Overview 41

Rationale 41

Experimental design 42

Results 43

Chapter 7: Validation of 4 anti-TAA autoantibodies in a separate Asian sera cohort 51

Overview 51

Rationale 51

Experimental design 52

Results 52

Chapter 8: Discussion 55

References 60

Abbreviation 71

Curriculum Vita 73

List of Tables

Table 1. The properties of autoantibodies in the positive sera ········ 29

Table 2. Identification of protein targets by mass spectrometry from SERPA ········ 36

Table 3. Candidate TAA targets identified from HCC driver genes ········ 39

Table 4. Frequency of the 19 anti-TAA autoantibodies in Hispanic sera cohort ········ 47

Table 5. Sequential addition of antigens to the panel of 4 TAAs ········ 50

Table 6. Frequency of the 4 anti-TAA autoantibodies in Asian sera cohort ········ 53

List of Figures

Figure 1. Positive sera screening by western blotting ········· 26

Figure 2. Autoantibody pattern and intensity of positive sera and negative sera samples by indirect immunofluorescence (IIF) ········· 27

Figure 3. Identification of differentially expressed protein spots by SERPA in Hispanic HCC sera and NHS sera ········· 31

Figure 4. Enrichment analysis of potential TAA target identified by SERPA ········· 34

Figure 5. Molecular weights and protein concentrations of 19 recombinant proteins on SDS-PAGE gel ········· 44

Figure 6. Autoantibody titer distribution of 19 potential TAA targets ········· 45

Figure 7. Autoantibody titers of the 4 TAAs in the Hispanic HCC and NHS groups ········· 46

Figure 8. ROC analysis of anti-TAA autoantibodies in sera from patients with Hispanic HCC and normal healthy control individuals ········· 49

Figure 9. Autoantibody titers of the 4 TAAs in the Asian sera cohort ········· 54

Chapter 1: Introduction

1.1 Hepatocellular carcinoma

Liver cancer is the 6th most common and 4th leading deadly cancer worldwide. It is estimated that by 2025, more than 1 million individuals will be affected by liver cancer pre year [1]. To be the most common form of liver cancer, hepatocellular carcinoma, takes up more than 80% of the overall liver cancer cases [2]. HCC incidence and mortality trends vary globally, partially due to timing and level of exposure to environmental and infectious factors, the ability to access healthcare, and the early-stage HCC diagnosis and screening. HCC tends to emerge in the later stages of life in Japan, North America, and European countries. In contrast, in some Asian and African countries, HCC is commonly diagnosed between the ages 30 to 60, much lower than in developed areas [3]. Overall survival rates also demonstrate geographical variances, where Japan and Taiwan sustain the highest median survival, partly due to their outstanding surveillance programs and investment in public health centers [4].

When focusing on the United States, research shows that epidemiological trends of HCC have changed over the decades. From 2000 to 2010, the overall age-adjusted incidence and mortality rates have increased [5]. Although the 35-49 years old population shows decreased incidence and mortality rates, the elderly people whose age is greater than 50 years suffer more from HCC, rendering 9.6% per year incidence rate and 5.6% per year mortality rate. The mortality rate of HCC also varies among states. The southern area, including Louisiana, Mississippi, and Texas, suffer the most from HCC, which has an average of 6.2 age-adjusted death rates per 100,000. Those previous studies have pointed out the importance of regional discrimination in further investigations of HCC in the United States.

Every cancer comes with stress, the same is true for HCC. Any etiological agent leading to chronic liver injury, inflammation, fibrosis, and liver cirrhosis will be considered risk factors for HCC [6]. The major risk factors are viral infection, alcohol consumption, and metabolic disorder leaded non-alcoholic fatty liver disease (NAFLD). Other sources of stress can broadly be defined as minor risk factors, such as aflatoxin exposure, alpha 1-antitrypsin deficiency, autoimmune hepatitis, hereditary hemochromatosis, etc. [7].

Viral infection is responsible for most liver cirrhosis and HCC cases. Research shows that hepatitis B virus (HBV) infection accounts for about 60% of HCC cases in Asia and Africa and 20% of Western patients. In contrast, the hepatitis C virus (HCV) is the most common etiological factor of HCC in North America, Europe, and Japan [8]. HBV is a double-stranded DNA virus that could encode the surface antigen HBsAg and integrate its genetic information into the host genome, thus inducing insertional mutagenesis and activating the tumorigenesis procedure [9]. HCV, on the other hand, is a single-stranded RNA virus that produces structural and nonstructural viral proteins [10]. Hepatitis viruses themselves are not the major attributions of HCC. Evidence shows that hepatitis B virus carriers suffer an incidence of HCC of 0.4% to 0.6% per year, while hepatitis B cirrhosis or hepatitis C cirrhosis retain 3%-8% and 3%-5% HCC incidence rates separately [6].

Alcohol consumption causes alcoholic liver diseases, which induce liver cirrhosis and HCC. Alcohol-induced HCC takes up about 30% of all HCC incidences [11]. Compared to non-drinkers, the risk of HCC increases by 46% for consuming 50g of ethanol per day and 66% for consuming 100g of ethanol per day in drinkers [12]. Alcohol consumption dramatically promotes the occurrence of alcohol-associated fatty liver, and through the vulnerable steatotic liver, it will develop secondary liver inflammation and further liver cirrhosis.

Metabolic disorder-induced liver diseases are now gaining intense traction. Attributing to the increasing prevalence of diabetes and obesity, non-alcoholic steatohepatitis (NASH) has become one of the most influential factors to cirrhotic HCC [13]. A previous study has indicated that the proportion of total HCC cases by NASH almost doubled in 2014-2016 compared to 2008-2010 [14]. Stresses such as dyslipidemia, high glucose diet, or sedentary lifestyles will lead to liver steatosis, which will sequentially cause liver inflammation. With hepatocyte damage, DNA damage, ER stress, oxidative stress, steatosis and lobular inflammation of the liver will be transmitted into NASH and thus result in further fibrosis and HCC [15]. NASH also constitutes the most significant etiological proportion of non-cirrhotic HCC in the United States [16]. Compared with other risk factors such as alcohol, HBV, and HCV, NASH-related HCC possess the highest proportion of noncirrhotic HCC and worse overall survival than alcohol-induced HCC [14].

The diagnosis of HCC is an integrated method that includes clinical and histopathological tests. Generally, the diagnosis algorithm m of HCC has 2 pathways, one is the size-based pathway and the other is the non-size-based pathway [17]. In the size-based pathway, repeated ultrasound (US) test or serological alpha-fetoprotein (AFP) test will be executed if the tumor size is less than 1 cm, while dynamic imaging examinations are carried out if the tumor size is larger than 1 cm. In both size-based and non-size-based pathways, imaging tests and biopsies are the critical factors in the final diagnosis of HCC.

US is the most common and financial-efficient test for HCC diagnosis. Different US techniques generate various clinical applications. Routinely executed gray-scale US can detect early-stage liver lesions with high sensitivity, while color Doppler flow US imaging can provide visualized blood supply within the liver mass. The noninvasiveness of the US makes it a resultful

technique in the early screening of potential HCC patients. Comprehensive US tests allow preoperative diagnosis, intraoperative localization, and postoperative evaluation of HCC [18]. Apart from the US system, the imaging system, including computerized tomography (CT) and magnetic resonance imaging (MRI) also plays a vital role in HCC diagnosis. Patients with an abnormal liver nodule under the US or a high serum AFP level (usually > 20 ng/ml) are considered a high-risk population and in need of dynamic imaging evaluation [6]. Although research shows that the sensitivity of MRI for HCC diagnosis could exceed that of CT, it is still believed that both imaging tests are highly recommended for the initial diagnosis of HCC because CT can generate intrahepatic vascular distribution while MRI has excellent tissue resolution [6, 19]. It should be noted that the accuracy of the imaging test of HCC depends on the risk of the population. For those with pre-HCC liver cirrhosis, the capability of imaging tests is guaranteed. But for those with atypical features such as benign liver nodules, the imaging test itself is not fully convincing.

HCC and other hepatic tumors such as intrahepatic cholangiocarcinoma are associated with tumors in vein, resulting in difficulty in the differential diagnosis. When the HCC is atypical and losing its imaging hallmarks, the histopathological test or the biopsy test is needed to distinguish the features of HCC [20]. Nonetheless, the biopsy is always bound with risks. The major risks of liver biopsy are bleeding and needle tract implantation [21]. In addition, the sensitivity of biopsy goes down to about 70% or even lower in tumor size less than 2 cm due to the potential for missing lesions and losing the ability to differentiate HCC from dysplastic nodules [2]. In general, the clinical diagnosis of HCC requires applying multiple techniques, including laboratory and non-laboratory tests.

1.2. Hispanic HCC

The impact of HCC varies among ethnicities, but the Hispanic population is affected more than any other population in the United States. Previous research of 205,369 cancer deaths from 2008-2012 was analyzed, suggesting that the Hispanic ethnicity has a higher liver cancer risk than the non-Hispanic white group [22]. The lowest overall survival rate is observed in the Hispanic group compared to the white and the African American groups [23]. Evidence shows that the age-adjusted incidence rates of HCC in the Hispanic population have surpassed those of Asians and American African HCC patients, suffering the lowest 1- and 3-year overall survival rates regardless of age, sex, time of diagnosis, or other factors[24, 25]. The incidence and mortality rate of Hispanic HCC also varies geographically. It is shown that the southern part of the United States, such as Texas and Louisiana states suffer higher incidence and mortality rates than other states [5]. From 1995 to 2010, Hispanics in south Texas had the highest incidence rate of HCC (12.1/100,000), even higher than Hispanics in the rest of Texas (10.9/100,000) and the United States average (8.4/100,000) [26]. Different Hispanic groups also show additional resistance against HCC. Some researchers investigated that Cubans and Puerto Ricans might have higher mortality rates than other Hispanic ethnicities, and the US-born Hispanics are at more significant risk than those foreign-born Hispanics [22, 24, 27].

Whereas the HVC infection is the dominant etiological factor for HCC in the United States, the Hispanic population preserves another circumstance. It has been found that Mexican Americans retain an intermediate HCV prevalence (1.9%) and a low HCV prevalence among South American Hispanics (0.4%) [28]. However, within 149,407 HCV-infected U.S. veterans, only 6% were Hispanics, but Hispanics have a higher annual incidence rate of cirrhosis and HCC compared to non-Hispanic white, suggesting that other etiological risk factors are ascribed to

Hispanic HCC [29]. A considerable amount of literature has shown that metabolic risk factors and infectious liver disease may be the leading cause of HCC incidence in Western society, especially in the Hispanic population. It was suggested that the prevalence of NAFLD is highest within the Hispanic-white population, and having no less than 2 metabolic risk factors will be significantly associated with an increased risk of HCC and hepatic decompensation [30]. Compared to White and African American groups, the Hispanic population suffers a higher incidence of metabolic-hepatic disorders such as non-alcoholic steatohepatitis (NASH) and non-alcoholic fatty liver disease (NAFLD) [23]. Studies have shown that a yearly cumulative incidence of HCC from NASH cirrhosis could be approximately 2.6%, and a prevalence of 58.3% of NAFLD and 19.4% of NASH has been found in a retrospective study regarding the adult Hispanic population [31, 32]. The metabolic disorders, unique biologic gene-environment interactions, and more aggressive cirrhosis progression might contribute to higher HCC incidence and mortality rate and shorter overall survival (OS) for Hispanic HCC patients[23, 33]. One possible reason for this association might be that insulin resistance promotes hepatocyte injury, proliferation, and collagen production, creating a pro-inflammatory status in the liver[34]. Apart from insulin resistance, other factors such as oxidative stress and malfunction of Toll-like receptor proteins within hepatocytes could also be involved in the transition of liver steatosis to NAFLD [35]. Those could be proportionally associated with lifestyle. Research shows that the prevalence of obesity strongly affects the Hispanic population [36]. South Texan Hispanics suffer over 83.1% of the population being either overweight (32.4%) or obese (50.7%) [37]. Obesity and diabetes mellitus could be the major factors contributing to the development of Hispanic HCC.

Even though African Americans suffer higher rates of obesity and hypertension, they appear to have lower NAFLD risk than the Hispanic population [18], suggesting that metabolic

disorders and other risk factors might also be crucial to Hispanic HCC prevalence. Alcohol consumption, genetic variance, and access to health care play a vital role in the prevalence and survival of Hispanic HCC. According to previous study, Hispanics had the most prominent orientation toward heavy drinking compared to other ethnicities [38]. This might be partially ascribed to the degree of acculturation and awareness of ethnic stigma [39]. Likewise, genetic variance can be another contributing factor to Hispanic HCC. An epidemiological study has investigated that 70% of children with NAFLD were of Hispanic origin, indicating the molecular foundation of the high prevalence of NAFLD in the Hispanic population. [40]. Previous research demonstrated that the rs738409[G] allele in the patatin-like phospholipase domain-containing 3 (PNPLA3) gene is more commonly seen among Hispanics, which is highly correlated with increased hepatic fat levels [41]. TP53R249S mutation is also found in Hispanic HCC in South Texas, leading to younger HCC occurrence and shorter overall survival [42]. Lack of early healthcare intervention and decreased invasive treatment utilization were noted among Hispanic HCC patients, leading to higher rates of HCC mortality among Hispanics compared to non-Hispanic whites and Asians [43].

1.3. Tumor-associated antigen (TAA) and anti-TAA autoantibodies

Tumor-associated antigens (TAAs) refer to the proteins by which the immunosurveillance system recognizes the aberrant changes in the tumor microenvironment. The antigenic proteins produced by tumorigenesis trigger immune reactions and thus generate autoantibodies against TAAs. The concept of TAA was first brought out in the 1960s by Baldwin, and hundreds of novel TAAs towards cancer were investigated from then on [44]. TAAs can be sorted into several categories: viral antigens, antigens caused by genetic mutation or dislocation, and ectopically

expressed antigens [45]. Remodeling of tumor cells caused by chronic inflammation can lead to dysregulated protein expression originating from the aberrant tumor microenvironment and results in the secretion of dysregulated proteins. There are 4 phases of how tumor cells interact with immune systems: elimination, escape, equilibrium, and immune subversion [46, 47]. The immune system first recognizes the abnormal cells and pre-cancer cells. Antigens comprising pathogenic molecules are presented to the host's immune system and trigger the immune response by inducing natural killer group 2D (NKG2D) ligands on cancer cells. This process is called tumor elimination [48, 49]. After the elimination phase, the tumor cells with the impaired immunogenic property might be able to get away from immune attacks and reach an equilibrium state with the host immune system. Eventually, the tumor cells start to over-compete host immune systems and suppress immune responses. Dendritic cell differentiation and maturation are suppressed by a high level of tumor-originated vascular endothelial growth factor (VEGF). VEGF also stimulates immature dendritic cells and recruits them from bone marrow to tumor, resulting in immune subversion by malfunctioning T-cells [50]. TAAs need to be released into the extracellular environment to access the lymph node and sequentially produce autoantibody. This process can be achieved through classical secretion pathways, active release of extracellular vesicles, and ectodomain shedding of transmembrane proteins [51]. In addition, tumor cell necrosis and autophagy could spill intracellular contents into the circulating blood, exposing the antigens to the immune system [52] . Once TAAs reached the lymph node, they will be captured by antigen-presenting cells (APCs), and though major histocompatibility complex II activates helper T cells. After releasing cytokines, helper T cells will activate mature B cells and thus differentiate into plasma B cells and further produce autoantibodies [45].

Autoantibody against TAA has its natural advantages of being a cancer biomarker. Compared to short-lived tumor antigens in blood, the antibody molecule is more stable, and antigen-antibody response is more enduring in body fluid. Also, the nature of antibody amplification response makes a small amount of antigen able to trigger a more significant amount of immune response, which is reflected in relative antibody concentrations [53]. With the application of time- and cost-efficient analysis tools such as enzyme-linked immunosorbent assay (ELISA), autoantibody against TAAs can be easily detected in patients' blood samples. It can be easily assembled to establish a multiple TAAs panel for better diagnostic performance. These days the profiling of TAAs has been more accessible with the development of immunoproteomics techniques than before. The general workflow of profiling TAAs includes an exhibition of candidate antigens, immune detection, and identification of target antigens [54].

1.4. TAA and anti-TAA autoantibodies in HCC

The liver is a central department of metabolism, blood regulation, and immune surveillance. The dysfunction of liver cells will lead to an impaired immune microenvironment, which will eventually cause HCC. The unique architecture of the liver and its rich blood supply system makes it easy to deliver the degraded immune signal. Liver sinusoids are formed by a monolayer of liver sinusoid endothelial cells that lack a basement membrane and let the blood flow directly reach the hepatocytes, passing immune cells to hepatocytes through sinusoids [55]. Also, risk factors such as alcohol or metabolic disorders will trigger inflammatory cell death and cause liver injury, which will lead to an influx of immune cells and lead to liver cirrhosis and HCC [56]. All the anatomical and microbial features of the liver have provided an appropriate opportunity for the immune diagnosis and treatment of HCC.

