**3. Research Strategy**

**Overall Significance**

Our application uses mobile health (mHealth), lab-on-a-chip (LOC ), biosensors, new portable ultrasound imaging technologies, and nanopore-sequencing technologies to develop low cost, convenient, user-friendly and portable devices or assays for screen and early detection of hepatocellular carcinoma (HCC) in Chinese population

**3. 1. Background and Significance**

**3.1.1. Develop a new generation of high-performance, low-cost, non-invasive and portable ultrasound scanner for early detection and screen of hepatocellular carcinoma which can be simply operated with limited infrastructure and by less trained healthcare staff will shift the paradigm from curative medicine to predictive medicine and create great potential to substantially improve early diagnosis and reduce mortality of HCC. This will also open a new avenue for early detection and screen of all cancers in China.**

In 2009, approximately 380,000 new liver cases and 350,000 liver cancer deaths occurred all over the China as a result of the high prevalence of chronic viral hepatitis. The crude incidence and crude mortality of liver cancer was 28.71/100,000 and 26.04/100,000, respectively. In some area incidence can be as high as 70-80/100,000 (Yuen et al. 2009). Analysis on the basis of geographic location showed that the incidence rate was higher in rural areas than in urban areas. China accounts for 55% of the roughly 650,000 new HCC cases worldwide (International Agency 2012). The crude and age-standardized incidence rates were 35.78/100 000 and 34.34/100 000 for rural areas. HCC is the second most common cancer in urban areas and first most common in rural areas (Song et al. 2013). Every minute, six people in China are diagnosed with cancer and every five minutes, there is one death from cancer according to the National Central Cancer Registry’s 2012 Annual Report in China (Chen et al. 2013). There are increasingly growth in cancer incidence in China due to rapid industrialization and population growth, environment deterioration, poor living habit, and aging. Pollution is the key reason behind the rise. We also observe that the cancer mortality in China is on the rise due to limited resources for cancer screening, early detection and treatment and lack of clinical symptoms in early stage HCC. Unfortunately, most HCC patients are diagnosed at an incurable advanced stage which leads to the almost equal incidence and mortality rate of HCC (Hu et al. 2013). HCC ranks third in annual global cancer mortality rates. More seriously, HCC has the shortest survival time of any cancer in both males and females (Kew 2012). Major population at risk of developing HCC in China was people with hepatitis B virus (HBV) infection (Song et al. 2013). It is reported that 93 million Chinese are HBV carriers, and about 20 million of them have chronic HBV infection. Early detection followed by treatment may is a key strategy to reduce mortality of HCC. Screening high-risk patients with HBV-or HCV-related chronic liver disease will increase the rate of early HCC detection and improve the rate of curative treatment (Song et al. 2013; Kudo et al. 2010). However, in China, it is lack of a government-funded national program to screen for high-risk patients with HBV-related chronic liver disease.

