A METHOD OF PREDICTING SURVIVAL OUTCOME OF A SUBJECT HAVING HEPATOCELLULAR CARCINOMA (HCC) OR CHOLANGIOCARCINOMA (CCA)

Field of Technology

The present disclosure relates to a method for generating a predictive survival outcome of a subject suffering from liver cancer. More particularly, the predictive survival outcome is generated based on expression profile of two genes or proteins, namely, polo-like kinase 1 (PLK1) and epithelial cell transforming 2 (ECT2), exhibited in a carcinoma tissue specimen acquired from the subject. The present disclosure also offers an apparatus equipped for generating the predictive survival outcome by way of the disclosed method.

Background

Primary liver cancer consists of two main histologically-distinct subtypes, i.e., hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (CCA), whose diagnoses and treatment decisions are uniquely based on their baseline clinical features. The extensive tumor heterogeneity of HCC or CCA is attributed to the presence of complex, multifactorial etiologies, including environmental factors such as hepatitis B virus (HBV), hepatitis C virus (HCV), parasitic infections and chemical carcinogens. Other risk factors include an unhealthy lifestyle, such as cigarette smoking, excess alcohol intake and dietary factors [3], in addition to sex and race/ethnicity disparities wherein liver cancer mainly affects men and is highly prevalent in Asian populations (http://globocan.iarc.fr/). HBV and HCV are the major causative etiological factors for HCC, accounting for up to 90% of liver cancer globally, while CCA is uncommon except in South-East Asia such as northeastern Thailand, where infection with liver fluke (*O. viverrini*) is endemic and approximately 60% of liver cancers are CCA [4]. These global disparities may be attributed to the presence of different etiological factors among different ethnic groups. One hypothesis is that various causative factors can evoke distinct molecular mechanisms to independently initiate malignant transformation, which results in inter-tumor genomic heterogeneity. Like many other solid cancers, biological and genetic heterogeneities in each type of the referred HCC and CCA cause them highly resistant to treatment, ranking them as the second most lethal malignancies worldwide [1,2]. Therefore, efforts have been put in to identify the underlying genes or proteins expressed in driving cancer development, such that patients can be better stratified according to the identified cancer drivers thus facilitating treatment through precision medicine.

For example, Chen et al. [45] found ECT2 expression is closely associated with the activation of the Rho/ERK signaling axis in promoting early HCC recurrence. Sun et al. [46] has also demonstrated that another gene, PLK1, was expressed significantly higher in HCC sample and can be an independent prognostic factor for HCC applicable for diagnosis and treatment. Similar conclusions were drawn by Takai et al. [47] in their publication supporting findings of Sun et al [46]. Researchers have further devised several diagnostic or prognostic means capitalizing on the association established between liver cancer and PLK1 or ECT2 expression. For instance, United States patent publication no. 2011/03119280 discloses a DNA biochip for HCC diagnosis by scanning multiple gene markers including PLK1. Most of the aforesaid approaches centered on cancer prognosis correlating to individual expression of the candidate genes or proteins instead of in-depth information generated according to correlation between or among the candidate genes or proteins, though derivation of diagnostic outcome based upon interrelationship of multiple expressed candidate genes has been proposed in some publications relating to other cancer types. Somma et al. in their United States patent publication no. 2012/0197540 provide a method to predict mortality risk of a breast cancer patient by measuring expression level of multiple genes including both PLK1 and ECT2 in a tissue sample of the cancer patient; the method subsequently employs each of the measured genes expression level to compute a prognostic score, through a predetermined algorithm, being indicative of the mortality risk of the subject. However, no similar method or platform exploiting on interrelationship of co-expressed genes or protein for prognostic or diagnostic of liver cancer have been disclosed. The prognostic outcome derived from such method or platform should be applicable to facilitate better therapeutic results or allotment of medical resources.