Tan and Zhang have suggested that there will be a sudden rise of autoantibody titer within a chronic hepatitis patient before the diagnosis of HCC, which is earlier than the appearance of AFP [52]. Based on that consensus, numerous anti-TAA antibodies have been investigated as diagnostic biomarkers for HCC patients, such as IMP (insulin-like growth factor 2 mRNA binding protein) family, CIP2A/p90, MDM2, RalA, c-myc, etc. [57-60]. In 1999, we first investigated IMP2/p62 as a promising biomarker in the early diagnosis of HCC [61]. In the subsequent molecular studies, we found that the expression of IMP2/p62 mRNA and protein are elevated in the HCC tissue compared to the adjacent normal liver tissue. In addition, the overexpression of IMP2/p62 is regulated by the Wnt/β-Catenin signaling pathway, which induces enhanced HCC cell mobility and genomic instability [62]. Our lab has investigated the heat shock protein family members as HCC diagnostic biomarkers, such as HSPA5, HSP60, and HSP70 [63, 64]. The combination of different TAAs has proved to be a more powerful tool in diagnosing HCC than single TAAs. Dai et al. suggested that compared to the single biomarker CAPERα, the sensitivity of a 14-TAA panel has risen from 21.1% to 69.7% [58]. Zhang et al. also investigated that the positive antibody reaction of an 8-TAA mini-array reaches a frequency of 59.8%, which is significantly higher compared to the single TAA, which varies from 9.9% to 21.8% [65].

1.5. Driver gene and TAA

Over the decades, the inflammatory turbulence of cancer has attracted massive attention, and autoantibodies are produced during the tumorigenesis progression like autoimmune disease. It is noted that 30% of all cancer patients have circulating anti-nuclear antibodies (ANAs) in their sera [66]. Zhang et al. investigated that there will be an influx of ANA level before the clinical diagnosis of hepatic cancer [67]. Though the exact cause of cancerous autoantibody production

remains to be elucidated, there are 4 possible theories that autoantibody is generated in cancer, which are protein expression level changes, tolerance and inflammation defection, protein structure alteration, and cellular death mechanisms [68]. The previous study has shown that most TAAs are overexpressed or mutated proteins, of which 42% are cytoplasmic proteins, 26.1% are nucleus proteins, 21.4% are membrane-bound proteins, and 10.3% are extracellular proteins [69]. Driver genes refer to the mutational genes that are strongly linked to tumorigenesis and cancer progression [70]. Previous researches have suggested that new antigenic epitope of TAAs could be produced via somatic mutations, therefore being detected by the immune system and bound by autoantibody in both mutated and non-mutated ways [45]. There is research suggests that the expression of neoantigens in tumor cells by genetic instability will reversely trigger the immune response against tumor tissue [71]. Interestingly, the autoantibody which is elicited as a result of modified proteins will bind to both modified and unmodified proteins, possible due to the epitope spreading, showing its prominent stability and sensitivity in serological analysis [72]. It is worth noting that missense point mutations are highly correlated to the production of antibodies but not other mutations such as stop, splice, and frameshift [69].

1.6. Identification of TAAs

A dysfunctional immunosurveillance system usually appears in cancer patients, and humoral immunity will recognize and respond to changes in tumor cells. Since the first research by Baldwin in the 1960s [44], the TAAs and anti-TAA antibodies have been drastically studied, and many have been applied as diagnostic or prognostic biomarkers. Not only the anti-TAA antibodies are easily accessible in the circulating blood samples, but the amplification of antigen-antibody response makes it easy to trigger a more significant immune response with relatively low

antigen quantity in the early diagnosis process. Along with the development of mass spectrometry technology, many antigens have been investigated during the decade, and various TAA profiling techniques have drawn interest by their edges.

Serological analysis of expression cDNA libraries (SEREX) refers to the technique that identifies TAAs by establishing a cDNA expression library obtained from tumor tissues and incubating it with patients' sera samples [73]. Ideally, the proteome of the tumor cells will be exposed, and a comprehensive autoantibody profile will be investigated through serological analysis. However, the coverage of antigens is limited due to the restricted tumor tissue source, and post-translational modifications (PTMs) are not shown in SEREX results. Another technique named multiple affinity protein profiling (MAPPing) involves a combination of 2D immunoaffinity chromatography with mass spectrometry [74]. By depleting antigens recognized by autoantibodies from healthy controls, the antigen candidates that bind to antibodies in cancer subjects will be purified and subsequently identified by mass spectrometry analysis [75]. Although the 3D structure of the proteins is retained, low-affinity autoantibodies are hard to find considering the holding capacity of the affinity column. Protein microarrays have been widely utilized in the TAA discovery field. The microarray will be incubated with specific sera and screen candidate antigens with a time-efficient pattern by assembling a series of recombinant proteins on a platform. Yet such an approach is financially consuming, and some recombinant proteins might not be able to fully exhibit PTMs of natural proteins [76].

In the present study, we are trying to excavate novel TAAs for Hispanic HCC patients by serological proteome analysis (SERPA). Unlike other methodologies, SERPA provides a combination of two-dimensional protein electrophoresis, western blotting analysis, and mass spectrometry analysis [77]. Proteins from tumor cell lysates are separated by isoelectric point (PI)

and molecular weight, which will be subsequently transferred onto membranes and probed with sera from cancer patients or healthy controls. Compared to SEREX, SERPA maintains the post-translational modification of the proteins, which might influence the antigenicity. Also, SERPA provides the reactivity of specific sera with various antigens between cases and controls [78]. As a time- and financially-efficient technique, the SERPA will provide new insight into autoantibody discovery for the present study.

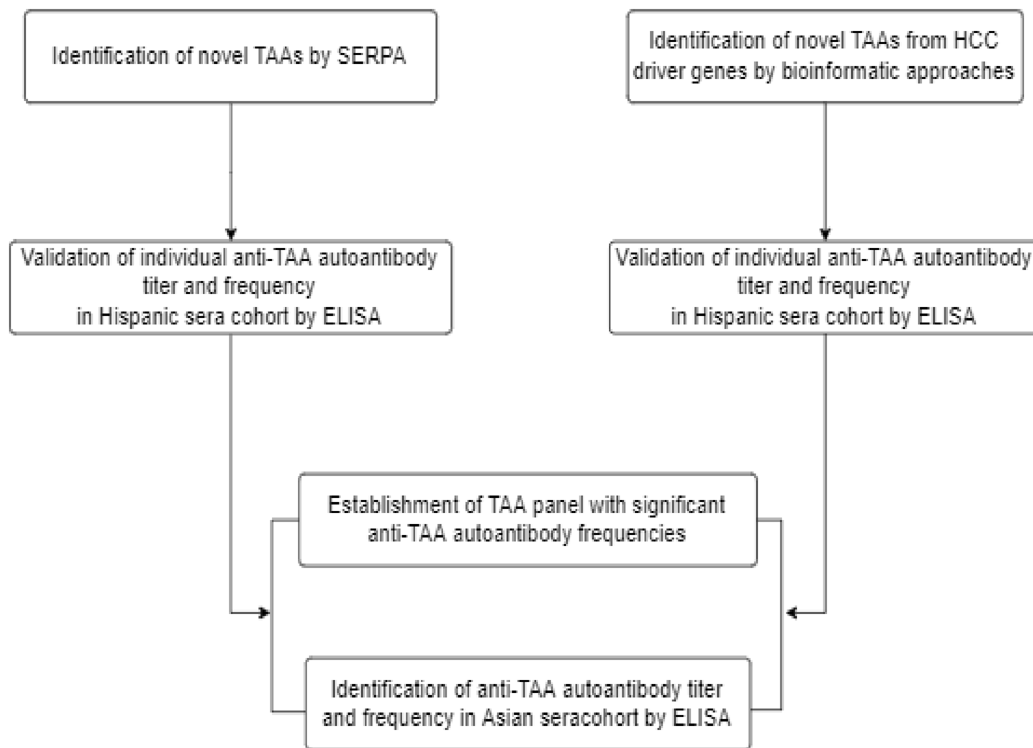
1.7. TAA and anti-TAA autoantibody panel

Compared to the features of rapid degradation and clearance in the circulating serum in TAAs, the presence of the anti-TAA autoantibodies are more persistent and stable [79]. Meanwhile, the wide prevalence of serological antibody identification and verification techniques guarantees the robustness of autoantibody biomarkers in diagnosing cancer. Sometimes, however, individual anti-TAA autoantibodies cannot show enough diagnostic value due to their low frequency, sensitivity, and specificity. One way to overcome this is to assemble multiple TAAs on a single platform complementarily, thus achieving an improved diagnostic value with a higher sensitivity rate [65]. Previous research from our lab shows that compared to the antibody frequency of individual TAAs, which varies from 6.6% to 21.1%, the sensitivity of the TAA panel will gradually ascend along with a stepwise increase of TAAs, and eventually reaches a sensitivity and specificity of 69.7% and 83.0%, respectively [58]. Additionally, individual anti-TAA antibodies might not be exclusive to specific cancers, such as p53 is proven to be overexpressed in multiple malignancies. A combination of different TAAs will reflect the distinct feature of autoantibody distribution in specific types of cancer, which can be beneficial for clinical precision diagnosis.

Chapter 2: Hypothesis & Specific aims

Hypothesis

Using TAAs as biomarkers have long been verified to be a powerful tool in HCC with high diagnostic and prognostic value. Yet very few studies focused on specific Hispanic populations, which comprise a significant proportion of worldwide liver cancer prevalence. Growing case numbers and particular metabolic dysfunction make it worthwhile to unearth more precise and accurate biomarkers in Hispanic HCC early diagnosis. In the present project, we aim to investigate novel diagnostic individual TAA biomarkers and TAA panels exclusively from sera of Hispanic HCC. The diagnostic value of potential protein hits from 2D gel electrophoresis encoded by driver genes will be examined. A TAA panel model with high sensitivity and specificity will be established. We hypothesize that novel TAA targets can be found in sera of Hispanic HCC groups and will contribute to a promising and powerful TAA panel in enhancing diagnostic value. The flow chart of the present study is listed below:



Specific aims

Specific aim 1: To identify potential TAA targets in Hispanic HCC by SERPA

Specific aim 2: To identify potential TAA targets from HCC driver genes by bioinformatic approaches

Specific aim 3: To validate anti-TAA autoantibodies from potential targets by ELISA and to establish an effective TAA panel in Hispanic HCC diagnosis

Specific aim 4: To validate the anti-TAA autoantibodies in a separate Asian sera cohort

Chapter 3: Material and methods

Serum samples

The serum samples were obtained from the serum bank of the Cancer Autoimmunity Research Laboratory at the University of Texas at El Paso. The Hispanic sera cohort includes 24 samples with hepatocellular carcinoma (HCC), 20 samples with liver cirrhosis (LC), 26 samples with chronic hepatitis (CH), and 40 normal healthy sera (NHS) samples. Asian sera cohort includes 53 HCC samples, 20 LC samples, and 44 CH samples. This study was approved by the Institutional Review Board of UTEP.

Cell culture

The HCC cell lines SNU449 (poorly differentiated cells) and HepG2 (well-differentiated cells) were purchased from American Type Culture Collection (ATCC, Manassas, VA), and cultured in RPMI 1640 (HyClone, Logan, UT) and DMEM (Dulbecco's modified Eagle's medium), (HyClone, Logan, UT), respectively, which were supplemented with 10% fetal bovine serum (FBS), 100 units/mL penicillin and 100 unite/mL streptomycin. Cells were grown in 75 cm2 tissue culture flasks (Corning, Glendale, AZ) and allowed to reach 95% confluency. When collecting cells, cells were rinsed with DPBS (Gibco, Carlsbad, CA) 3 times and then incubated with trypsin-EDTA solution (Gibco, Carlsbad, CA) and harvested in a 15mL centrifuge tube for further study.

Two-dimensional Gel Electrophoresis (2DE analysis)

HepG2 and SNU449 cell lysates were prepared and lysed in the rehydration sample buffer (Bio-Rad, Hercules, CA). After being vigorously vortexed for 90 minutes at room temperature (RT), the samples were next centrifuged at 16,000 g under 4 °C temperature for 20 minutes. The supernatant was collected, and the protein concentration was measured by Bradford Protein Assay kit (Bio-Rad, Hercules, CA). In the first-dimensional (1DE) analysis, 125 μL of 150 μg protein was loaded onto the isoelectric focusing (IEF) strip (Non-linear, pH 3–10, 7 cm, Bio-Rad, Hercules, CA). Then, 1DE was performed in the PROTEAN IEF system (Bio-Rad, Hercules, CA) with linearly increased voltage from 0 to 250V in 30 minutes, then gradually increased voltage to 400 V in 90 minutes, and finally kept at 4000 V for 25,000 kvh. The prepared IEF strips were subsequently used for the second-dimensional (2DE) gel electrophoresis. Strips were docked into 12% SDS-polyacrylamide (SDS-PAGE) gel by Overlay agarose (Bio-Rad, Hercules, CA), and the electrophoresis was accomplished under a constant voltage of 80 V for 2 hours. The SDS-PAGE gels were then transferred onto the nitrocellulose (NC) membrane (Osmonics Inc., MA) or stained with 0.1% Coomassie blue R-250 reagent. The protein spots were identified by PDQuest 2-DE analysis software v. 8.0(Bio-Rad, Hercules, CA).

Western blotting and proteomic analysis

After the 2DE was conducted, the SDS-PAGE gel was transferred onto the NC membrane under a constant current of 250 mA for 70 minutes. Non-fat milk (Bio-Rad, Hercules, CA) was diluted in 0.1% tris-buffered saline tween (TBST) to 5% to be used as the blocking reagent. The NC membranes were blocked overnight at 4 °C. The positive sera pool or the negative sera pool was diluted with 5% non-fat milk at a concentration of 1:200 and incubated with the NC membranes for 2 hours under RT. Horse-radish peroxidase (HRP)-conjugated goat anti-human IgG (Caltag

Laboratories, San Francisco, CA) was diluted with 5% non-fat milk at a concentration of 1: 10,000 and was subsequently incubated with the NC membrane for 1 hour at RT. The blotting spots were detected by the enhanced chemiluminescence (ECL) kit (Amersham, Arlington Heights, IL) with the iBright system (ThermoFisher, Waltham, MA).

Mass spectrometry analysis

The mass spectrometry services were completed by Applied Biomics (Hayward, CA). The target gel spots were digested in gel with modified porcine trypsin protease (Trypsin Gold, Promega). The digested tryptic peptides were desalted by Zip-tip C18 (Millipore). Peptides were eluted from the Zip-tip with 0.5 ul of matrix solution (a-cyano-4-hydroxycinnamic acid (5 mg/ml in 50% acetonitrile, 0.1% trifluoroacetic acid, 25 mM ammonium bicarbonate) and spotted on the AB SCIEX MALDI plate (Opti-TOFTM 384 Well Insert). MALDI-TOF MS and TOF/TOF tandem MS/MS were performed on an AB SCIEX TOF/TOF™ 5800 System (AB SCIEX, Framingham, MA). MALDI-TOF mass spectra were acquired in reflectron reflection positive ion mode, averaging 4000 laser shots per spectrum. TOF/TOF tandem MS fragmentation spectra were acquired for each sample, averaging 4000 laser shots per fragmentation spectrum on each of the 10 most abundant ions present in each sample (excluding trypsin autolytic peptides and other known background ions). Both the resulting peptide mass and the associated fragmentation spectra were submitted to the GPS Explorer workstation equipped with MASCOT search engine (Matrix Science) to search the Swiss-Prot database. Searches were performed without constraining protein molecular weight or isoelectric point, with variable carbamidomethylation of cysteine and oxidation of methionine residues. One missed cleavage was also allowed in the search parameters. Candidates with either protein score C.I.% or Ion C.I.% greater than 95 were considered significant.

Indirect Immunofluorescence Assay (IIF)

An indirect immunofluorescence assay was performed on HEp-2 antinuclear antigen tissue slides (Bion Enterprises, Des Plaines, IL). The sera were diluted at 1:40 in phosphate-buffered saline (PBS), pH 7.4, and incubated with the slides for 30 min at RT. After extensive washing with PBS buffer, the slides were incubated with fluorescein isothiocyanate (FITC) conjugated goat-antihuman IgG (Santa Cruz Biotechnology, CA) as secondary antibody diluted 1:1000 in PBS for 30 min at RT. The slides were then washed three times with PBS and added mounting media containing 1.5 μg/mL 4′,6′-diamidino-2phenylindole (DAPI) (Vector Laboratories Inc. Burlingame, CA) to prevent photobleaching. The slides were then examined under confocal fluorescence microscopy at 400× magnification. Images were acquired using the software Zeiss Zen 3.2 (Zeiss, San Diego, CA). The immunofluorescent intensity were measured by Zeiss Zen software, which an average intensity of 20 cells within one sample represents its overall intensity.