Many studies show that serological tests using protein biomarkers have a low sensitivity and hence do not provide adequate information for screening (Sherman 2001; Ayuso et al. 2011). Imaging techniques are a key tool for the early detection and the management of HCC (Bruix et al. 2011). The non-invasive ultrasonography (US), computed tomography (CT) and magnetic resonance imaging (MRI) allow detection and diagnosis of HCC at an asymptomatic stage. Unlike in the developed world where teams of specialists use sophisticated imaging technologies such as Dynamic multiphasic multidetector-row CT (MDCT) and magnetic resonance imaging (MRI) to examine patients (Lee et al. 2012), CT and MRI are expansive to use for a large proportion of patients in China. we are often facing shortages of trained medical practitioners, limited facilities, and lack of expansive image resources (Cheng et al. 2013). Ultrasonography (US) that is inexpensive, not invasive and can be repeated without risk is recommended to use for surveillance program of HCC (Ayuso et al. 2011). HCC merges as a small nodule and progresses into a larger nodule (Kojiro et al. 2006). The main purpose of screening for HCC by US in the population at risk is to detect tumor <20mm because they have low probability of vascular invasion (Bruix et al. 2010; Kojiro and Roskams 2005). The sensitivity and specificity of US for detection of HCC in cirrhosis patients range from 60% to 80%, and from 5% to 94%, respectively (Ayuso et al. 2011; Bolondi 2003). The problem with easily accessible US for screening of HCC is that it is often difficult to distinguish small HCC tumors from the nodularity of the cirrhotic liver. It requires expertise to improve the sensitivity of the US techniques (Lee et al. 2010). MRI has higher sensitivity than US (89%-100%), but will generally not be used for screening because its cost is high and it is not widely available (Lencioni et al. 2005). The major barriers in promoting to use US for the early detection of HCC and establishment of national program to screen for HCC in China are (1) lack of low-cost and durable US, (2) lack of professional local healthcare staff, (3) lack of US that can be operable in locations with limited or no medical infrastructure and provides rapid results, and (4) lack of portable US. The current US scanners are fairly large, expensive and situated in hospitals and clinics. To overcome these problems, in this application we will use the MobiUS SP1 System and MobiUS TC2 System as major tools for screening and early detection of HCC. The proposed approach has several remarkable features. First, the MobiUS is small and portable. It weights less than 12 ounces and can fit in our pocket and it’s light enough to take any locations where the patient needs immediate care. Second, it is easy to use. The touch screen user interface is extremely straightforward. Third, the connected Images are readily stored in 8 or 64 GB capacity and quickly shared via USB port or transmitted over a mobile network or WiFi. Fourth, since the system has built-in network connectivity which allows easy sharing of images from the device for second opinion or remote diagnosis or automatic detection by advanced analytic methods. Fifth, it is affordable. Its cost is very low. The sixth, since it is battery powered, it can work off the electric grid. The simple, portable, low cost MobiUS that plugs into smartphone for early detection of HCC provide instant scan image on the mobile device’s screen and a breakthrough in early diagnosis of cancer and will revolutionize screen of HCC and Medicare in China.

**3.1.2. Circulating miRNAs in blood which** **are a class of small noncoding RNAs to regulate gene expression as noninvasive biomarkers are emerging as promising and** **easily accessible biomarkers for screen of HCC. Perform a large HCC study with a novel, low cost, enzymatic color reaction-based miRNA biochip for further validating miRNAs as markers for screening and early detection of HCC in China will help collect necessary information that can help establish a government-funded nationwide screening and surveillance program for high-risk patients with hepatitis B virus (HBV)-related chronic liver disease and facilitate the use of miRNA as a non-invasive screening tool and early detection of HCC in China.**

The widely used serological markers, α-fetoprotein (AFP) test and des--carboxy prothrombin (DCP) have long been used as an early diagnostic biomarker of HCC, but have low sensitivity and specificity. AFP has specificity (75%) and sensitivity (68%),whereas elevated DCP activity is only observed in 44-47% of HCC (Abdalla and Haj-Ahmad 2012a; El-Houseini et al. 2005; Sheng et al. 2013; Collier and Sherman 1998), They are inadequate for screening (Ayuso et al. 2011) and are used the sole screen for HCC has been questioned (Sherman 2005). The development of non-invasive biomarkers for early detection of HCC with high sensitivity and specificity would be highly beneficial and urgently needed (Abdalla and Haj-Ahmad 2012b). miRNAs that are approximately 22-nucleotide noncoding RNAs regulate the molecular pathways underlying carcinogenesis (Garzon et al. 2006; Thorgeirsson et al. 2002).

Several recent studies found that miRNA were surprisingly informative (Lu et al. 2005). Aberrant miRNA expression promotes pathologic conditions of hepatocellular HCC. miRNA expressions are tissue specific and tumor specific (Volinia et al. 2006). miRNAs acting as oncogenes or tumor suppressors contribute to HCC development (Guo and Friedman 2013). Comprehensive analysis of the miRNA expression demonstrated that miRNAs have much higher cancer diagnostic and prognostic accuracy than mRNAs. It was reported that the sensitivity and specificity of circulating miR-483-5p for classifying HCC were, respectively, 75.5% and 89.8% and suggested miR-483-5p as a potential HCC biomarker (Sheng et al. 2013). In three independent cohort studies with 934 samples including healthy, chronic hepatitis B, cirrhosis, and Hepatitis B-Virus (HBV)-related HCC), our group identified seven promising diagnostic plasma miRNA biomarkers for HCC (miR-122, miR-192, miR-21, miR-223, miR-26a, miR-27a and miR-801). Merging seven miRNAs as a panel to diagnose HCC, we can reach a high classification accuracy of HCC (AUC=0.864 and 0.888 for training and validation data set, respectively). We also found that the miRNA panel can well differentiate HCC from healthy (AUC=0.941), chronic hepatitis B (AUC=0.842), and cirrhosis (AUC=0.884), respectively. Based on this preliminary results, we will develop a novel, low cost, enzymatic color reaction-based miRNA biochip detection assay for early detection of HCC. The detection system realizes a high level of automation. The monitoring and measurement of the hybridization and color reaction operate automatically. Specific software is incorporated into the system to interpret and transform data from the scanner and to generate test reports. This automation of detection and data analysis substantially reduces variability in the process. The biochip system has good scalability. New probes specific for other miRNAs can easily be added to the system. Therefore, we can easily expand the initially discover miRNA panel to adapt tests to other different subpopulations in China. The assay is rapid and simple to use. The whole testing progress can be finished in a short time.