Summary

The present disclosure is directed to a method for measuring or quantifying expression of PLK1 and ECT2 in a tissue specimen or sample acquired from a subject. Particularly, the expression described herein can refer to at least one of transcriptome or protein expression in the specimen. Preferably, the tissue sample is of HCC- or CCA-affected.

An object of the present disclosure is to offer a method for generating a prognostic or diagnostic result regarding a subject affected by HCC or CCA using the measured or quantified PLK1 and ECT2 expression in a tissue specimen. The prognostic or diagnostic result generally relates to predictive survival outcome or mortality risk of the subject. The prognostic or diagnostic result can be referred by a physician for carrying out precision medicine.

A further object of the present disclosure is to provide an apparatus being configured to, preferably in a fully or semi-automated fashion, quantify expression of PLK1 and ECT2 then generate the prognostic or diagnostic results with respect to medical condition of a subject affected by liver cancer.

At least one of the preceding objects can be met, in whole or in part, by the present invention, in which one of the embodiments includes a method for generating a predictive survival outcome of a subject affected by liver cancer, comprising: providing a tissue specimen of the subject, the tissue specimen including or being one of hepatocellular carcinoma or cholangiocarcinoma; performing at least one nucleic acid-based and/or proteomic-based assay to quantify transcriptome and/or protein expression of polo-like kinase 1 (PLK1) and epithelial cell transforming 2 (ECT2) in the provided specimen; deriving a ratio of PLK1 expression over ECT2 expression within a defined area of the provided specimen; and generating the predictive survival outcome of the subject according to the derived ratio, the subject being regarded with a first predictive survival outcome when the derived ratio is less than a predetermined value or a second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value. More particularly, the first predictive survival outcome corresponds to more than 40%-60% of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40%-60% or lower of overall survival percentage within the predetermined period in the survival analysis. Typically or preferably, the predetermined period is, but not limited to, 12 to 36 months.

In a number of embodiments, the nucleic-acid based assay is any one or combination of microarray-based technologies and polymerase chain reaction based technology including digital PCR.

In several embodiments, the proteomic-based assay includes or is a set of tissue microarrays and immunohistochemistry staining of human tissues.

In another aspect of the present disclosure, an apparatus for generating a predictive survival outcome of a subject having liver cancer is disclosed. The apparatus comprises a first module capable of identifying, detecting or determining transcriptome and/or protein expression of polo-like kinase 1 (PLK1) and epithelial cell transforming 2 (ECT2) in a tissue specimen of the subject; and a second module being configured for deriving a ratio of PLK1 expression over ECT2

expression within a defined area on the provided specimen and generating the predictive survival outcome of the subject according to the derived ratio, the second module configured to regard the subject as having a first predictive survival outcome when the derived ratio is less than a predetermined value, or having a second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value. More specifically, the first predictive survival outcome corresponds to more than 40% of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40% or lower of overall survival percentage within the predetermined period in the survival analysis. Typically or preferably, the nucleic-acid based assay abovementioned can refer to any one or combination of microarray-based technologies such as microarray and polymerase chain reaction based technology including digital PCR. Likewise, the proteomic-based assay can be tissue microarrays and immunohistochemistry staining of human tissues.

For a plurality of embodiments, the disclosed apparatus further comprises a test module capable of performing at least one nucleic acid-based or proteomic-based assay to couple expressed transcriptome or protein of PLK1 and ECT2 with a signaling moiety fashioned to continuously or periodically emit a signal detachable by the first module.

In some embodiments, the subject is preferably of Asian descent to attain results of greater accuracy. Also, the tissue specimen of the subject includes or is carcinoma being one of hepatocellular carcinoma or cholangiocarcinoma.