Bioinformatic analysis

Driver genes of HCC were acquired from the IntoGen database and were subsequently screened by the Oncomine database. The genes in HCC groups retaining an mRNA expression greater than 1.5-fold change and *p* value less than 0.001 in no less than 2 cohorts were deemed as hits. The hits were validated in a separate TCGA database via the UALCAN algorithm with the same criteria as the Oncomine database mining. Enrichment analysis of proteins from SERPA results was implemented by R software (V. 4.1.1). The codes are listed below:

```
library(clusterProfiler), library(org.Hs.eg.db), library(enrichplot), library(ggplot2)

select_feature=read.table("gene names.txt",header=F,sep=",")
```

```
common=select_feature[[1]]

entrezid=bitr(common,fromType = "SYMBOL", toType = "ENTREZID",OrgDb = "org.Hs.eg.db")

ego <- enrichGO(OrgDb="org.Hs.eg.db", gene = entrezid$ENTREZID, ont = "MF", pAdjustMethod="BH",pvalueCutoff = 0.05, readable= TRUE)

barplot(ego,showCategory = 10)

summary(ego)

as.data.frame(ego)

cnetplot(ego, categorySize="pvalue",showCategory = 15)

heatplot(ego)

goplot(ego)

x2 <- pairwise_termsim(ego)

emapplot(x2)

emapplot(x2, showCategory = 10, color = "p.adjust", layout = "nicely")

ekk <- enrichKEGG(gene= entrezid$ENTREZID,organism = 'human', pvalueCutoff = 0.05, pAdjustMethod="BH",qvalueCutoff = 0.05)

dotplot(ekk,showCategory = 5,font.size=15)

edox <- setReadable(ekk, 'org.Hs.eg.db', 'ENTREZID')

cnetplot(edox, categorySize="pvalue",colorEdge=TRUE)
```

```
terms <- ego$Description[1:3]

pmcplot(terms, 2010:2018, proportion=FALSE)

pool=rbind(ego1[1:10],ego2[1:10],ego[1:10])

dat=data.frame(name=pool$Description,number=pool$Count,Ontology=c(rep("Molecula
r function",10),rep("Cellular component",10),rep("Biological process",10)))

ggplot(dat,aes(x=factor(name,levels=(name)),y=number,fill=Ontology))+geom_bar(stat=
"identity",position=position_dodge(0.5),width=0.8)+coord_flip()+theme_bw()+xlab("")+
ylab("Number of genes")
```

Enzyme-linked immunosorbent assay (ELISA)

Recombinant proteins were commercially purchased: Merlin, SETDB1, BRG1, RNA helicase A, CALR, and TPM3 were purchased from Signalway Antibody (Greenbelt, MD); HSPA5, ACTG1, ENO1, NRAS, ERK2, p16, and GMPS were purchased from Lifespan Biosciences (Seattle, MA); HSP60, HSP70, and TPI were purchased from Abcam (Boston, MA); GNAS and P4HB were purchased from Aviva Systems Biology (San Diego, CA); DNMT3A was purchased from Novus Biologicals (Centennial, CO). The proteins' purity was > 90%. When coating onto the polystyrene 96-well plates, the proteins were diluted to a final concentration of 0.75 μg/mL in pH 7.2 phosphate-buffered saline (PBS) (Gibco, Waltham, MA), each well was loaded with 100 μl of diluted protein. The 1:200 diluted serum samples were used as primary antibodies for 100 μL/well and stayed at 37 °C for 1.5 hours. Next, the plates were rinsed with PBS 3 times, 1:1000 HRP-conjugated goat anti-human IgG was used as the secondary antibody

for 100 µL/well and stayed at 37 °C for 1 hour. After another 3 times of PBS washes, the chromogenic substrate ABTS (2,2′-azinobis (3-ethylbenzthiazoline-6-sulfonic acid)) (Alfa Aesar, Tewksbury, MA) was applied and an average optical density (OD) of 405 nm was used for plate reading.

Statistical analysis

The cutoff value in the results of ELISA tests was defined as the mean of optical density (OD) value plus 2 times of standard deviation of OD value from the NHS group. Non-parametrical Mann-Whitney test was used to compare the mean OD value between 2 different groups. The diagnostic value of anti-TAA autoantibodies was represented by receiver operating characteristic (ROC) analysis of single variables, leading to the estimates of area under ROC curve (AUC) value with a 95% confident interval (CI). Statistical analysis was accomplished by SPSS v.27.0 and GraphPad v. 9.0 software. Differences were considered statistically significant with a significant level (p) less than 0.05.

Chapter 4: 9 potential TAA targets were identified by SERPA in Hispanic HCC

Overview

Liver cancer has been a global challenge for decades, attributed to its high prevalence and mortality rates. It is estimated that by 2025 there will be over 1 million victims affected by liver cancer annually [1]. In the United States, the incidence of HCC has doubled over the past 2 decades due to increasing hepatitis C virus-infected and non-alcoholic fatty liver disease patients [80]. Among all the ethnic groups in the United States, the Hispanic group surpassed the Asian/Pacific group in 2012 and had the highest incidence rate [81]. Although retaining such a high prevalence rate, the Hispanic group gained little attention for its diagnostic and prognostic surveillance research. The distinct risk factor and pathogenesis of Hispanic HCC patients make it a critical issue in HCC diagnosis and precision medical treatment [30, 31]. In this study, we are trying to identify novel TAA targets exclusively for Hispanic HCC patients with SERPA. By incubating HCC sera and normal healthy control sera (NHS) with proteins from HCC cell lysate, the differentially expressed proteins will be identified by mass spectrometry and deemed as potential targets. We found 9 potential TAA targets for Hispanic HCC, which are Endoplasmic reticulum chaperone (HSPA5), 60 kDa heat shock protein (HSP60), Triosephosphate isomerase (TPI), Protein disulfide-isomerase (P4HB), Actin cytoplasmic 2 (ACTG1), Heat shock 70 kDa protein (HSP70), Alpha-enolase (ENO1), Tropomyosin alpha-3 chain (TPM3), and Calreticulin (CALR). We also did enrichment analysis and functional annotation on the identified targets, showing that most proteins have functions in protein folding and ubiquitin ligase binding, which partially unveiled the mechanism of Hispanic HCC tumorigenesis.

Rationale

Anti-TAA autoantibodies have long been proven effective in early tumor diagnosis. Many TAAs are potential biomarkers in different malignancies, including hepatocellular carcinoma (HCC). Our lab has published several studies on HCC TAAs and anti-TAA autoantibodies as a diagnostic tool. IMP2/p62 was first investigated as a diagnostic biomarker in HCC by Zhang et al. in 2001 [82]. Shao et al. identified HSPA5 as a diagnostic biomarker for HCC and has a sensitivity of 71.4% together with alpha fetal protein (AFP) [64]. HSP60 and HSP70 were also validated as appropriate biomarkers for HCC early diagnosis by Looi et al. [63]. Although numerous TAA biomarkers were investigated, most studies focused on Asian patient samples; few have drawn attention to other ethnicities. The Hispanic population has a higher incidence and mortality rates of HCC in the United States than other ethnic groups, which urgently needs novel biomarkers for Hispanic HCC diagnosis.

Previous studies from our lab have confirmed two-dimensional gel electrophoresis (2DE) as a time- and cost-efficient assay for discovering novel TAA biomarkers from serum samples. Bo et al. used HepG2 cell lysates as proteome base together with liver fibrosis and normal healthy sera, and identified ENO1 as a potential biomarker for liver fibrosis diagnosis [83]. Dai et al. also found ENO1 as an effective diagnostic biomarker based on SERPA in non-small cell lung cancer, proving the feasibility of SERPA [84].

Experimental design

To identify novel TAA and anti-TAA autoantibody targets, we first need to screen out positive sera with relatively high autoantibody expression from the Hispanic HCC group. To this end, we tested all the HCC sera with western blotting. By incubating with HepG2 cell lysates, the

intensity of immune blotting with HepG2 proteome will be shown. Those sera with strong bands will be deemed as positive sera with significant autoantibody expression while NHS sera with low autoantibody expression as negative sera. Next, we will execute an indirect immunofluorescence assay (IIF) to validate the result. By incubating individual sera with fixed Hep-2 cells on the slides, the quantities of anti-nuclear antibodies will be quantified by immunofluorescent intensity.

In previous HCC biomarker studies using two-dimensional electrophoresis (2DE), only HepG2 cell line was used, representing well-differentiated HCC status, showing an incomplete picture of HCC development. To address this issue, we decided to choose 2 separate cell lines, HepG2 and SNU449, representing well-differentiated and poor-differentiated HCC conditions [85, 86]. Using more than one cell line will increase potential TAA spots in electrophoresis gels, providing a more comprehensive picture of differentially expressed proteins within cancerous and pre-cancerous stages. For the first-dimension electrophoresis, we will use gel strips containing the isoelectric point (IEF) from 3 to 10, representing the IEF of most of the expressed protein. After identifying the potential protein targets, we are trying to investigate the internal correlation of the proteins using bioinformatic tools. By showing the functional annotation of the specific genes of the protein targets, we will understand the internal mechanisms better and possibly provide insight into the signaling correlation of the proteins in the tumorigenesis of Hispanic HCC.

Results

Western blotting was used to detect evident immunoproteomic signals in Hispanic HCC sera. HepG2 and SNU449 cell lines were cultured, and whole protein was extracted. Subsequently, the proteome was electrophoresed in SDS-PAGE gel and transferred onto NC membranes. Membrane strips were incubated with 25 Hispanic HCC sera and 25 NHS sera, the location and concentration of the protein band were shown on the NC membrane. The results of positive HCC sera and NHS sera control are exhibited. (Fig 1.)

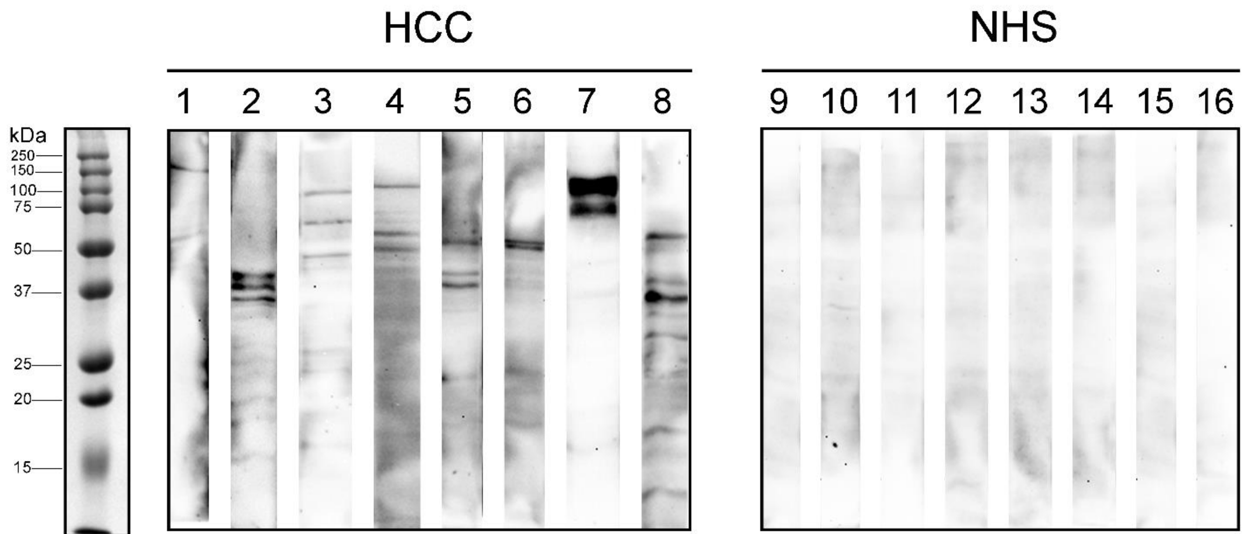
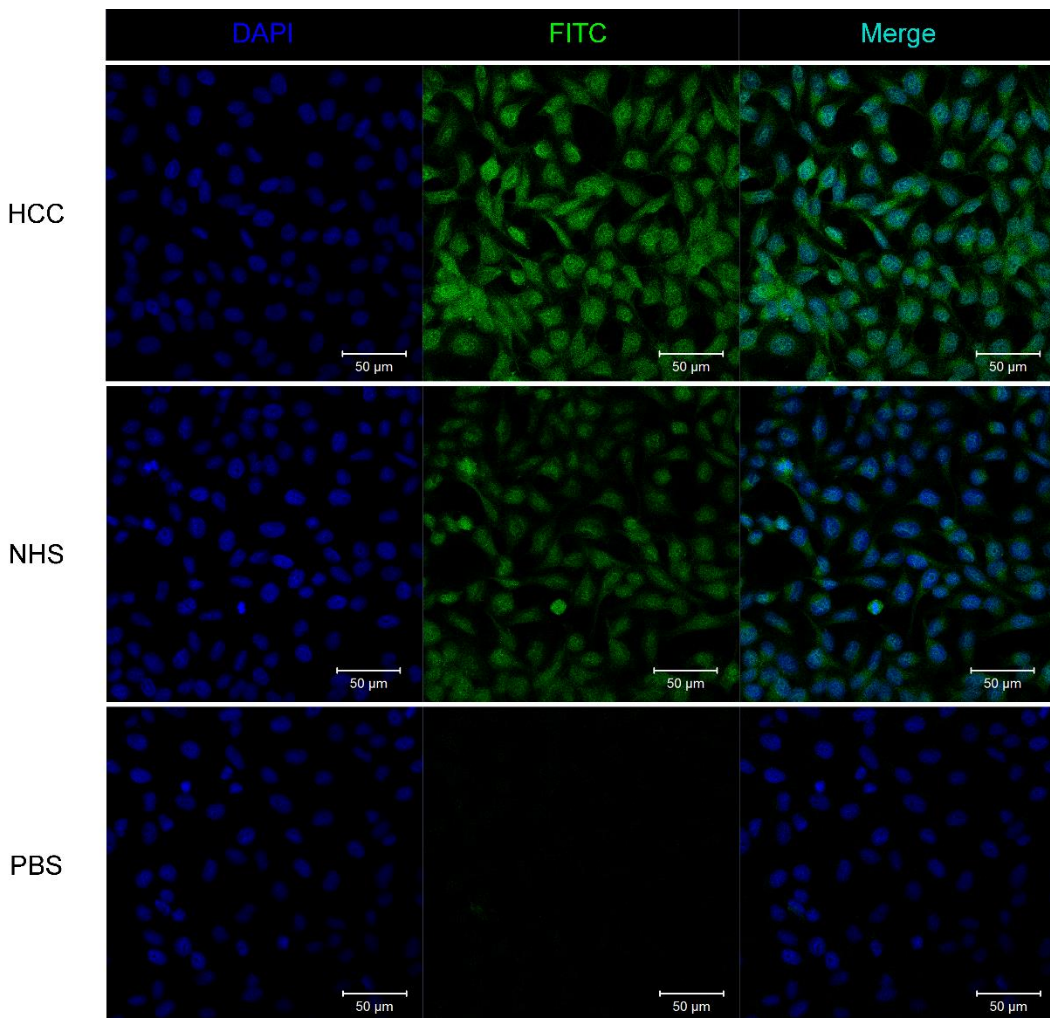


Fig 1. Positive sera screening by western blotting. Detection of autoantibodies against cellular protein antigens extracted from Hep G2 cell lysates. Sera sample from Hispanic patients with HCC and NHS were incubated with the whole proteome of the antigen. Western blotting results show 8 positive sera with strong immunoreactive bands from HCC and 8 negative sera with low or no immunoreactive bands from NHS.

The IIF assay was performed apart from the western blotting assay to evaluate the overall autoantibody intensity within HCC sera and NHS sera on Hep-2 cell slides (Fig 2.). While PBS and NHS sera show relatively no or low staining intensity, positive HCC sera show higher intensity in both cytoplasmic and perinuclear staining patterns. HCC, NHS, and PBS groups showed a median intensity of 33.11, 20.90, and 4.01 separately. The average intensity of HCC sera is stronger than NHS sera and PBS negative control (p<0.01), showing overall auto-antibody frequency is higher in HCC sera.



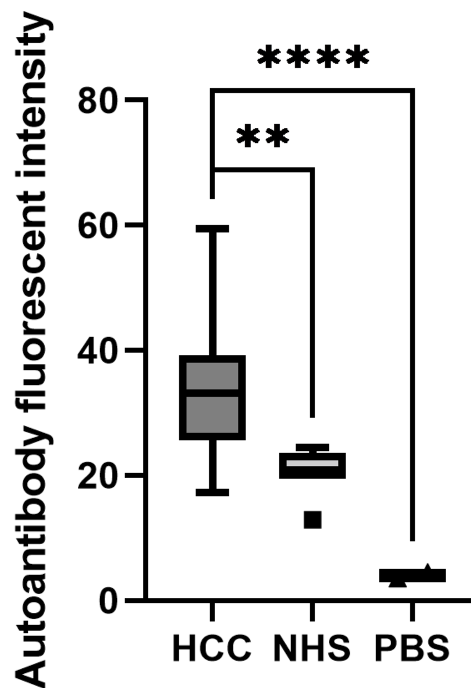


Fig 2. Autoantibody pattern and intensity of positive sera and negative sera samples by indirect immunofluorescence (IIF). While PBS, as a negative control, shows a low autoantibody frequency, HCC and NHS sera have higher auto-antibody frequency shown on IIF slides. Likewise, auto-antibody frequency in HCC sera is significantly higher than in NHS sera. ($p<0.05$)

To select positive HCC sera with high autoantibody expression, we choose samples with both prominent protein bands on western blotting and relatively high intensity on IIF slides. Eventually, 8 HCC sera were designated as positive sera with high autoantibody expression. 8 NHS sera with low auto-antibody intensity were found as a counterpart to form negative sera. The information on positive sera was included in Table 1, including the molecular weight of positive WB band, intensity, and staining pattern of IIF (Table 1.).

Table 1. The properties of autoantibodies in the positive sera

Positive HCC sera number	WB band (kD)	IIF intensity	Staining Pattern
#2	150， 53	49.23	Cytoplasmic
#4	40， 37， 32	23.48	Cytoplasmic
#7	90， 61， 48	26.43	Cytoplasmic
#11	92， 55， 47	27.42	Cytoplasmic
#19	51， 40， 37	32.28	Cytoplasmic
#20	51， 49	37.45	Cytoplasmic
#24	97， 70	34.88	Nuclear
#26	51， 37， 33， 30， 18， 13	25.48	Cytoplasmic

Information of positive HCC sera. The molecular weight of prominent band in WB, intensity, and staining patterns of IIF were recorded for each serum. WB: western blotting; IIF: indirect immunofluorescence.