To perform a large HCC study with newly developed miRNA biochip for further validating and selecting miRNAs as markers for screening and early detection of HCC in China will help collect necessary information that can help establish a government-funded nationwide screening and surveillance program for high-risk patients with hepatitis B virus (HBV)-related chronic liver disease and facilitate the use of miRNA as a screening tool and early detection of HCC.

**3.1.3. Develop novel sparse sufficient dimension reduction algorithm for efficiently extracting image information and miRNA variants and classification method based on matrix completion and functional data analysis techniques for image and miRNA data analysis will shift the analytic paradigm for disease risk prediction and diagnosis**

**from multivariate analysis to functional data analysis and will substantially improve our ability to accurately detect HCC at its early stage. The proposed novel powerful data reduction and classification methods will facilitate to develop automatic HCC detection system that requires minimum training for local healthcare staff.**

Image data are complex and high dimensional. An image is often represented in the form of a m dimensional vector of each pixel. Typically, there are hundreds of thousands of pixel per dataset (Landgrebe 2002). Most effective image features are at high dimensional space (Lee et al. 2013). The dimension of image data is very high. However, the source of high-dimensional data in reality often has limited number of dimensions which contain meaningful and essential information, which is usually defined as the intrinsic dimensionality of the data. Most dimensions that are outside the intrinsic dimensionality are either redundant or provide no useful information. They are irrelevant with disease and phenotypes. The high dimensional image data that are full of redundant information and noises will lead to reducing classification accuracy, compromising automation of image analysis and posing great challenges in developing ultrasound scanner for early detection of HCC.

To address these challenges, we proposal three novel approaches. Dimension reduction of image features in which high dimensional data are projected to low dimensional space by minimizing the loss of the information of the high-dimensional data plays an essential role in image analysis (Yao et al. 2013). The classical data reduction techniques are principal component analysis (PCA) and multidimensional scaling (MDS), among others. These data reduction methods are based on multivariate statistical analysis. The first novel approach is to shift data reduction techniques from multivariate analysis to functional data analysis. The image measurements between neighboring regions are highly correlated. The current multivariate data reduction techniques ignore space order and dependent information of image data. To overcome these limitations, we plan to explore functional data analysis. In stead of using PCA for data reduction we use functional principal component analysis (FPCA). Unfortunately, the current FPCA methods can only be applied to one dimensional data. However, image data are two or three dimensional. To address this issue, we extend one dimensional FPCA to the two or three dimensional FPCA and apply them to image data reduction. The second novel approach is sparse sufficient dimension reduction. Since dimension of image data is very high and the most image measurements across the whole region are irrelevant with disease, we replace the original image predictor features with its prediction onto a low dimensional space of the predictor space – a minimal set of linear combinations of the original predictors which is referred to as a central subspace (CS), without loss of information on response (Li 1991;Cook 1994; 1998; Fukumizu et al. 2006). The third novel approach is matrix completion. Ultrosound image often corrupted with various types of noises such as impulsive, Poisson and Gaussian noises (Pizurica et al. 2006) and hence needs to be denoised as a pre-processing step. Denoising is one of the most significant tools in ultrasound imaging (Achim et al. 2001; Roozgard et al. 2012). However, the widely used ultrasound image denoising algorithms assume the additive white Gaussian noise models which may not be realistic (Barzigar et al. 2013). Existing other types of noises will degrade the performance of these denoising methods. Matrix completion is a recently developed powerful method for removing noises from the data (Hu et al. 2013; Candes and Plan 2010). The key intuition of the matrix completion is to keep only the reliable pixels and remove all unreliable pixels that are corrupted by noises (Barzigar et al. 2013; The US for detection of HCC is based on the characterization of the nodular hepatic lesion, morphometric and topographic information and detection of changes in tissue density due to pathological changes (Ayuso et al. 2011). Extracting information on nodular hepatic lesion and anatomic structure changes for HCC from background image can be formulated as salient object detection in image analysis and can also be solved via matrix completion (Shen and Wu, 2012). In other words, an image is represented as summation of a low-rank matrix and sparse noises in a certain feature space, where the non-salient regions (or background) can be characterized by the low-rank matrix, and the salient regions are identified as the sparse noises. The proposed high dimensional data reduction and classification methods can be applied to joint image and miRNA data analysis, which will provide a general framework for screen and early detection of HCC using combined ultrasound image and miRNA information. Combination of these approaches not only substantially improve accuracy of HCC diagnosis, but also allow developing completely automatic scanner for early detection of HCC where its operation is simple, does not need human intervention and requires minimum training for local healthcare staff.