Brief Description of Drawings

Figure 1 shows results of consensus clustering with a hierarchical clustering algorithm with left panel: Empirical cumulative distribution (CDF) plots corresponding to the consensus matrices in the range of cluster number (K) = 2-8, middle panel: Corresponding change in area under CDF, right panel: Consensus matrices corresponding to a number K of clusters ranging between 2 and 5. Results for CCA or HCC samples are shown on the top or bottom row, respectively;

Figure 2 (a) is a heatmap of CCA subtypes based on consensus clustering and hierarchical clustering of the most variable genes that the x-axis represents CCA subtype consensus clusters, with CCA samples being represented in columns, grouped by the dendrogram into 4 main clusters and genes being represented in rows, the

expression of 1189 genes, identified as differentially expressed between C1 and C2 by t-test (FDR <0.05, fold change >2) are shown in log_2 from -3 to 3; (b) is heatmap of HCC subtypes based on 1020 differentially expressed genes (FDR <0.05, fold change >2) shown as in (a) that the x-axis represents HCC subtype consensus clusters; and (c) is a Subclass Mapping of CCA and HCC subtypes that significant relationships between clusters are represented by Bonferroni adjusted p-values;

Figure 3 shows VENN diagrams indicating the number of genes differentially expressed between C1 or C2 samples of CCA and HCC as defined by Student's T-test (FDR<0.05, fold change >2), with the left panel showing the number of genes upregulated in C1 in CCA (n=578) and HCC (n=656) and the overlapping genes (n=218), in which the probability for overlapping genes is statistically significant (p=3.20x10-97; hypergeometric test), and the right panel showing similar results for C2 upregulated genes;

Figure 4 shows Kaplan-Meier survival analysis of CCA subtypes (top panel) or HCC subtypes (bottom panel);

Figure 5 shows significant pathways, identified by GSEA analysis, of HCC or CCA C1 or C2 subtypes, represented by log10 p-values from 2 to 0 (p-value from 0.01 to 1);

Figure 6 shows (a) a comparison between Thai CCA subtypes and the published signatures indicated on the y-axis with each column representing a tumor sample, positivity of each signature is represented by grey bars, negativity by black bar and cases with FDR >0.05 are indicated in white, and (b) a comparison between Thai HCC subtypes and the published signatures indicated on the y-axis;

Figure 7 shows (a) the frequency of chromosomal aberrations for HCC (upper panel) or chromosomal aberrations in HCC with frequency >30% (lower panel), and (b) the like results for CCA;

Figure 8 shows (a) the frequency of chromosomal aberrations for the CCA C1 subtype (top panel) or CCA C2 subtype (bottom panel) that copy number gain or loss is shown

in light grey or black respectively, and (b) the frequency of chromosomal aberrations for the HCC C1 subtype (top panel) or HCC C2 subtype (bottom panel);

Figure 9 shows the relationship between high concordant genes and tumor subtypes along with the frequency of samples with copy number variation (CNV);

Figure 10 is a Kaplan-Meier plot of all 378 HCC cases;

Figure 11 shows representative images of CCA and HCC cases at 200x magnification based on immunohistochemical staining for PLK1 (left panels) or ECT2 (right panels);

Figure 12 reveals (a) correlation of PLK1 (left panels) or ECT2 (right panels) array expression and TMA score in CCA cases (upper panel) or HCC cases (lower panel), and (b) correlation of PLK1 and ECT2 array expression or TMA score in CCA or HCC cases;

Figure 13 shows results of Kaplan-Meier survival analysis of CCA cases (left panels) or HCC cases (right panels) based on a median cutoff of PLK or ECT2;

Figure 14 shows (a) Kaplan-Meier survival analysis of all CCA and HCC cases based on the ratiometric combination of protein expression (ECT2/PLK1), with low and high cutoff defined by ECT2/PLK1 ratio $\leqslant 1$ for low and >1 for high group, and (b) Kaplan-Meier survival analysis of HCC cases is shown as in (a);

Figure 15 shows subclass mapping and associated Kaplan-Meier survival analysis of (a) Thai HCC versus Chinese HCC subtypes, (b) Asian American (AsA) subtypes, (c) Thai CCA versus Japanese CCA subtypes, (d) Thai HCC versus European American (EA) HCC subtypes, and (e) Subclass Mapping of Thai CCA versus Caucasian CCA subtypes, with significant relationships between clusters are represented by Bonferroni-adjusted p values from 0 to 1 and subtypes in cohorts, no matching subtype in the Thai cohort being indicated with an X and significant associations between clusters being indicated in light grey with $p<0.05$;