HepG2 and SNU449 cell lines were cultured, and whole protein was extracted subsequently. After 1st dimensional IEF electrophoresis and 2nd-dimensional SDS-PAGE gel electrophoresis, protein spots were separated on the SDS-PAGE gel. Three gels were acquired at the same time for further processing. One gel was stained with Coomassie blue solution while the other 2 gels were transferred onto NC membranes and further immunoblotted with HCC sera pool and NHS sera pool, respectively. Protein spots with remarkable staining on NC membranes were tracked back to Coomassie blue gel. (Fig 3.) Assays were repeated 3 times, and all the susceptible protein spots were exercised for further MALDI-TOF/TOF analysis.

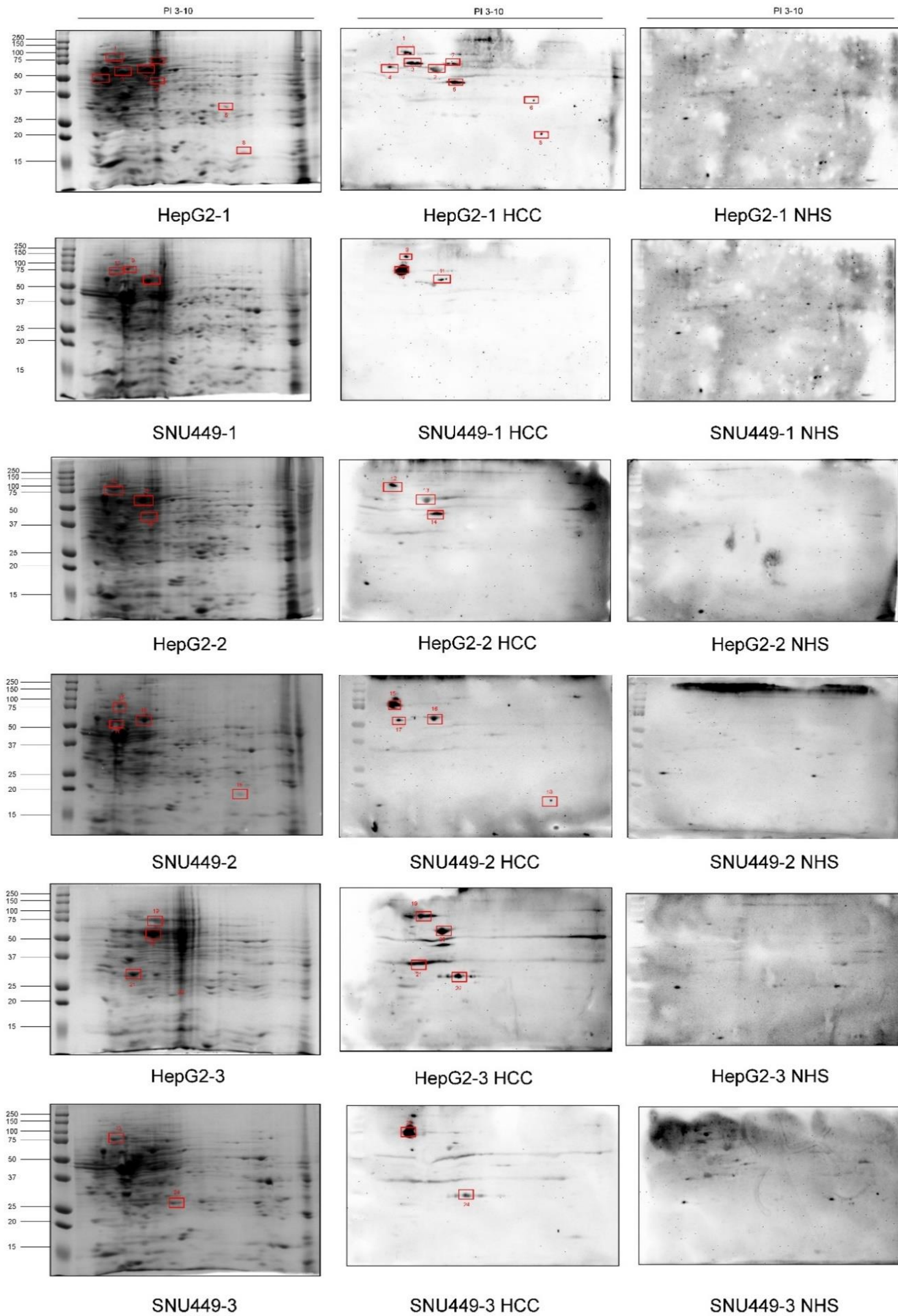

PI 3-10
PI 3-10
PI 3-10
250
150
100
75
50
37
25
20
15
HepG2-1
HepG2-1 HCC
HepG2-1 NHS
SNU449-1
SNU449-1 HCC
SNU449-1 NHS
HepG2-2
HepG2-2 HCC
HepG2-2 NHS
SNU449-2
SNU449-2 HCC
SNU449-2 NHS
HepG2-3
HepG2-3 HCC
HepG2-3 NHS
SNU449-3
SNU449-3 HCC
SNU449-3 NHS

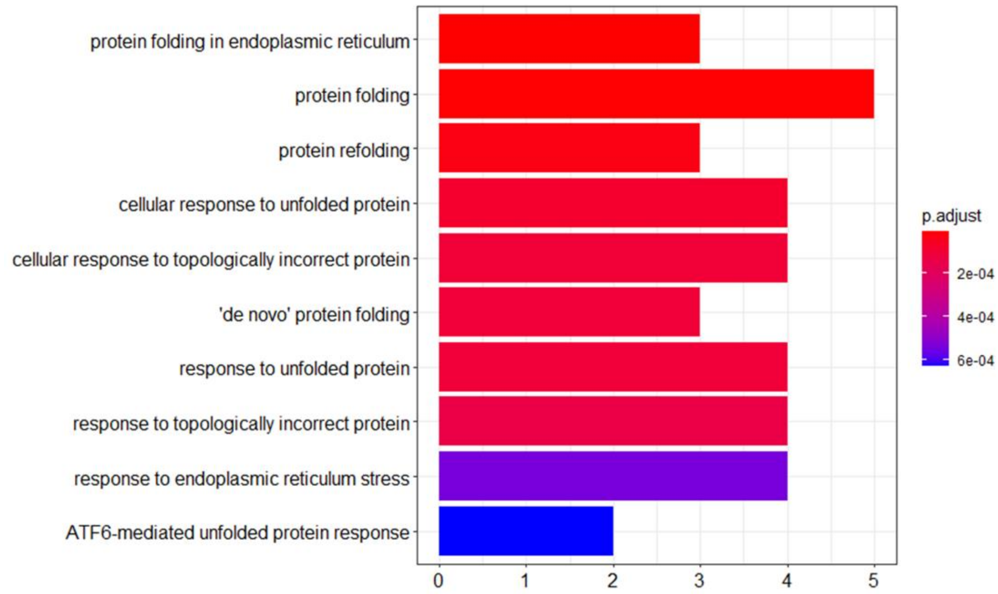
Fig 3. Identification of differentially expressed protein spots by SERPA in Hispanic HCC sera and NHS sera. Proteins extracted from HepG2 and SNU449 cell lysates were first electrophoresed on isoelectric focusing gel strips and then separated on SDS-PAGE gel by western blotting. Protein spots with strong signals either in the HCC group or in the NHS group are defined as potential differentially expressed antigens.

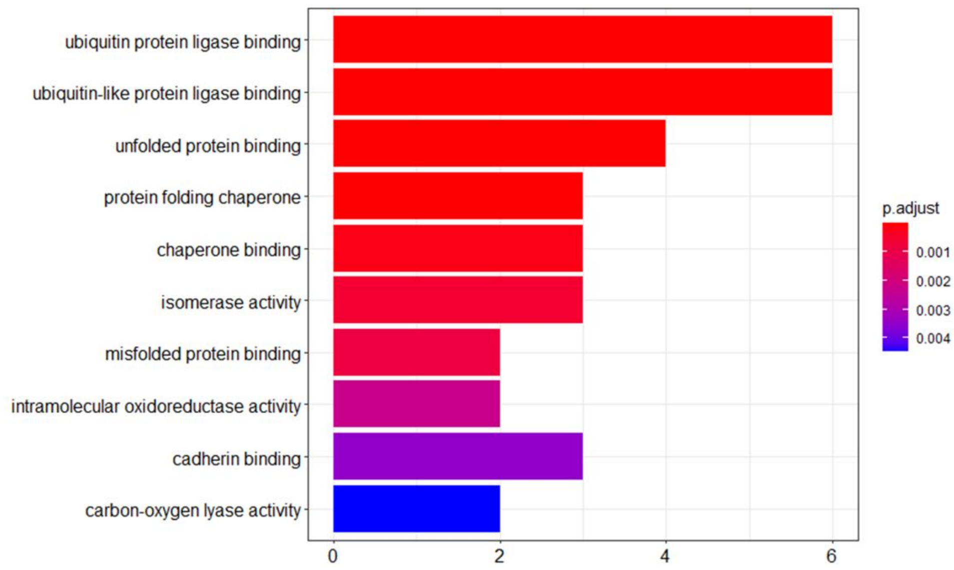
After matching the Coomassie-stained 2DE gels with the corresponding NC membrane, identified 24 immunoreactive protein spots (16 from HepG2 cell line and 8 from SNU449 cell line) were excised and sent to MALDI-TOF/TOF mass spectrometry analysis after trypsin digestion. The resulting MS/MS spectra were analyzed by GPS Explorer software equipped with the MASCOT search engine using the SwissProt database. The filtering process for top scores of protein hits in the excised immunoreactive spots was based on molecular weight, isoelectric point, percent coverage, number of unique peptides, and the total number of proteins identified for each spot. Protein spots with mismatched molecular weight and/or protein PI (isoelectric point) or confidence score (C.I.%) lower than 90% will be rejected.

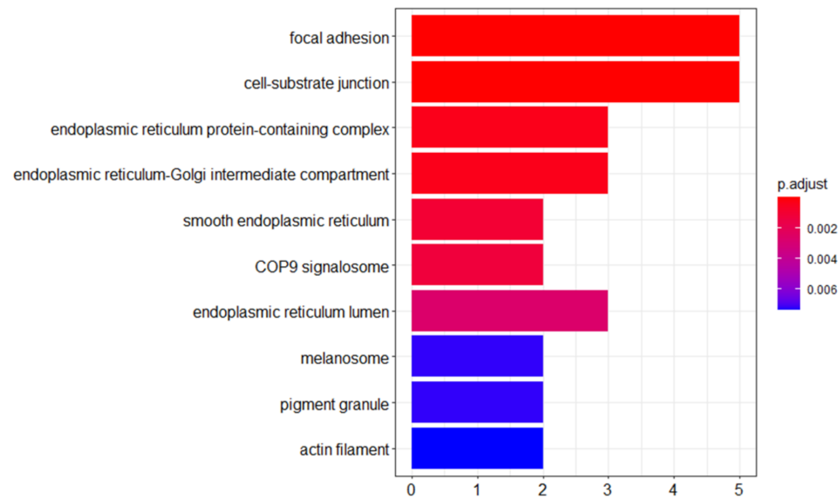
In the 24 samples, all of them have a 100-protein confidence score. After rejecting 3 protein spots with unexpected molecular weight and/or protein PI, 21 protein spots were deemed as potential proteins of TAAs with a high level of credibility (Table 2.). 9 protein hits (Endoplasmic reticulum chaperone BiP /HSPA5, 60 kDa heat shock protein /HSP60, Triosephosphate isomerase /TPI, Protein disulfide-isomerase /PDI, Actin, cytoplasmic 2 /ACTG1, Calreticulin/CRP55, Alpha-enolase /ENO1, Heat shock 70 kDa protein /HSP70, Tropomyosin alpha-3 chain/TPM3) were uncovered, among which the HSPA5 protein has the highest interest with 5 repetitions.

HSPA5, HSP60, TPI, and PDI were found both in HepG2 cell lines and SNU449, indicating their potentiality as TAA targets in further analysis.

We next investigated the GO analysis on the genes of differentially expressed hits. We showed the top 10 ranks of biological process, cellular component, and molecular function within 9 hits based on the minimum values of false discovery rate (FDR) and maximum counts of GO (Fig 4A-4C). The results indicate that most of the hits' functions centralize in protein folding, mainly involving ubiquitin-protein ligase folding and ubiquitin-like protein ligase folding. The main cellular structures in the process are located at the extracellular matrix and endoplasmic reticulum (ER). The enrichment map also shows protein binding and metabolic processes such as oxidation-reduction reaction (Fig 4D). These results show potential relationships between Hispanic HCC and specific metabolic disorders.


A
protein folding in endoplasmic reticulum
protein folding
protein refolding
cellular response to unfolded protein
cellular response to topologically incorrect protein
'de novo' protein folding
response to unfolded protein
response to topologically incorrect protein
response to endoplasmic reticulum stress
ATF6-mediated unfolded protein response
0
1
2
3
4
5
p.adjust
2e-04
4e-04
6e-04


B
ubiquitin protein ligase binding
ubiquitin-like protein ligase binding
unfolded protein binding
protein folding chaperone
chaperone binding
isomerase activity
misfolded protein binding
intramolecular oxidoreductase activity
cadherin binding
carbon-oxygen lyase activity
0
2
4
6
p.adjust
0.001
0.002
0.003
0.004



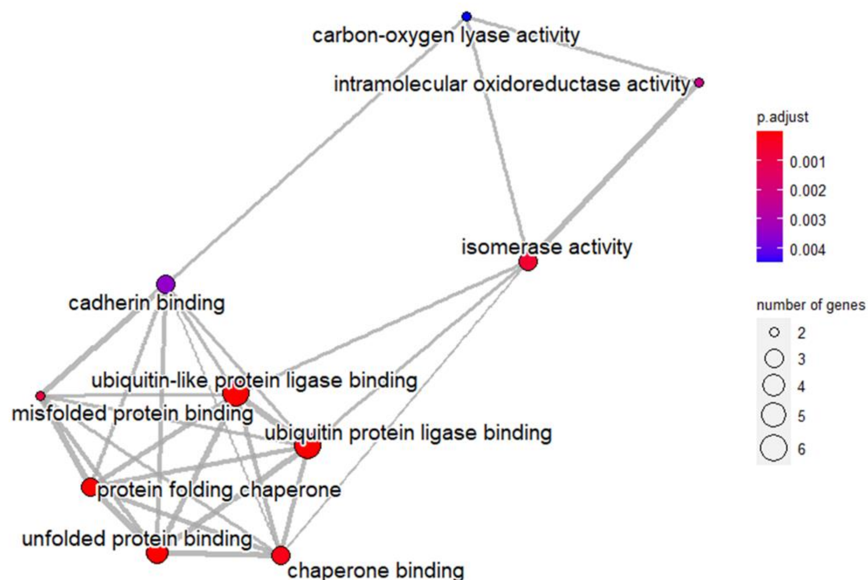


Fig 4. Enrichment analysis of potential TAA target identified by SERPA. Significantly enriched GO annotations of 2D protein hits genes, including HSPA5, HSPD1, TPI, P4HB, ACTG1, CALR, ENO1, HSPA1A, TPM3. (A) Biological process annotation of targeted genes. (B) Molecular function annotation of targeted genes. (C) Cellular component annotation of targeted genes. (D) Enrichment map of targeted genes.

Table 2. Identification of protein targets by mass spectrometry from SERPA

Protein spots	Name of protein	MW (Da)	PI	Protein score C.I.%
HepG2 - 1	Endoplasmic reticulum chaperone BiP/HSPA5	72,288	5.07	100
HepG2 - 2	60 kDa heat shock protein/HSP60	61,016	5.70	100
HepG2 - 3	Protein disulfide-isomerase/P4HB	57,081	4.76	100
HepG2 - 4	Calreticulin/CALR	48,112	4.29	100
HepG2 - 5	Actin, cytoplasmic 2/ACTG1	41,766	5.31	100
HepG2 - 6	Triosephosphate isomerase/TPI	26,653	6.45	100
HepG2 - 7	Alpha-enolase/ENO1	47,139	7.01	100
HepG2 - 9	Endoplasmic reticulum chaperone BiP/HSPA5	72,288	5.07	100
HepG2 - 10	60 kDa heat shock protein/HSP60	61,016	5.70	100
HepG2 - 11	Actin, cytoplasmic 2/ACTG1	41,766	5.31	100
HepG2 - 14	60 kDa heat shock protein/HSP60	61,016	5.70	100
HepG2 - 15	Tropomyosin alpha-3 chain/TPM3	32,930	4.68	100
HepG2 - 16	Triosephosphate isomerase/TPI	26,653	6.45	100
SNU449 - 1	Heat shock 70 kDa protein/HSP70	94,271	5.11	100
SNU449 -2	Endoplasmic reticulum chaperone BiP/HSPA5	72,288	5.07	100
SNU449 - 3	Endoplasmic reticulum chaperone BiP/HSPA5	72,288	5.07	100
SNU449 - 4	60 kDa heat shock protein/HSP60	61,016	5.70	100
SNU449 - 5	Protein disulfide-isomerase/P4HB	57,081	4.76	100
SNU449 - 6	Triosephosphate isomerase/TPI	26,653	6.45	100
SNU449 - 7	Endoplasmic reticulum chaperone BiP/HSPA5	72,288	5.07	100
SNU449 - 8	Triosephosphate isomerase/TPI	26,653	6.45	100

Convincible protein spots from 2D gels with confidence value over 95%. 21 out of 24 protein spots were identified as potential biomarkers with compatible molecular weight and isoelectric point. MW: molecular weight, PI: protein isoelectric point, C.I.: confidence value.