**3.2. Preliminary data**

**3.2.1. Ultrasound imaging for HCC diagnosis studies**

Miniature, portable and low cost ultrasound equipment can generate accurate and precise image. However, the ultrasound image is often corrupted by noise. The major limitation of ultrasound as a tool for screening HCC is that ultrasound is an operator dependent, thus its reproducibility is reduced. The quality of ultrasound image dependents on the experience and expertise of the healthcare staff. To overcome this limitation, we develop automatic HCC detection system that requires minimum training and has high accuracy. To evaluate the feasibility of the US-based automatic HCC detection system (For detailed description of the system, **please see approach section**), as a pilot study, the proposed approach was applied to HCC diagnosis study in Jinling Hospital affiliated with Medical School of Nanjing University in China. We included 249 individuals. From these, 101 cases with HCC **by imaging equipment . With patients having one lesion, patients two lesions, patients multiple lesions. Most patients being diagnosed in what stage?.** The sensitivity, specificity and accuracy of diagnosis of HCC were 79.21%, 87.84 and 84.34, respectively. This showed that using US for detection of HCC, we can reach good accuracy. To alleviate the impact of sample structures on the accuracy and provide reliable results, we also carry out 5-fold cross validation analysis. The sensitivity, specificity and classification accuracy of the US for diagnosis of HCC were listed in Table 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 1. Sensitivity, specificity, accuracy of the ultrasound for detection of HCC by 5 fold cross validation | | | | | | |
|  | Training Data Set | | | Test Data Set | | |
| Fold of the Data | Sensitivity | Specificity | Accuracy | Sensitivity | Specificity | Accuracy |
| CV-1 | 0.74074 | 0.74545 | 0.74346 | 0.75000 | 0.78947 | 0.77586 |
| CV-2 | 0.70000 | 0.74803 | 0.72947 | 0.71429 | 0.71429 | 0.71429 |
| CV-3 | 0.69136 | 0.74380 | 0.72277 | 0.75000 | 0.74074 | 0.74468 |
| CV-4 | 0.63953 | 0.81148 | 0.74038 | 0.66667 | 0.80769 | 0.75610 |
| CV-5 | 0.73684 | 0.74107 | 0.73936 | 0.72000 | 0.86111 | 0.80328 |
| Average | 0.70170 | 0.75800 | 0.73509 | 0.72019 | 0.78266 | 0.75884 |

**3.2.2. Plasma miRNA panel for HCC diagnose study**

Many miRNAs have been shown to be associated with various biological processes, including development, cellular proliferation, apoptosis, oncogenesis, and tumor metastasis (Lee et al. 1993; Garzon et al. 2006; Calin and Croce 2006; Volinia et al. 2006; Borel et al. 2012; Bushati and Cohen 2007). It is also found that a large percentage of mRNAs may be subject to regulation by miRNAs (Lewis et al. 2005). Many studies report that a large amount of stable miRNAs are circulated in human serum/plasma and show differential expressions between HCC and normal tissues (Li et al. 2010; Pineau et al. 2010; Chen et al. 2008; Mitchell et al. 2008). Plasma miRNA can be potential biomarkers for detection of HCC (Borel et al. 2012). To evaluate the feasibility of miRNA as biomarkers for diagnosis of HCC and identify a panel of miRNA for detection of HBV-related HCC, we investigated expression profiles of 723 miRNAs in a large cohort of 934 participants.