Figure 16 shows box plots of BMI with median and standard deviation for C1 and C2 subtypes of Thai HCC, Asian American HCC and Thai CCA cases;

Figure 17 shows (a) a heatmap represents the correlation of metabolite abundance and gene expression in CCA samples, and (b) a heatmap is shown representing the global correlation of metabolite abundance and gene expression in HCC samples, with light grey and black bars indicating positive or negative correlation respectively, based on the Pearson R value from -1 to 1;

Figure 18 shows (a) hierarchical clustering of 81 metabolites separates HCC-C1 (light grey bar) and C2 cases (black bar), and (b) hierarchical clustering of 77 metabolites separates CCA-C1 (light grey bar) and C2 cases (black bar), in which samples are represented in columns, metabolites are represented in rows, and metabolite abundance is represented in log2;

Figure 19 shows (a) Ingenuity Pathway Analysis of the highly concordant metabolite/gene network, and (b) Ingenuity Pathway Analysis of the highly concordant metabolite/gene network, with upregulated metabolites in C1 subtype or the C2 subtype being denoted in light grey and dark grey respectively;

Figure 20 shows (a) box-plots of the abundance of three representative bile-acid-related metabolites in C1 (n=15) and C2 (n=14) HCC samples with Student's T-test p-values, and (b) box-plots of the abundance of three representative bile-acid-related metabolites in C1 (n=33) and C2 (n=18) CCA samples is shown with Student's T-test p-values;

Figure 21 shows (a) CIBERSORT analysis of the HCC C1 versus the HCC C2 subtype, and (b) CIBERSORT analysis of the CCA C1 versus the CCA C2 subtype, with high to low associations between cell types being scaled from 1 to -1, and the size of circles being indicative of the significance of the association that larger circles represent higher significance;

Figure 22 shows (a) box-plots of the abundance of three leukocyte types in C1 and C2 HCC samples with standard deviation and Student's T-test p-values, and (b) box-plots of

the abundance of three leukocyte types in C1 and C2 CCA samples with standard deviation and Student's T-test p-values.

Detailed Description

Hereinafter, the invention shall be described according to representative or preferred embodiments of the present invention and by referring to the accompanying description and drawings. However, it is to be understood the description and drawings corresponding to such embodiments is for purpose of clarity and to aid understanding, and it is envisioned that individuals having ordinary skill in the relevant art may devise various modifications without departing from the scope of the invention as defined by the appended claims.

The term "gene" as used herein may refer to a DNA sequence with functional significance. It can be a native nucleic acid sequence, or a recombinant nucleic acid sequences derived from natural source or synthetic construct. The term "gene" may also be used to refer to, for example and without limitation, a cDNA and/or an mRNA encoded by or derived from, directly or indirectly, genomic DNA sequence.

The term “transcriptome” used herein means a collection of RNA transcripts transcribed in a specific tissue, whether coding or non-coding, and preferably contains all or substantially all of the RNA transcripts generated in the tissue. These transcripts include messenger RNAs (mRNA), alternatively spliced mRNAs, ribosomal RNA (rRNA), transfer RNAs (tRNAs) in addition to a large range of other transcripts, which are not translated into protein such as small nuclear RNAs (snRNAs), antisense molecules such as short interfering RNA (siRNA) and microRNA and other RNA transcripts of unknown function. The transcriptome can also refer to proteins translated from the RNA transcripts as an extension of gene transcription.