Chapter 5: 10 potential TAA targets were identified from HCC driver genes

Overview

Cancer driver genes are those capable of driving tumors with mutations. The mutated form of these genes could affect a set of key cellular functions, thus disturbing homeostatic development and causing tumors. Due to its relationship with carcinogenesis, driver genes with various mutations could be valid biomarkers in clinical investigation. It was shown that proteins with modified structures by point mutations are highly correlated with autoantibodies production [69]. Several studies have identified driver genes' effects as biomarkers in cancer diagnosis [87, 88]. To enlarge the pool of potential TAA targets for Hispanic HCC, we are trying to identify differentially expressed genes from IntoGen, the driver gene database. To accomplish this, we measured the mRNA expression of the HCC driver genes in the Oncomine database and verified the results in a separate TCGA database. Eventually, we discovered potential TAAs encoded from 10 driver genes, which are CDKN2A/p16, SETDB1/SETDB1, DHX9/DHX9, SMARCA4/SMARCA4, GNAS/GNAS, NF2/Merlin, DNMT3A/DNMT3A, NRAS/NRAS, GMPS/GMPS, and MAPK1/MAPK1.

Rationale

Although SERPA provides a time- and financially- efficient approach for discovering diagnostic biomarkers, there are still some drawbacks. The 2D electrophoresis limited the identification of certain classes of proteins, such as membrane proteins. Only sequential epitopes are detected due to the properties of the western blotting technique [89]. Therefore, an enlarged

pool for TAA targets is needed without being restricted in gel electrophoresis. Driver genes mutation can be accumulative and thus differentiate different cancerous and pre-cancer stages, which helps early diagnosis of cancer patients. Though gene signatures of driver genes are potential biomarkers, insights are yet to be provided into the correlation between TAAs and proteins encoded by driver genes. As a powerful diagnostic tool in clinics with accessible testing samples such as serum, TAAs reflect the picture of how proteome and genome modify in the tumor tissue and the secretion afterward, which is primarily associated with genomic mutation status. In the present study, we aim to define proper proteins encoded by driver genes with great anti-TAA autoantibody in serum, thus providing a wholesome portrait for HCC diagnosis.

Experimental design

Multiple bioinformatic tools have been established to investigate mutational driver genes in cancer. A state-of-the-art review has shown a rather integrated diver gene maps in pan-cancer using Integrative OncoGenomics pipeline [90]. By integrating numerous online databases, it shows 75 driver genes in hepatic cancer. Among those selected driver genes, we are supposed to screen the most differentially expressed driver genes between HCC patients and normal healthy patients by bioinformatical algorithms. The mRNA expression will be first investigated by the Oncomine database. From 8 cohorts in the Oncomine database, those driver genes with an mRNA fold change greater than 1.5 between liver cancer tissues and normal liver tissues and a p-value less than 0.001 will be deemed potential hits. All the hits will subsequently be verified in TCGA databases via visualized algorithm named UALCAN. Suppose the hits keep consistency with the same criteria in the Oncomine. In that case, those driver genes will be considered actual hits, and the proteins encoded by these driver genes will be deemed to be potential TAA targets.

Results

Driver gene dataset was downloaded from the IntoGen pipeline. Within liver cancer, there are a total of 75 driver genes out of 1616 samples from 9 cohorts. Those driver genes were screened within the Oncomine database for differential mRNA expression. Driver genes with more than a 1.5-fold change in HCC compared to normal tissue in no less than 2 cohorts will be deemed as hits. The mRNA expression of the hits will be validated in UALCAN by screening the TCGA database. Finally, 10 hits were identified by this filtering mechanism. (Table 3.). Mutations refer to the mutation numbers of the driver genes in HCC, which include truncation, missense, and synonymous mutations. Samples refer to the number of patient samples in 1616 HCC samples from the IntoGen database. Generally speaking, the higher the mutation number and patient number, the greater possibility that a driver gene will lead to the tumorigenesis of HCC.

Table 3. Candidate TAA targets identified from HCC driver genes

Gene Name/ Protein symbol	Mutations	Samples	Cohorts with Fold Change > 1.5	Functions of Protein
CDKN2A/ p16	34	29	2	Induce cell cycle arrest in G1 and G2 phases
SETDB1/ SETDB1	32	12	3	Histone methyltransferase

DHX9/ RNA helicase A	33	11	2	Nucleic acid helicase that unwinds DNA and RNA
SMARCA4/ BRG1	30	9	3	Transcriptional activator
GNAS/ GNAS	26	8	2	The essential component of adenylyl cyclase signal transduction pathways
DNMT3A/ DNMT3A	25	4	2	DNA methylation
NF2/ Merlin	14	6	3	Regulator of Sav/Wts/Hpo signaling pathway
NRAS/ NRAS	12	4	2	Bind GDP/GTP and possess intrinsic GTPase activity
MAPK1/ ERK2	10	3	2	Serine/threonine kinase in MAP signal transduction pathway
GMPS/ GMPS	9	4	4	Catalyzes the conversion of XMP to GMP

Proteins encoded by driver gene hits to be potential biomarkers in Hispanic HCC.

Chapter 6: Autoantibody prevalence of the potential TAA targets and establishment of anti-TAA autoantibody panel

Overview

The enzyme-linked immunoassay (ELISA) is an easy to operate and cost-efficient assay for testing antibody-antigen reactions. It has been validated multiple times that ELISA provides robustness and reproducibility in anti-TAA autoantibody detection. To test the titer and frequency of the anti-TAA autoantibody of the 19 potential TAA targets from SERPA and driver genes, we used ELISA to test the autoantibody. The titer of the autoantibodies was quantified as the optical density value by ELISA. The protein products of the 19 targets were commercially purchased and incubated with 4 Hispanic sera groups, including HCC, liver cirrhosis (LC), chronic hepatitis (CH), and NHS. The cutoff values were set according to the OD values of the NHS sera group, and the frequencies of the individual protein targets were statistically analyzed. As a result, 4 targets showed significant differences between the HCC group and NHS group, including DNMT3A, p16, HSP60, and HSP70. Nest, we assembled the 4 targets as an anti-TAA autoantibody panel, which showed a sensitivity of 75%, compared to the highest single autoantibody sensitivity of 45.8%. We identified 4 TAAs as diagnostic biomarkers for Hispanic HCC patients in the present study. The panel of the 4 anti-TAA autoantibodies shows greater diagnostic value than single ones.

Rationale

A serological assay that is applied to clinics should meet several prerequisites: high robust, easy to perform, cost-effective, and suitable for high-volume testing, and a stable shelf-life of the

testing reagents [45]. Among all the immune assays that target antibody-antigen reactions, ELISA shows prominent advantages in autoantibody detection compared to other counterparts such as immunohistochemistry and immunoprecipitation [91]. The substrate of recombinant protein also provides solid support and is minimally invasive by using patients' sera samples. All the properties make the ELISA an effective immunoassay in the clinical translation of TAA biomarkers. Nonetheless, in the clinical diagnosis, single TAAs might not be able to provide enough sensitivity because they only reflect part of the autoantibodies distribution. A combination of multiple TAAs has been proved to be more sensitive and cover more potential patients. Okada et al. combined 6 autoantibodies in the diagnosis of HCC, and pointed out an increase of sensitivity of 56% in the TAA panel, compared to 11~19% in single TAAs and 41% in AFP [92]. They also found that positivity for the TAA panel is associated with poor prognosis. Hoshino et al. also showed better discrimination of early-stage gastric cancer from normal control by applying a multiple autoantibodies panel [93].

Experimental design

The recombinant proteins of all the 19 potential TAA targets were commercially purchased as substrates in ELISA tests. The feasibility of the recombinant proteins will be validated by SDS-PAGE gel to ensure the correct molecular weight and protein concentration. A series of standard bovine serum albumin (BSA) with volumes of 0.25 μg, 0.5 μg, and 0.75 μg will be used as controls, and a volume of 0.5 μg recombinant proteins will be loaded into the gel. To investigate the autoantibody titer among different tumorigenesis stages, we will use 4 sera groups, including HCC sera, LC sera, CH sera, and NHS sera. In the present study, all the sera samples will be from Hispanic individuals. The Hispanic sera samples will be acquired from the sera bank from Dr.

Jianying Zhang's lab in the University of Texas at El Paso. We will use the same substrate, sera, and anti-human IgG concentrations to maintain consistency. In the statistical analysis, the cutoff value will be the mean value plus 2 times of standard deviation value of the OD values from each NSH group. Positive sera from HCC, LC, or CH are those with OD values above the cutoff values. The Ki-square test accomplishes the comparison of positive frequencies from different sera groups.

Results

To investigate the autoantibody titer and frequency against potential TAAs screened from the immunoproteomic approach and driver genes pipeline, quantitative enzyme-linked immunoassay (ELISA) was applied. The recombinant protein products of the potential TAA targets were used as coating antigens, and the molecular weight and protein concentration were ascertained by SDS-PAGE gel electrophoresis (Fig 5.). All the 19 protein bands showed consistent signals with 0.5 μg BSA, and the molecular weight positions corresponded to the product manuals, indicating that the 19 recombinant proteins are qualified for subsequent ELISA tests. A cohort of Hispanic sera was used as testing samples, which include 24 HCC sera, 20 liver cirrhosis (LC) sera, 26 chronic hepatitis (CH) sera, and 40 normal healthy control sera (NHS). The cutoff value was established as the mean optical density (OD) value of 40 NHS sera plus two standard deviations (SD). The titers of autoantibodies against 19 antigens in different groups are shown in Figure 6. The high titers are displayed in several HCC sera and many LC sera, compared to lower titers in CH sera and NHS sera. To investigate the frequency of autoantibodies against all the antigens, the positive rates of each sera group from each antigen are listed in Table 4. Among the 19 potential TAA targets, 4 showed significantly higher autoantibody frequency in the HCC sera compared to the NHS sera, which are DNMT3A, CDKN2A, HSP60, and HSPA5. We next

measured the titer distribution within the 4 TAAs. All the TAAs showed significantly higher OD values in the HCC group compared to the NHS group ($p < 0.05$) (Fig 6.). These results indicate that these 4 TAAs could be biomarkers for Hispanic HCC patients by distinguishing Hispanic HCC and normal healthy controls.

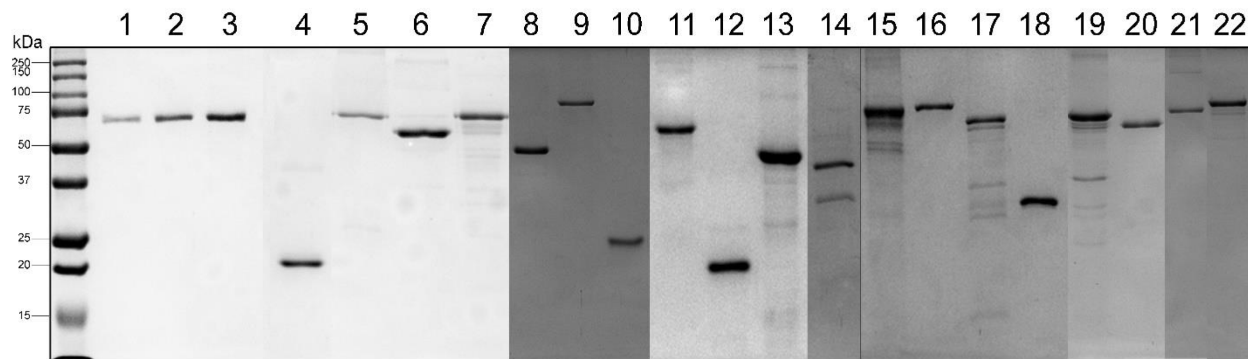


Figure 5. Molecular weights and protein concentrations of 19 recombinant proteins on SDS-PAGE gel. BSA samples were loaded as controls with concentrations of 0.25 µg, 0.5 µg, and 0.75 µg. The recombinant proteins were loaded with a concentration of 0.5 µg. The marker ladder confirms the molecular weights of the recombinant proteins. 1~3: 0.25~0.75 µg BSA; 4: SETDB1; 5: NF2; 6: SMARCA4; 7: DHX9; 8: MAPK1; 9: GMPS; 10: NRAS; 11: GNAS; 12: CDKN2A; 13: ACTG1; 14: DNMT3A; 15: CALR; 16: HSP70; 17: P4HB; 18: TPI; 19: ENO1; 20: TPM3; 21: HSP60; 22: HSPA5.

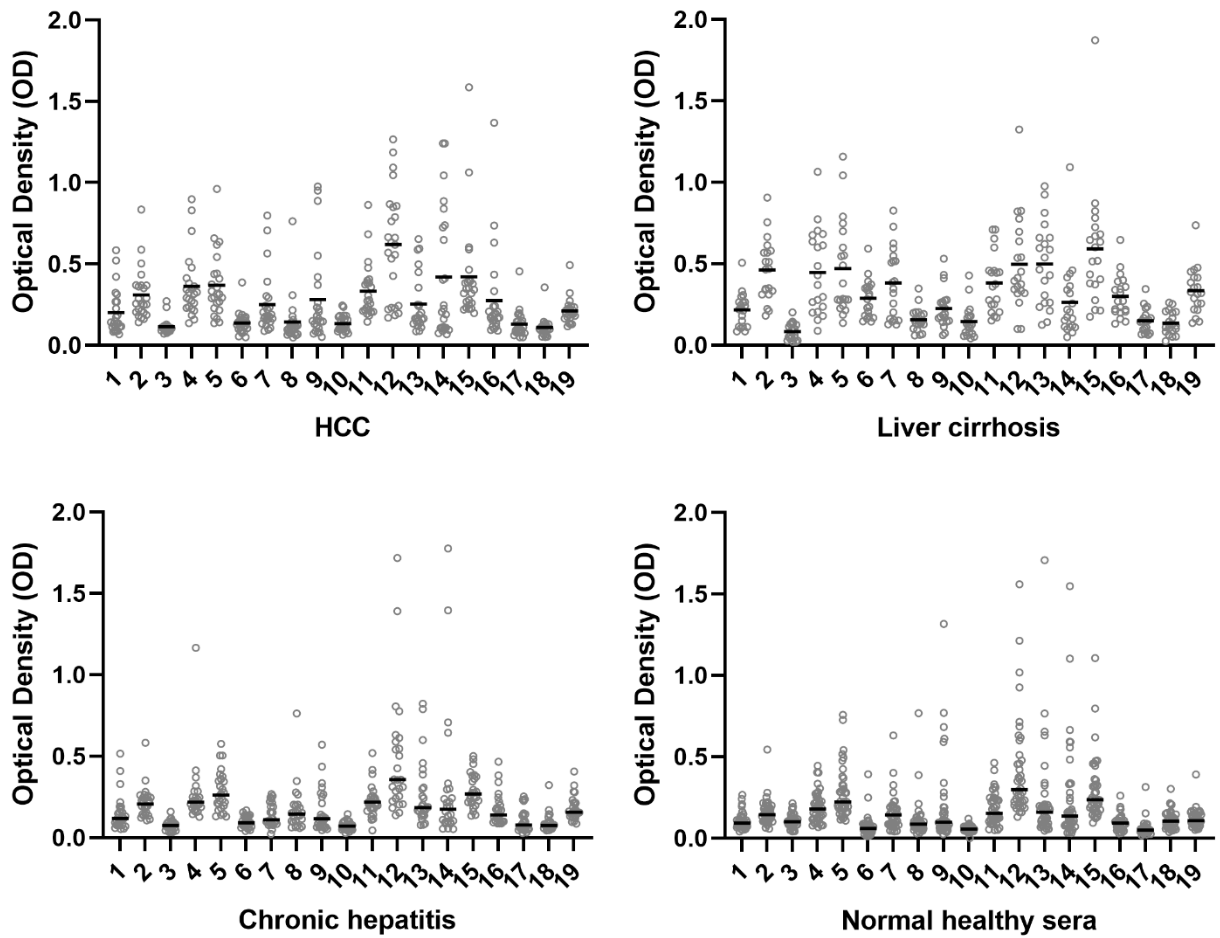


Figure 6. Autoantibody titer distribution of 19 potential TAA targets. The ranges of autoantibody titers are shown as absorbance units obtained from enzyme immunoassays. The mean optical density (OD) is shown for every condition. 1, HSPA5; 2, HSP60; 3, TPI; 4,P4HB; 5, ACTG1; 6, HSP70; 7, ENO1; 8, TPM3; 9, CALR; 10, p16; 11, SETDB1; 12, DHX9; 13, SMARCA4; 14, GNAS; 15, NF2; 16, DNMT3A; 17, NRAS; 18, GMPS and 19, ERK2.