One aspect of the present disclosure relates to a method for generating a predictive survival outcome of a subject affected with liver cancer. Essentially, the method comprises the steps of providing a tissue specimen of the subject, the tissue specimen being one of affected with hepatocellular carcinoma or cholangiocarcinoma; performing at least one nucleic acid-based and/or proteomic-based assay to quantify transcriptome and/or protein expression of polo-like kinase 1 (PLK1) and epithelial cell transforming 2 (ECT2) in the provided specimen; deriving a ratio of PLK1 expression over ECT2 expression within a defined area on the provided specimen; and generating the predictive survival outcome of the subject according to the derived ratio, the

subject being regarded with a first predictive survival outcome when the derived ratio is less than a predetermined value or a second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value. Preferably, the first predictive survival outcome corresponds to more than 40%-60% of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40%-60% or lower of overall survival percentage within the predetermined period in the survival analysis. It was found by the inventors of the present disclosure that relative expression of PLK1 and ECT2, tumor markers, in diseased tissue carrying or affected with hepatocellular carcinoma or cholangiocarcinoma correlates to survival outcome of the tested subject or acuteness of the liver cancer adversely affecting the subject. The acquired predictive survival outcome can be an effective tool to facilitate planning of strategic treatment and corresponding arrangement of medical resources according to medical needs of the tested subject.

According to several embodiments, the tissue specimen affected with hepatocellular carcinoma or cholangiocarcinoma in the present disclosure can be acquired through any operation(s) or medical procedure(s) known in the field. Preferably, the tissue specimen is extracted by way of liver biopsy, which can be, but is not limited to, a percutaneous-, transjangular- or laparoscopic-based procedure. For a number of embodiments, the type of liver biopsy carried out is dependent upon the downstream assay for quantifying the expression of transcriptome and/or protein relating to PLK1 and/or ECT2. For instance, the tissues specimen acquired through laparoscopy is more suitable for proteomic-based assay such as tissue microarray requiring sufficient area in the specimen to be prepared for immunochemistry staining and signal detection. On the other hand, it is preferable to subject tissue specimen obtained from percutaneous biopsy for nucleic acid-based assay that the mRNAs or transcriptomes corresponding to expression of PLK1 and/or ECT2 extracted from the specimen, proportionally amplified, and finally quantified.

For a number of embodiments, the nucleic-acid based assay is any one or a combination of microarray-based technologies and polymerase chain reaction based technology including digital PCR. The nucleic-acid based assay generally involves extraction of the overall mRNAs or transcriptomes from the tissue specimen of predetermined dimension followed by selectively quantifying the mRNAs or transcriptomes relating to PLK1 and/or ECT2 expression, with or without selective amplification of the mRNAs or transcriptomes of PLK1 and/or ECT2. Preferably, the transcriptomes of PLK1 and/or ECT2 hybridize on substantially complementary probes anchored on a platform. The hybridization process is configured to emit a signal, preferably a light

signal, detectable by one or more sensors located approximate to or within the platform. The probes are tagged with a signaling moiety which is fashioned to generate the light signal upon hybridization and forming of the double-stranded nucleic acids. The signals emitted are preferably proportional to the number of hybridization reactions that have occurred in the platform that enable subsequent quantification of the PLK1 and/or ECT2 expression by referring to a pre-established standard.