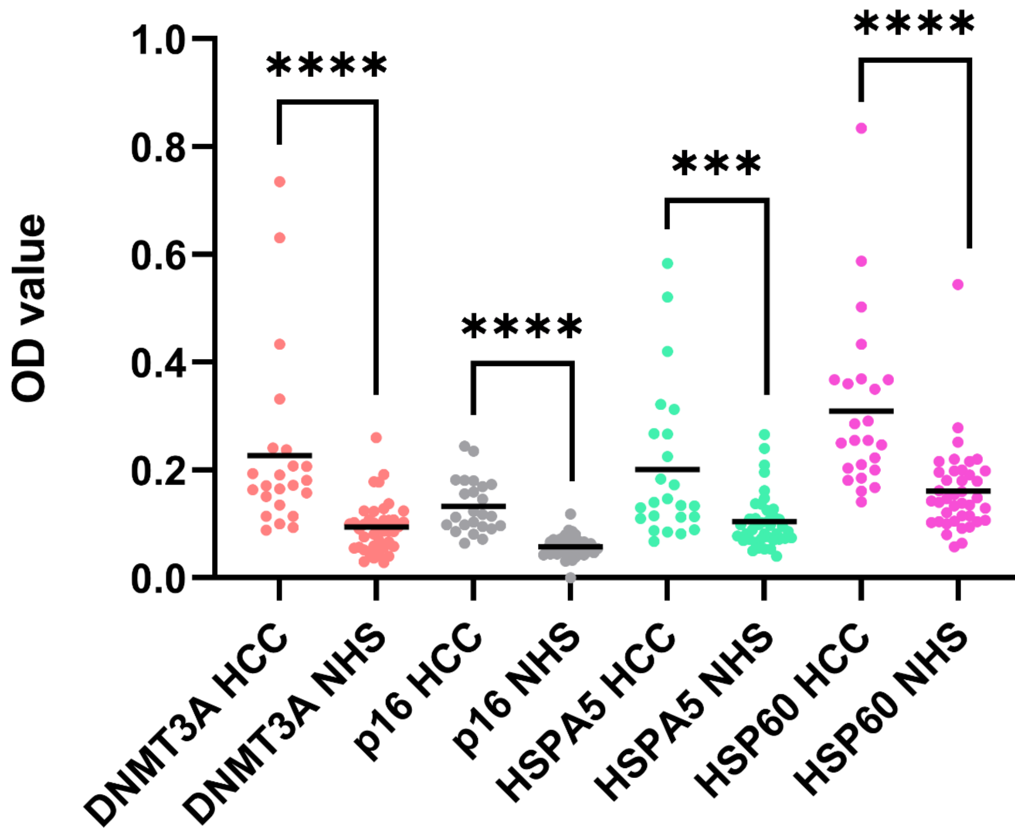


Figure 7. Autoantibody titers of the 4 TAAs in the Hispanic HCC and NHS groups. Significant differences of optical density (OD) values are observed in all 4 anti-TAA autoantibodies. *** $p < 0.001$; **** $p < 0.0001$.

Table 4. Frequency of the 19 anti-TAA autoantibodies in Hispanic sera cohort

Antigens	**HCC (n=24)**	**LC (n=20)**	**CH (n=26)**	**NHS (n=40)**
HSPA5	8 (33.3%) *	12 (60.0%) **	4 (15.4%)	3 (7.5%)
HSP60	9 (37.5%) **	15 (75.0%) **	2 (7.7%)	1 (2.5%)
TPI	2 (8.3%)	1 (5.0%)	0 (0.0%)	0 (0.0%)
P4HB	6 (25.0%)	10 (50.0%) **	2 (7.7%)	3 (7.5%)
ACTG1	4 (16.7%)	6 (30.0%) *	1 (3.8%)	2 (5.0%)
HSP70	1 (4.2%)	14 (70.0%) **	0 (0.0%)	2 (5.0%)
ENO1	4 (16.7%)	10 (50.0%) **	0 (0.0%)	2 (5.0%)
TPM3	5 (20.8%)	2 (10.0%)	1 (3.8%)	2 (5.0%)
CALR	3 (12.5%)	0 (0.0%)	0 (0.0%)	2 (5.0%)
p16	10 (41.7%) **	9 (45.0%) **	6 (23.1%) *	1 (2.5%)
SETDB1	3 (12.5%)	4 (20.0%)	1 (3.8%)	1 (2.5%)
DHX9	4 (16.7%)	1 (5.0%)	2 (7.7%)	3 (7.5%)
SMARCA4	0 (0.00%)	2 (10.0%)	1 (3.8%)	1 (2.5%)
GNAS	4 (16.7%)	1 (5.0%)	2 (7.7%)	2 (5.0%)
Merlin	2 (8.3%)	5 (25.0%)	0 (0.0%)	2 (5.0%)
DNMT3A	11 (45.8%) **	17 (85.0%) **	5 (19.2%)	2 (5.0%)
NRAS	5 (20.8%)	7 (35.0%) **	4 (15.4%)	2 (5.0%)
GMPS	1 (4.2%)	2 (10.0%)	1(3.8%)	1 (2.5%)
MAPK1	3 (12.5%)	15 (75.0%) **	4 (15.4%)	2 (5.0%)

HCC: hepatocellular carcinoma; LC: liver cirrhosis; CH: chronic hepatitis; NHS: normal healthy sera. * $p < 0.05$, ** $p < 0.01$, cutoff value = mean of 40 NHS + 2 SD.

To better display the diagnostic values of 4 autoantibodies, the area under the receiver operator characteristic curve (AUC) was analyzed. As shown in Figure 7., ROC analysis showed that the anti-TAA autoantibody could differentiate the Hispanic HCC group from normal Hispanic controls, with AUC values varying from 0.7505 to 0.8885. Next, we evaluated the accuracy of the 4-autoantibodies panel as a diagnostic platform for Hispanic HCC. To enhance the antibody detection for diagnosis, we sequentially add the antigen to the TAA panel, from DNMT3A, which has the highest frequency of antibody, to HSPA5 with the lowest frequency (Table 4.). A stepwise sensitivity increase reaches 75%, compared to the most elevated single anti-TAA antibody (45.8%). The specificity decreased from 95% to 85%. The TAA panel's positive predictive value (PPV) and negative predictive value (NPV) were 83.3% and 77.3%, respectively. The likelihood ratio (LR) reflects the potentiality of a diagnostic platform and how likely a patient has a disease or condition. The higher the LR+, the more assurance that the person is a valid patient; the lower the LR-, the more confidence that the person does not retain the specific health problem. For the 4-TAA panel, the LR+ and LR- were 5.00 and 0.29, respectively, indicating that the positive result is likely to originate from patients with HCC, and the negative consequence is insufficient to rule out HCC.

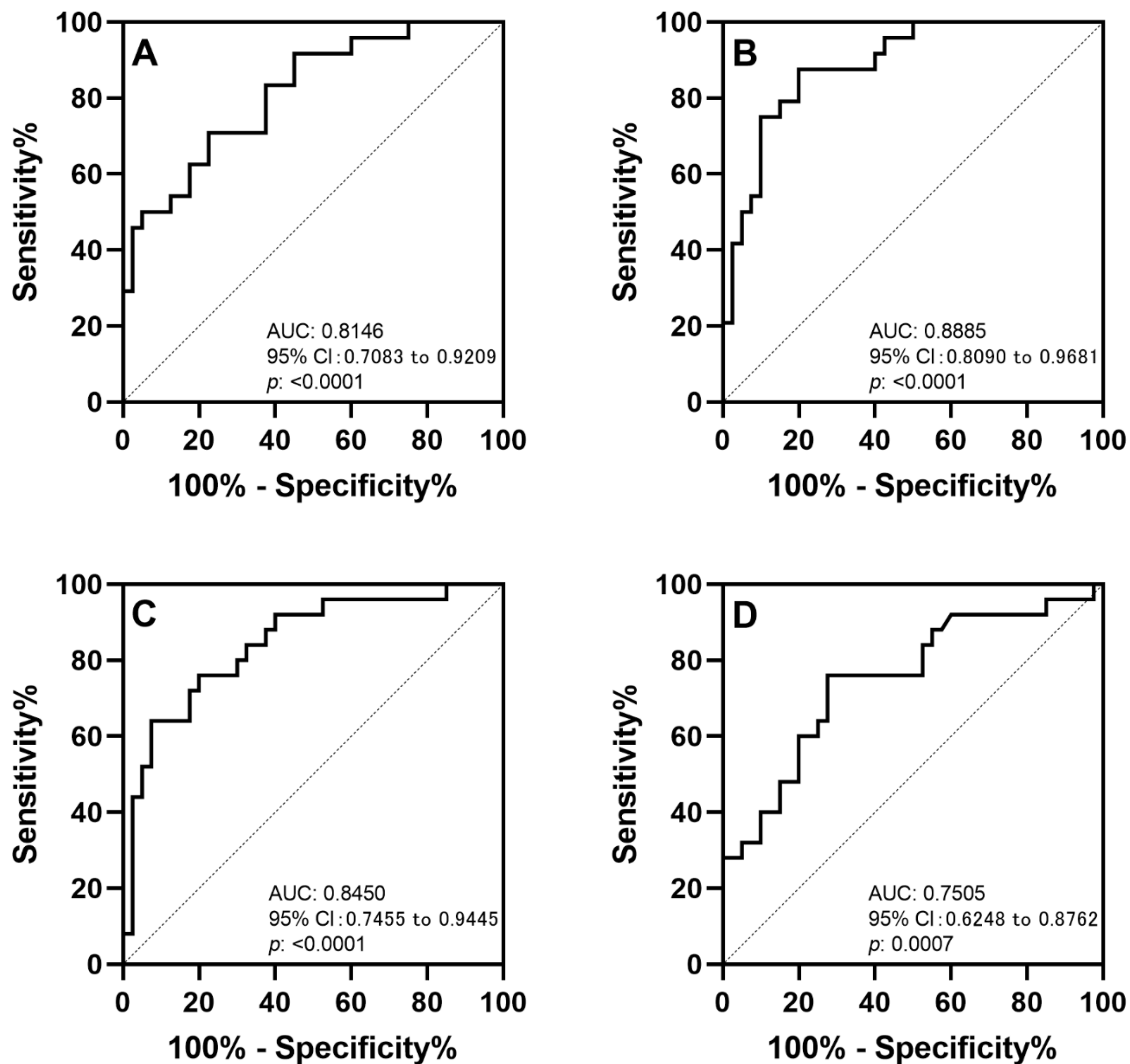


Figure 8. ROC analysis of anti-TAA autoantibodies in sera from patients with Hispanic HCC and normal healthy control individuals. (A): autoantibody to p16; (B): autoantibody to DNMT3A; (C): autoantibody to HSP60; (D): autoantibody to HSPA5. AUC: area under the receiver operating characteristic (ROC) curve.

Table 5. Sequential addition of antigens to the panel of 4 TAAs

Antigens		HCC (n=24)	NHS (n=40)	Se	Sp	FP	FN	PPV	NPV	LR+	LR-
1 TAA	DNMT3A	11 (45.8%)	2 (5%)	45.8%	95%	5%	54.2%	90.2%	63.7%	9.16	0.57
2 TAA	DNMT3A or P16	14 (58.3%)	3 (7.5%)	58.3%	92.5%	7.5%	41.7%	88.6%	68.9%	7.77	0.45
3 TAA	DNMT3A or P16 or HSP60	17 (70.8%)	4 (10%)	70.8%	90%	10%	29.2%	87.6%	75.5%	7.08	0.32
4 TAA	DNMT3A or P16 or HSP60 or HSPA5	18 (75%)	6 (15%)	75%	85%	15%	25%	83.3%	77.3%	5.00	0.29

HCC: Hepatocellular carcinoma; NHS: Negative healthy serum; Se: sensitivity = positive/ number of HCC; Sp: specificity = negative/ number of NHS; FP: false positive = 1 – Sp; FN: false negative = 1 – Se; PPV: positive predictive value = Se/(Se + FP); NPV: negative predictive value = Sp/(Sp + FN); LR+: positive likelihood ratio = Se/(1 – Sp); LR-: negative likelihood ratio = (1 - Se)/Sp.

Chapter 7: Validation of 4 anti-TAA autoantibodies in a separate Asian sera cohort

Overview

Different ethnic groups retain different pathogenesis properties in HCC development, attributing to various etiological risk factors. After identifying anti-TAA autoantibody biomarkers within Hispanic HCC groups, we need to verify the efficacy of those biomarkers in other groups, to see if the biomarkers are exclusive to the Hispanic HCC patients. In the present study, we aim to evaluate the titer and frequency of the individual autoantibodies in a separate sera cohort from Asian population. The Asian sera groups include HCC, LC, and CH, representing different stages of Asian HCC tumorigenesis. It is shown that the titer or frequency of anti- HSPA5 and p16 autoantibody have significant difference between the Asian HCC group and NHS group. The other autoantibodies did not show significant differences, suggesting that anti- DNMT3A and HSP60 autoantibodies could be diagnostic biomarkers exclusive to Hispanic HCC. HSPA5 and p16 shows potentiality as a biomarker in both Hispanic and Asian HCC groups.

Rationale

In the development of HCC, different risk factors and lifestyles could lead to various ways of tumorigenesis and precancerous conditions. Attributing to a high prevalence of metabolic disorders such as Diabetes Mellitus II and non-alcoholic liver fatty liver disease, the Hispanic population retains an increasing incidence and mortality rate of HCC [22, 94]. Unlike the Hispanic group, more than 50% of Asian HCC cases evolved from Hepatitis B virus infection in the United

States [95]. Those specific etiological risk factors could lead to distinct changes in either structure of tissues or the internal signaling pathways, producing altered proteins and autoantibodies. To test if the identified biomarkers are exclusive to Hispanic HCC, we compared the distribution of the 4 autoantibodies in Asian cohorts. The autoantibodies that are significantly expressed only in Hispanic HCC can be deemed as the exclusive biomarkers.

Experimental design

To keep the consistency with the previous part, we are trying to use ELISA to test the titer of each Asian sera cohort, including 53 cases of HCC sera, 20 cases of LC sera, and 44 cases of CH sera. In the experiment, we will simultaneously test the autoantibodies titers of Hispanic HCC sera and NHS sera and select Hispanic NHS sera as control groups. In this way can we reduce the bias to the lowest level, owing to the same control group that both cohorts share. The cutoff value will be defined as the mean value plus 2 times of standard deviation value of the NHS group, which is the same as the previous part.

Results

To validate the robustness of the 4 anti-TAA autoantibodies as exclusive Hispanic HCC biomarkers, we used another cohort containing 53 HCC, 20 LC, and 44 CH Asian sera samples, to see if those autoantibodies are significantly different. To maintain consistency, we follow the same NHS sera control group and the criteria for setting up cutoff value in the precious part. The frequencies of the anti-TAA autoantibodies are shown in Table 6. Among all the Asian HCC sera groups, only anti-HSPA5 autoantibody showed a significantly higher positive rate (28.3%) than

the NHS control group (p = 0.003). Concerning the other groups, the anti-p16 and anti-HSPA5 autoantibody showed greater expression for the LC groups, while no significant higher autoantibody frequency was detected in the CH groups. We next compare the OD values from the Asian HCC, LC, and CH groups with the NHS control groups using the non-parametrical Mann-Whitney test (Fig 8.). The results showed that anti-p16 and anti-HSPA5 showed significantly different OD values in HCC and LC groups compared to the control group. No significant differences were found between HCC and NHS groups in anti-DNMT3A or anti-HSP60 autoantibodies, suggesting that these 2 can be the exclusive diagnostic biomarkers for Hispanic HCC.

Table 6. Frequency of the 4 anti-TAA autoantibodies in Asian sera cohort

Antigens	Asian HCC (n=53)	Asian LC (n=20)	Asian CH (n=44)	NHS (n=40)
DNMT3A	5 (9.4%)	5 (25%)	3 (6.8%)	2 (5%)
P16	9 (17.0%)	7 (35.0%) *	4 (9.1%)	3 (7.5%)
HSPA5	15 (28.3%) **	8 (40%) ***	4 (9.1%)	1 (2.5%)
HSP60	6 (11.3%)	3 (15%)	3 (6.8%)	2 (5%)

HCC: hepatocellular carcinoma; LC: liver cirrhosis; CH: chronic hepatitis; NHS: normal healthy sera. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, cutoff value = mean of 40 NHS + 2 SD.

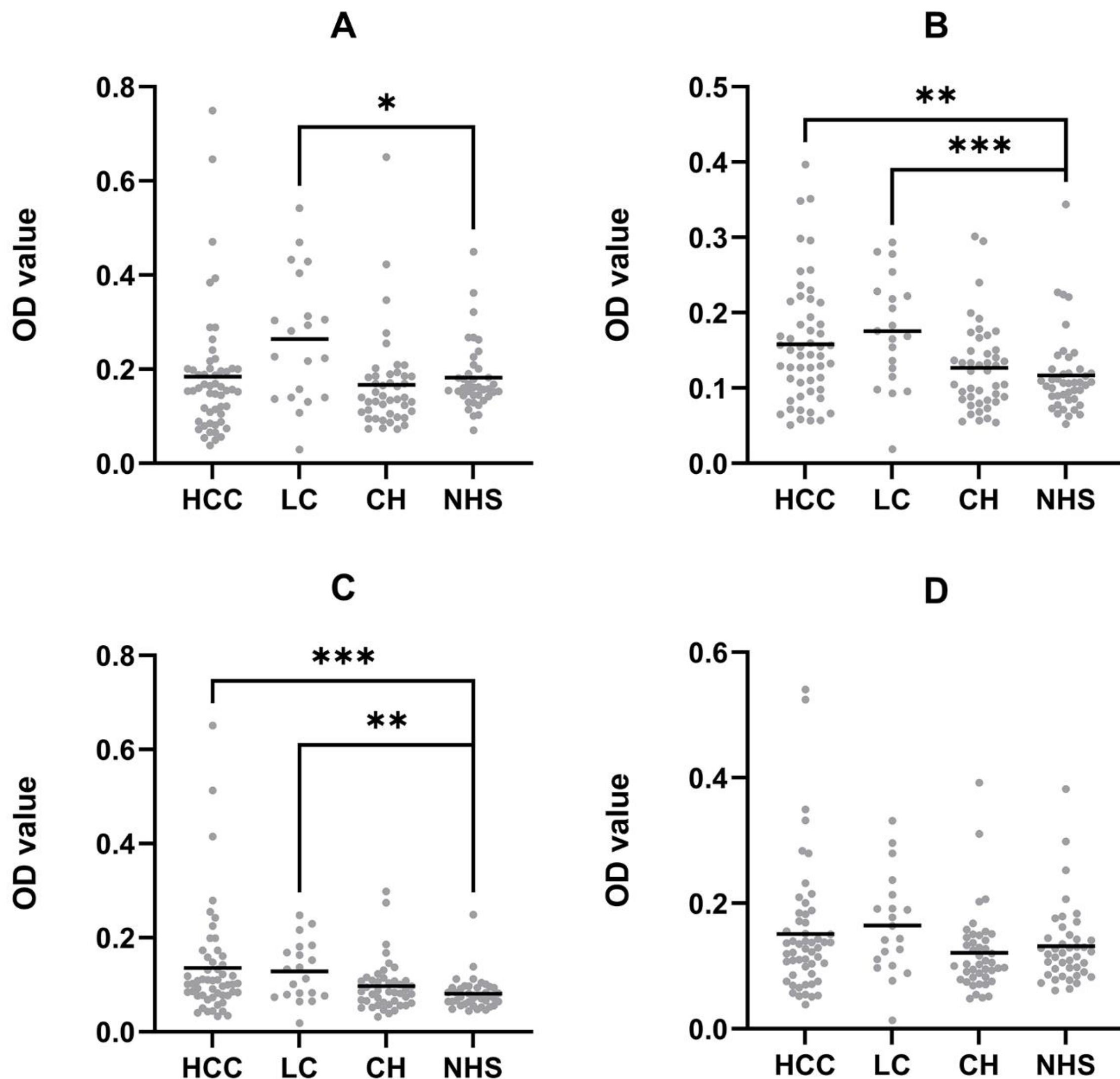


Figure 9. Autoantibody titers of the 4 TAAs in the Asian sera cohort. The autoantibody titers of the 4 TAAs are shown in the Asian HCC, LC, and CH sera groups and NHS control sera group. Mann-Whitney tests show significant differences in anti-p16 and anti-HSPA5 autoantibodies titers when comparing the HCC and LC groups with the NHS control group. Anti-DNMT3A autoantibody shows a significantly higher titer than NHS controls. A: DNMT3A; B: p16; C: HSPA5: D: HSP60. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Chapter 8: Discussion

Numerous studies have noted the importance of cancer biomarkers in the early diagnosis. TAAs being efficient and robust biomarkers have been applied in the HCC diagnosis for decades. Nonetheless, there is no research focused on Hispanic HCC patients. In the United States, the highest incidence rate and lowest overall survival rate of HCC were observed in the Hispanic population compared to other ethnic groups [23, 24]. According to previous research, non-alcoholic fatty liver disease (NAFLD) and other metabolic risk factors such as obesity and diabetes might be the major etiological cause of Hispanic HCC [23, 30]. The distinct pathogenesis of Hispanic HCC can lead to altered protein, and antigen epitopes expression, therefore may produce unique autoantibodies exclusive to the population. This study is the first one that investigated the anti-TAA autoantibody biomarkers catering to the diagnosis of the Hispanic HCC group. To identify specific TAA targets with high autoantibody intensity in Hispanic HCC, an immunoproteomic approach was used in our study to discover the differentially expressed antigens in Hispanic HCC sera compared to the Hispanic normal healthy control sera. Such an approach was named serological proteome analysis (SERPA) and has been successfully applied in our lab to discover novel biomarkers within different malignancies. In the 2DE part of the SERPA analysis, we used not one but two other HCC cell lines, including HepG2 and SNU449, compared to previous SERPA studies, which only used HepG2 cell line [63, 83]. Combining the two cell lines will include both well differentiated and poorly differentiated HCC, creating a more integrated landscape of HCC proteome and metabolic pattern [85, 86]. Further mass spectrometry analysis will identify the differentially expressed antigens on the SDS-PAGE gel. Meanwhile, we are trying to enlarge the potential target pool by screening differentially expressed driver genes in HCC.

Driver genes refer to the mutational genes which have a strong linkage with tumorigenesis and cancer progression [70]. Previous research has suggested that a new antigenic epitope of TAAs could be produced via somatic mutations, therefore being detected by the immune system and bound by autoantibody in both mutated and non-mutated ways [45, 68]. To investigate the linkage between driver gene and Hispanic HCC TAA, we are aiming to filter proper TAA candidates from the driver gene database IntoGen [90]. With the acknowledgment that differentially expressed genes in the tumor tissue and adjacent normal tissue can be considered as proper TAA candidates, we verified all the 75 driver genes from the IntoGen database via the Oncomine database together with the TCGA database and found another 10 candidates, thus providing a more integrated scenery for novel TAA discovery. Next, we did the enzyme-linked immunoassay (ELISA) to test the autoantibody frequency in 4 sera groups originating from the Hispanic population, including HCC, liver cirrhosis (LC), chronic hepatitis (CH), and normal healthy control (NHS). Throughout all 19 potential anti-TAA autoantibodies, 4 of them showed a significant difference between the HCC sera group and NHS sera group in the positive frequency, indicating those 4 autoantibodies can be novel biomarkers for Hispanic HCC diagnosis. Interestingly, many targets showed remarkable high autoantibody frequency in the LC group, while few showed significant frequency in the CH group. We were expecting a continuous increase of the specific autoantibodies as the illness progresses. However, the LC group but not the HCC group retains the most significant number of positive autoantibody biomarkers. This can be attributed to several reasons. First is the low level of the sera sample number. In the present study, only 24 HCC sera and 20 LC sera were used, which can not represent a larger scale of the autoantibodies distribution in the Hispanic population. Secondly, liver cirrhosis represents the closest pre-cancer stage of HCC tumorigenesis. The emergence of the tumor may have existed in the LC patients without the exact diagnosis of

HCC. Last but not least, this phenomenon might result from the distinct feature of Hispanic HCC etiology. Multiple studies have suggested that the Hispanic population gains a higher prevalence rate of non-alcoholic fatty liver disease (NAFLD) than non-Hispanic white and African American populations [96-99]. It is noted that it is more prone to see Mallory body and advanced fibrosis in Hispanic people than non-Hispanic white and African American people [98]. Hispanic NAFLD patients also show a high level of aminotransferases, suggesting a great inclination to have inflammation progressions such as fibrosis and cirrhosis [100]. More research is needed on the biomarker of Hispanic liver cirrhosis and the transition of Hispanic liver cirrhosis to HCC.

Among all the 4 TAAs we identified, anti-DNMT3A autoantibody retains the highest sensitivity. As a family member of DNA methyltransferases, DNMT3A can catalyze de novo DNA methylation in CpG islands of the gene promoter [101], resulting in the silencing of the tumor suppressor genes in multiple cancers. It is suggested that DNMT3A has the potential to be an HCC biomarker, Hassona et al. have indicated an elevated mRNA expression of DNMT3A in HCC patients and retaining an AUC of 0.958 in discriminating HCC from hepatitis cirrhotic patients [102]. Yan et al. suggest that DNMT3A inhibits Tyrosine-protein phosphatase nonreceptor type 13 (PTPN13) by binding onto its gene promoter and promotes the development of HBV-related HCC [103]. Upregulated DNMT3A was also shown to be strongly correlated with non-coding RNA alteration in HCC development, including lncRNA HCP5, miR-137, and LncRNA34a via multiple signaling pathways [104-106]. In addition, DNMT3A-mediated epigenetic silencing of PTEN was found in NAFLD-induced HCC, hinting the linkage between DNMT3A expression and the specific etiology of Hispanic HCC [107]. P16, the product of the CDKN2A gene, is another TAA biomarker we identified from the Hispanic HCC group. Studies have investigated that the methylation of the CDKN2A gene and the dysregulation of p16 expression could lead to the HCC

progression [108, 109]. It was believed that homozygous deletion of the CDKN2A gene would result in the inactivation of the G1 to S phase cell cycle and thus contribute to HCC tumorigenesis [110]. The capability of p16 being an HCC biomarker has been reported in many studies. Zhang et al. and Koziol et al. reported p16 as a member of the TAA microarrays, which have great performance in the early diagnosis of HCC [57, 65]. Wang et al. also reported an AUC value of 0.62 for circulating anti-p16 antibodies in the plasma of HCC patients, validating its diagnostic value in clinics [111]. Not only those 2 biomarkers but the heat shock protein (HSP) family has also been proved to correlate with multiple malignancies over the years. Known to function as molecular chaperones, HSP60 and HSPA5 assist in folding newly synthesized and misfolded proteins [112]. Our lab has found HSP60 and HSPA5 as valid TAA biomarkers in the diagnosis of HCC in multiple studies [63, 64]. Previous research also verified the efficacy of those two proteins as HCC diagnostic biomarkers with different sensitivity and specificity rates [113, 114]. Interestingly, hepatocyte ER stress, which is regulated by HSPA5, has been suggested to play an important role in the development of steatohepatitic HCC [115], which accounts for a great proportion of the Hispanic HCC. By combing all 4 anti-TAA autoantibodies, the sensitivity of the biomarker panel in the present study reached 75%, which was drastically elevated compared to the single autoantibodies, showing the effectiveness of the biomarker panel in future clinical application.

In the final part of the study, we were trying to assess the exclusiveness of the biomarkers to Hispanic HCC. A separate cohort of Asian sera was included as a counterpart to the Hispanic cohort. It is shown that although anti-p16 and HSPA5 autoantibodies showed significantly higher frequencies in the LC group, but only HSPA5 retains higher frequency in the HCC group, compared to the NHS control group. The comparison of antibody titer in Asian HCC and NHS

groups showed that anti-p16 and anti-HSPA5 autoantibody is significantly higher in HCC. By combining the results of antibody titers and frequencies, it suggests that within the 4 Hispanic HCC biomarkers, DNMT3A and HSP60 can be exclusive to Hispanic HCC, while p16 and HSPA5 showed efficacy in Asian HCC diagnosis. The underlying reason could be the etiological variances between different ethnic groups.

There remain some limitations within the present study. First, the sample number of the sera is not enough. In the current study, we used 110 sera samples in the Hispanic and 157 in the Asian cohorts, which is limited for research regarding an ethnic population. To further test the validity of the TAAs, a larger sera cohort is needed, which could reduce the bias and include a more integrated Hispanic HCC proteome. Another weakness of the current research is that it does not verify the 4 TAAs in ethnic populations other than Asian, for instance, Caucasian white, and African American. Like the Hispanic group, the African American harbors a high prevalence of metabolic disorders and a high incidence of HCC [23]. By comparing the Hispanics to other populations, the TAA biomarker exclusive to the Hispanic HCC will be investigated, and a complete biomarker disparity will be shown for future study.

References

1. Organization, W.H., *International agency for research on cancer.* 2019.
2. Llovet, J.M., et al., *Hepatocellular carcinoma.* Nature reviews Disease primers, 2021. **7**(1): p. 1-28.
3. Park, J.W., et al., *Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE Study.* Liver International, 2015. **35**(9): p. 2155-2166.
4. Yang, J.D., et al., *A global view of hepatocellular carcinoma: trends, risk, prevention and management.* Nature reviews Gastroenterology & hepatology, 2019. **16**(10): p. 589-604.
5. Altekruse, S.F., et al., *Changing hepatocellular carcinoma incidence and liver cancer mortality rates in the United States.* The American journal of gastroenterology, 2014. **109**(4): p. 542.
6. Marrero, J.A., et al., *Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases.* Clinical Liver Disease, 2019. **13**(1): p. 1.
7. Chidambaranathan-Reghupaty, S., P.B. Fisher, and D. Sarkar, *Hepatocellular carcinoma (HCC): epidemiology, etiology and molecular classification.* Advances in cancer research, 2021. **149**: p. 1-61.
8. Akinyemiju, T., et al., *The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015.* JAMA oncology, 2017. **3**(12): p. 1683-1691.
9. Liang, T.J., *Hepatitis B: the virus and disease.* Hepatology, 2009. **49**(S5): p. S13-S21.
10. Bukh, J., *The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control.* Journal of hepatology, 2016. **65**(1): p. S2-S21.

11. Ganne-Carrié, N. and P. Nahon, *Hepatocellular carcinoma in the setting of alcohol-related liver disease.* Journal of hepatology, 2019. **70**(2): p. 284-293.

12. Turati, F., et al., *Alcohol and liver cancer: a systematic review and meta-analysis of prospective studies.* Annals of oncology, 2014. **25**(8): p. 1526-1535.

13. Stepanova, M., et al., *Direct and indirect economic burden of chronic liver disease in the United States.* Clinical Gastroenterology and Hepatology, 2017. **15**(5): p. 759-766. e5.

14. Hester, C.A., et al., *Comparative analysis of nonalcoholic steatohepatitis–versus viral hepatitis–and alcohol-related liver disease–related hepatocellular carcinoma.* Journal of the National Comprehensive Cancer Network, 2019. **17**(4): p. 322-329.

15. Anstee, Q.M., et al., *From NASH to HCC: current concepts and future challenges.* Nature reviews Gastroenterology & hepatology, 2019. **16**(7): p. 411-428.

16. Mittal, S., et al., *Hepatocellular carcinoma in the absence of cirrhosis in United States veterans is associated with nonalcoholic fatty liver disease.* Clinical Gastroenterology and Hepatology, 2016. **14**(1): p. 124-131. e1.

17. Wen, N., et al., *The clinical management of hepatocellular carcinoma worldwide: A concise review and comparison of current guidelines: 2022 update.* BioScience Trends, 2022. **16**(1): p. 20-30.

18. Berzigotti, A., et al., *Novel ultrasound-based methods to assess liver disease: The game has just begun.* Digestive and liver disease, 2018. **50**(2): p. 107-112.

19. Roberts, L.R., et al., *Imaging for the diagnosis of hepatocellular carcinoma: a systematic review and meta-analysis.* Hepatology, 2018. **67**(1): p. 401-421.

20. Neoplasia, I.C.G.f.H., *Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia.* Hepatology, 2009. **49**(2): p. 658-664.

21. Zhou, J., et al., *Guidelines for the diagnosis and treatment of hepatocellular carcinoma (2019 Edition).* Liver Cancer, 2020. **9**(6): p. 682-720.

22. Pinheiro, P.S., et al., *Cancer mortality in Hispanic ethnic groups.* Cancer Epidemiology and Prevention Biomarkers, 2017. **26**(3): p. 376-382.

23. Venepalli, N.K., et al., *Features of hepatocellular carcinoma in Hispanics differ from African Americans and non-Hispanic Whites.* World Journal of Hepatology, 2017. **9**(7): p. 391.

24. Kulik, L. and H.B. El-Serag, *Epidemiology and management of hepatocellular carcinoma.* Gastroenterology, 2019. **156**(2): p. 477-491. e1.

25. Davila, J.A. and H.B. El–Serag, *Racial differences in survival of hepatocellular carcinoma in the United States: a population-based study.* Clinical Gastroenterology and Hepatology, 2006. **4**(1): p. 104-110.

26. Ramirez, A.G., et al., *Incidence of hepatocellular carcinoma in Texas Latinos, 1995–2010: an update.* PloS one, 2014. **9**(6): p. e99365.

27. Setiawan, V.W., et al., *Disparity in liver cancer incidence and chronic liver disease mortality by nativity in H ispanics: The M ultiethnic C ohort.* Cancer, 2016. **122**(9): p. 1444-1452.

28. Kuniholm, M.H., et al., *Prevalence of hepatitis C virus infection in US Hispanic/Latino adults: results from the NHANES 2007–2010 and HCHS/SOL studies.* The Journal of infectious diseases, 2014. **209**(10): p. 1585-1590.

29. El-Serag, H.B., et al., *Racial Differences in the Progression to Cirrhosis and Hepatocellular Carcinoma in HCV-Infected Veterans: Effect of race on cirrhosis and HCC risk.* Official journal of the American College of Gastroenterology| ACG, 2014. **109**(9): p. 1427-1435.

30. Wong, A., et al., *Higher risk of hepatocellular carcinoma in Hispanic patients with hepatitis C cirrhosis and metabolic risk factors.* Scientific reports, 2018. **8**(1): p. 1-8.

31. Saab, S., et al., *Nonalcoholic fatty liver disease in Latinos.* Clinical Gastroenterology and Hepatology, 2016. **14**(1): p. 5-12.

32. Ascha, M.S., et al., *The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis.* Hepatology, 2010. **51**(6): p. 1972-1978.

33. O'Connor, S., et al., *Hepatocellular carcinoma-United States, 2001-2006.* Morbidity and Mortality Weekly Report, 2010. **59**(17): p. 517-520.

34. Svegliati-Baroni, G., et al., *Insulin and insulin-like growth factor-1 stimulate proliferation and type I collagen accumulation by human hepatic stellate cells: differential effects on signal transduction pathways.* Hepatology, 1999. **29**(6): p. 1743-1751.

35. Anstee, Q.M. and C.P. Day, *The genetics of NAFLD.* Nature reviews Gastroenterology & hepatology, 2013. **10**(11): p. 645-655.

36. Ogden, C.L., et al., *Prevalence of childhood and adult obesity in the United States, 2011-2012.* Jama, 2014. **311**(8): p. 806-814.

37. Pan, J.-J., et al., *Prevalence of metabolic syndrome and risks of abnormal serum alanine aminotransferase in Hispanics: a population-based study.* PloS one, 2011. **6**(6): p. e21515.

38. Dawson, D.A., et al., *Changes in alcohol consumption: United States, 2001–2002 to 2012–2013.* Drug and alcohol dependence, 2015. **148**: p. 56-61.

39. Vaeth, P.A., M. Wang-Schweig, and R. Caetano, *Drinking, alcohol use disorder, and treatment access and utilization among US racial/ethnic groups.* Alcoholism: Clinical and Experimental Research, 2017. **41**(1): p. 6-19.

40. Schwimmer, J.B., et al., *Heritability of nonalcoholic fatty liver disease.* Gastroenterology, 2009. **136**(5): p. 1585-1592.

41. Romeo, S., et al., *Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease.* Nature genetics, 2008. **40**(12): p. 1461-1465.

42. Jiao, J., et al., *Prevalence of aflatoxin-associated TP53R249S mutation in hepatocellular carcinoma in Hispanics in South Texas.* Cancer Prevention Research, 2018. **11**(2): p. 103-112.

43. Mathur, A.K., et al., *Racial/ethnic disparities in access to care and survival for patients with early-stage hepatocellular carcinoma.* Archives of surgery, 2010. **145**(12): p. 1158-1163.

44. Baldwin, R., *Tumour-specific immunity against spontneous rat tumours.* International journal of cancer, 1966. **1**(3): p. 257-264.

45. Macdonald, I.K., C.B. Parsy-Kowalska, and C.J. Chapman, *Autoantibodies: opportunities for early cancer detection.* Trends in cancer, 2017. **3**(3): p. 198-213.

46. Dunn, G.P., et al., *Cancer immunoediting: from immunosurveillance to tumor escape.* Nature immunology, 2002. **3**(11): p. 991-998.

47. Dunn, G.P., L.J. Old, and R.D. Schreiber, *The three Es of cancer immunoediting.* Annu. Rev. Immunol., 2004. **22**: p. 329-360.

48. McGilvray, R.W., et al., *NKG2D ligand expression in human colorectal cancer reveals associations with prognosis and evidence for immunoediting.* Clinical Cancer Research, 2009. **15**(22): p. 6993-7002.

49. Waldhauer, I. and A. Steinle, *NK cells and cancer immunosurveillance.* Oncogene, 2008. **27**(45): p. 5932-5943.

50. Yigit, R., et al., *Ovarian cancer creates a suppressive microenvironment to escape immune elimination.* Gynecologic oncology, 2010. **117**(2): p. 366-372.

51. Gangoda, L., et al., *Extracellular vesicles including exosomes are mediators of signal transduction: are they protective or pathogenic?* Proteomics, 2015. **15**(2-3): p. 260-271.

52. Tan, E.M. and J. Zhang, *Autoantibodies to tumor-associated antigens: reporters from the immune system.* Immunological reviews, 2008. **222**(1): p. 328-340.

53. Hanash, S., *Harnessing immunity for cancer marker discovery.* Nature biotechnology, 2003. **21**(1): p. 37-38.

54. Heo, C.-K., Y.Y. Bahk, and E.-W. Cho, *Tumor-associated autoantibodies as diagnostic and prognostic biomarkers.* BMB reports, 2012. **45**(12): p. 677.

55. Giraud, J., et al., *Hepatocellular carcinoma immune landscape and the potential of immunotherapies.* Frontiers in immunology, 2021. **12**: p. 655697.

56. Benechet, A.P. and M. Iannacone, *Determinants of hepatic effector CD8+ T cell dynamics.* Journal of hepatology, 2017. **66**(1): p. 228-233.

57. Koziol, J.A., et al., *Early detection of hepatocellular carcinoma using autoantibody profiles from a panel of tumor-associated antigens.* Cancer Immunology, Immunotherapy, 2018. **67**(5): p. 835-841.

58. Dai, L., et al., *Using immunomic approach to enhance tumor-associated autoantibody detection in diagnosis of hepatocellular carcinoma.* Clinical Immunology, 2014. **152**(1-2): p. 127-139.

59. Liu, M., et al., *Autoantibody response to murine double minute 2 protein in immunodiagnosis of hepatocellular carcinoma.* Journal of Immunology Research, 2014. **2014**.

60. Wang, K., et al., *Immunogenicity of Ra1A and its tissue-specific expression in hepatocellular carcinoma.* International journal of immunopathology and pharmacology, 2009. **22**(3): p. 735-743.

61. Zhang, J.-Y., et al., *A novel cytoplasmic protein with RNA-binding motifs is an autoantigen in human hepatocellular carcinoma.* Journal of Experimental Medicine, 1999. **189**(7): p. 1101-1110.

62. Xing, M., et al., *Overexpression of p62/IMP2 can promote cell migration in hepatocellular carcinoma via activation of the Wnt/β-catenin pathway.* Cancers, 2019. **12**(1): p. 7.

63. Looi, K.S., et al., *Using proteomic approach to identify tumor-associated antigens as markers in hepatocellular carcinoma.* Journal of proteome research, 2008. **7**(9): p. 4004-4012.

64. Shao, Q., et al., *Autoantibodies against glucose-regulated protein 78 as serological diagnostic biomarkers in hepatocellular carcinoma.* International journal of oncology, 2012. **41**(3): p. 1061-1067.

65. Zhang, J.-Y., et al., *Antibody detection using tumor-associated antigen mini-array in immunodiagnosing human hepatocellular carcinoma.* Journal of hepatology, 2007. **46**(1): p. 107-114.

66. Torchilin, V.P., L.Z. Iakoubov, and Z. Estrov, *Antinuclear autoantibodies as potential antineoplastic agents.* Trends in immunology, 2001. **22**(8): p. 424-427.

67. Zhang, J.-Y. and E.M. Tan, *Autoantibodies to tumor-associated antigens as diagnostic biomarkers in hepatocellular carcinoma and other solid tumors.* Expert review of molecular diagnostics, 2010. **10**(3): p. 321-328.

68. Zaenker, P., E.S. Gray, and M.R. Ziman, *Autoantibody production in cancer—the humoral immune response toward autologous antigens in cancer patients.* Autoimmunity reviews, 2016. **15**(5): p. 477-483.

69. Reuschenbach, M., M. von Knebel Doeberitz, and N. Wentzensen, *A systematic review of humoral immune responses against tumor antigens.* Cancer immunology, immunotherapy, 2009. **58**(10): p. 1535-1544.

70. Waks, Z., et al., *Driver gene classification reveals a substantial overrepresentation of tumor suppressors among very large chromatin-regulating proteins.* Scientific reports, 2016. **6**(1): p. 1-12.

71. Joseph, C.G., et al., *Association of the autoimmune disease scleroderma with an immunologic response to cancer.* Science, 2014. **343**(6167): p. 152-157.

72. James, J.A. and J.B. Harley, *B-cell epitope spreading in autoimmunity.* Immunological reviews, 1998. **164**(1): p. 185-200.

73. Sahin, U., et al., *Human neoplasms elicit multiple specific immune responses in the autologous host.* Proceedings of the national academy of sciences, 1995. **92**(25): p. 11810-11813.

74. Hardouin, J., et al., *Usefulness of autoantigens depletion to detect autoantibody signatures by multiple affinity protein profiling.* Journal of separation science, 2007. **30**(3): p. 352-358.

75. Caron, M., G. Choquet-Kastylevsky, and R. Joubert-Caron, *Cancer immunomics using autoantibody signatures for biomarker discovery.* Molecular & Cellular Proteomics, 2007. **6**(7): p. 1115-1122.

76. Kobayashi, M., et al. *Development of autoantibody signatures for common cancers*. in *Seminars in immunology*. 2020. Elsevier.

77. Tan, H.T., et al., *Serum autoantibodies as biomarkers for early cancer detection.* The FEBS journal, 2009. **276**(23): p. 6880-6904.

78. Magdeldin, S., et al., *Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis.* Clinical proteomics, 2014. **11**(1): p. 1-10.

79. Anderson, K.S. and J. LaBaer, *The sentinel within: exploiting the immune system for cancer biomarkers.* Journal of proteome research, 2005. **4**(4): p. 1123-1133.

80. White, D.L., et al., *Incidence of hepatocellular carcinoma in all 50 United States, from 2000 through 2012.* Gastroenterology, 2017. **152**(4): p. 812-820. e5.

81. Kim, H.-s. and H.B. El-Serag, *The epidemiology of hepatocellular carcinoma in the USA.* Current gastroenterology reports, 2019. **21**(4): p. 1-8.

82. Zhang, J., et al., *De-novo humoral immune responses to cancer-associated autoantigens during transition from chronic liver disease to hepatocellular carcinoma.* Clinical & Experimental Immunology, 2001. **125**(1): p. 3-9.

83. Peng, B., et al., *Using immunoproteomics to identify alpha-enolase as an autoantigen in liver fibrosis.* Journal of proteome research, 2013. **12**(4): p. 1789-1796.

84. Dai, L., et al., *Serological proteome analysis approach-based identification of ENO1 as a tumor-associated antigen and its autoantibody could enhance the sensitivity of CEA and CYFRA 21-1 in the detection of non-small cell lung cancer.* Oncotarget, 2017. **8**(22): p. 36664.

85. Qiu, Z., et al., *Hepatocellular carcinoma cell lines retain the genomic and transcriptomic landscapes of primary human cancers.* Scientific Reports, 2016. **6**(1): p. 1-13.

86. Nwosu, Z.C., et al., *Liver cancer cell lines distinctly mimic the metabolic gene expression pattern of the corresponding human tumours.* Journal of Experimental & Clinical Cancer Research, 2018. **37**(1): p. 1-15.

87. Harris, K.L., et al., *Rationale and roadmap for developing panels of hotspot cancer driver gene mutations as biomarkers of cancer risk.* Environmental and molecular mutagenesis, 2020. **61**(1): p. 152-175.

88. Liu, W., Y. Peng, and D.J. Tobin, *A new 12-gene diagnostic biomarker signature of melanoma revealed by integrated microarray analysis.* PeerJ, 2013. **1**: p. e49.

89. Canelle, L., et al., *An efficient proteomics-based approach for the screening of autoantibodies.* Journal of immunological methods, 2005. **299**(1-2): p. 77-89.

90. Martínez-Jiménez, F., et al., *A compendium of mutational cancer driver genes.* Nature Reviews Cancer, 2020. **20**(10): p. 555-572.

91. Houen, G., *Autoantibodies as diagnostic tools*, in *Autoantibodies*. 2019, Springer. p. 1-11.

92. Okada, R., et al., *Six autoantibodies as potential serum biomarkers of hepatocellular carcinoma: A prospective multicenter study.* International Journal of Cancer, 2020. **147**(9): p. 2578-2586.

93. Hoshino, I., et al., *Panel of autoantibodies against multiple tumor-associated antigens for detecting gastric cancer.* Cancer science, 2017. **108**(3): p. 308-315.

94. Ha, J., et al., *Burden of hepatocellular carcinoma among hispanics in South Texas: a systematic review.* Biomarker Research, 2017. **5**(1): p. 1-6.

95. Gordon, S.C., et al., *Antiviral therapy for chronic hepatitis B virus infection and development of hepatocellular carcinoma in a US population.* Clinical Gastroenterology and Hepatology, 2014. **12**(5): p. 885-893.

96. Weston, S.R., et al., *Racial and ethnic distribution of nonalcoholic fatty liver in persons with newly diagnosed chronic liver disease.* Hepatology, 2005. **41**(2): p. 372-379.

97. Clark, J.M., *The epidemiology of nonalcoholic fatty liver disease in adults.* Journal of clinical gastroenterology, 2006. **40**: p. S5-S10.

98. Mohanty, S.R., et al., *Influence of ethnicity on histological differences in non-alcoholic fatty liver disease.* Journal of hepatology, 2009. **50**(4): p. 797-804.

99. Browning, J.D., et al., *Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity.* Hepatology, 2004. **40**(6): p. 1387-1395.

100. Alazmi, W.M., et al., *Predictors of cirrhosis in Hispanic patients with nonalcoholic steatohepatitis.* Digestive diseases and sciences, 2006. **51**(10): p. 1725-1729.

101. Wong, K.K., C.H. Lawrie, and T.M. Green, *Oncogenic roles and inhibitors of DNMT1, DNMT3A, and DNMT3B in acute myeloid leukaemia.* Biomarker insights, 2019. **14**: p. 1177271919846454.

102. Hassona, M.M., et al., *DNA methyltransferases as potential biomarkers for HCV related hepatocellular carcinoma.* Asian Pacific Journal of Cancer Prevention: APJCP, 2020. **21**(11): p. 3357.

103. Yan, Y., et al., *Anti-oncogene PTPN13 inactivation by hepatitis B virus X protein counteracts IGF2BP1 to promote hepatocellular carcinoma progression.* Oncogene, 2021. **40**(1): p. 28-45.

104. Zhou, Y., et al., *Long non-coding RNA HCP5 functions as a sponge of miR-29b-3p and promotes cell growth and metastasis in hepatocellular carcinoma through upregulating DNMT3A.* Aging (Albany NY), 2021. **13**(12): p. 16267.

105. Wang, J., et al., *Upregulation of miR-137 Expression Suppresses Tumor Growth and Progression via Interacting with DNMT3a Through Inhibiting the PTEN/Akt Signaling in HCC.* OncoTargets and therapy, 2021. **14**: p. 165.

106. Zhang, L., et al., *The molecular mechanism of LncRNA34a-mediated regulation of bone metastasis in hepatocellular carcinoma.* Molecular cancer, 2019. **18**(1): p. 1-15.

107. Liu, D., et al., *Squalene epoxidase drives NAFLD-induced hepatocellular carcinoma and is a pharmaceutical target.* Science translational medicine, 2018. **10**(437): p. eaap9840.

108. Csepregi, A., et al., *Promoter methylation of CDKN2A and lack of p16 expression characterize patients with hepatocellular carcinoma.* BMC cancer, 2010. **10**(1): p. 1-12.

109. Zhou, Y., et al., *CDKN2A promoter methylation and hepatocellular carcinoma risk: A meta-analysis.* Clinics and research in hepatology and gastroenterology, 2018. **42**(6): p. 529-541.

110. Zucman-Rossi, J., et al., *Genetic landscape and biomarkers of hepatocellular carcinoma.* Gastroenterology, 2015. **149**(5): p. 1226-1239. e4.

111. Wang, J., et al., *Further study of circulating antibodies to P16, CD25 and FOXP3 in hepatocellular carcinoma.* OncoTargets and therapy, 2019. **12**: p. 10487.

112. Hoter, A., S. Rizk, and H.Y. Naim, *Heat shock protein 60 in hepatocellular carcinoma: Insights and perspectives.* Frontiers in molecular biosciences, 2020. **7**: p. 60.

113. Liu, H., et al., *Screening of autoantibodies as potential biomarkers for hepatocellular carcinoma by using T7 phase display system.* Cancer epidemiology, 2012. **36**(1): p. 82-88.

114. Hong, Y., et al., *An analysis of immunoreactive signatures in early stage hepatocellular carcinoma.* EBioMedicine, 2015. **2**(5): p. 438-446.

115. Nakagawa, H., et al., *ER stress cooperates with hypernutrition to trigger TNF-dependent spontaneous HCC development.* Cancer cell, 2014. **26**(3): p. 331-343.

Abbreviation

1-D- One-dimensional gel electrophoresis

2-DE- Two-dimensional gel electrophoresis ABC- ATP-binding cassette

ACTG1- Actin, cytoplasmic 2

AUC- Area under the curve

CALR- Calreticulin

CDKN2A- Cyclin Dependent Kinase Inhibitor 2A

CH- Chronic hepatitis

CI- Confidence interval

DAPI- 4’,6’-diamidino-2-phenylindole

DHX9- DExH-Box Helicase 9

DNMT3A- DNA Methyltransferase 3 Alpha

DTT- Dithiothreitol

ECL- Enhanced chemiluminescence

ELISA- Enzyme linked immunosorbent assay

ENO1- Alpha-enolase

GMPS- Guanine Monophosphate Synthase

GNAS- Guanine Nucleotide Binding Protein Subunit Alpha S

GO- Gene ontology

HCC- Hepatocellular carcinoma

His- Hispanic

HSP60- Heat shock 60 kDa protein

HSP70- Heat shock 70 kDa protein

HSPA5- Endoplasmic reticulum chaperone BiP

IEF- Isoelectric focusing

IIF- Indirect immunofluorescence

LC- Liver cirrhosis

MAPK1- Mitogen-Activated Protein Kinase 1

MS- Mass spectrometry

NF2- Moesin-Ezrin-Radixin Like Tumor Suppressor

NRAS- Neuroblastoma RAS Viral Oncogene Homolog

NHS- Normal human sera

P4HB- Protein disulfide isomerase

PBS- Phosphate buffered saline

PBST- Phosphate buffered saline containing 0.05% Tween-20

PR- Positive rate

ROC- Receiver operating characteristic

RT- Room temperature

SDS-PAGE- SDS polyacrylamide

SERPA- Serological proteome analysis

SETDB1- SET Domain Bifurcated Histone Lysine Methyltransferase 1

SMARCA4- SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4

TAAs- Tumor-associated antigens

TPI- Triosephosphate isomerase

TPM3- Tropomyosin alpha-3 chain

Curriculum Vita

Yangcheng Ma completed his bachelor's degree in clinical medicine (equivalent to M.D.) from the Zhengzhou University in 2016 and completed his master's degree in clinical medicine-surgery majoring sports medicine from the Zhengzhou University in 2018. From 2018 he joined the doctoral program in biological sciences at the University of Texas at El Paso.

He has worked in the research field of cancer biomarkers in Dr. Jianying Zhang's lab. He has one first-author manuscript under reviewing and two co-author papers published. He attended several academic conferences: American Association for Cancer Research (AACR) in 2022, and Border Biomedical Research Center (BBRC) in 2020, and published abstracts for the conferences. He has also gained summer research funding provided by the University of Texas in El Paso in 2021. From 2018 to 2022 he has been teaching assistant in the biological department and has accumulated rich experience in teaching.

After graduation he will dedicate himself into biomarker research in the future path. His long-term goal is to put effort into clinical and translational research.

Contact email address:

yma3@miners.utep.edu

mayc921226@gmail.com