Similarly, for a plurality of embodiments, the proteomic-based assay is tissue microarrays and immunohistochemistry staining of human tissues by which the expressed protein or peptides of PLK1 and/or ECT2 are selectively coupled to one or more signaling moieties succeeding by detecting the signaling released from the coupled signaling moieties to result in quantification of the PLK1 and/or ECT2 expression. For example, in tissue microarrays (TMAs), the acquired tissue sample is used to prepare 0.5-2.0 mm cores of formalin fixed and paraffin embedded tissue sample. The prepared tissue sample proceeds to immunohistochemistry treatment at 1 to 10μm TMA sections in the form of slides. Further, the slides containing trimmed segment of the tissue specimen are deparaffinized in xylene, and rehydrated in graded alcohol. Antigen retrieval is then performed under pressure for 5 to 60 minutes with an acidic buffer. Anti-PLK1 (mouse monoclonal) is applied to the treated tissue sample at a 1:1000 dilution in a 2% non-fat milk solution at room temperature for 1 to 4 hours. Anti-ECT2 (mouse monoclonal) is applied at a dilution of 1:500 in a 2% non-fat milk solution. By bringing the antibody into contact with the targeted tumor marker or protein, the disclosed method results in yielding of antigen-antibody complexes detachable through a sensor or system such as those commercially available, for instance, Envision+ system (Dako) with concurrent use of 3,3'-diaminobenzidine (DAB). Preferably, the slides are counterstained with hematoxylin, dehydrated, cleared and coverslipped. Detection of the tagged or stained PLK1 and/or ECT2 can be performed at 200X or higher magnification for quantification of the PLK1 and/or ECT2 expression within a defined area. It is worth noting that the quantification can be performed manually by trained personnel or with the assistance of a signal sensor or image analyzer coupled to a computing device. Preferably, the treated tissue specimen can be scored for various image parameters or properties such as percentage of tumor cells and intensity of the indicating signal or staining. The scores of different image properties generated can be further manipulated such as by way of multiplying each other or being input into one or more algorithms to yield more insightful findings with respect to the subject affected by hepatocellular carcinoma or cholangiocarcinoma. In a number of embodiments, the expression level of PLK1 or ECT2 in the disclosed method corresponds to or is defined by

multiplying scores acquired from the treated specimen or image captured for the treated specimen. The expression level can be quantified by way of multiplying a score of percentage of tumor cells expressing PLK1 or ECT2 with a score of intensity of the signal or staining of the expressed PLK1 or ECT2, providing a computed score substantially representing the expression level of each tumor marker. For such embodiments, the expression level of the tumor marker under consideration corresponds to the computed score. The scale of the scores for each image property can range from 1-15, more preferably 1-10, and most preferably 1-4. Multiple areas of the treated specimen are scored independently; in some embodiments that the disclosed method computes a mean using these independent scores to attain better representation of the PLK1 and/or ECT2 expression, potentially yielding results of greater accuracy.

With the expressed transcriptomes and/or protein detected and quantified through the nucleic acid-based or protein-based assay, the disclosed method sets out to derive the ratio of the PLK1 expression over ECT2 expression within a defined area of the provided specimen. For example, the derived ratio is 1.5 when the expression level of the PLK1 is numerally represented as 12 and the expression level of ECT2 calculated is 8. For several embodiments, the ratio can be calculated or computed by inverting the numerator and denominator of the aforesaid ratio employing ECT2 and PLK1 expression levels respectively as numerator and denominator instead. Preferably, the area has a dimension of, but not limited to, 0.1 to 1.0mm in length, 0.1 to 1.0mm and 1 to 4um in thickness. To simplify the process in deriving the predictive survival outcome of the subject, the disclosed method regards the subject to have the first predictive survival outcome when the derived ratio is less than a predetermined value or the second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value. In various embodiments, the predetermined value can be any number within, but not limited to, the range of 0 to 16 in connection to a 0-4 scale of scores used. The predetermined number can range from 0.001 to 1 as well depending upon the scale of scores applied in the relevant embodiments. Also, the derived ratio can be analyzed or mathematically further associated with expression level of other genes or proteins relating to hepatocellular carcinoma or cholangiocarcinoma to produce a further or more conclusive prediction or observation.

As aforementioned, in accordance with some embodiments, the first predictive survival outcome corresponds to more than 40%-60% of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40%-60% or lower of overall survival percentage within the predetermined period in the survival analysis. In

addition, the survival percentage generated can be considered as an indicator offering empirical information with respect to potential responsiveness of the tested subject towards general treatment against hepatocellular carcinoma or cholangiocarcinoma. The tested subject may have to undergo alternative treatment option rather than conventional remedy when the subject has been associated with or assigned to second predictive survival outcome or relatively poor survival percentage. The inventors of the present disclosure also noticed that the disclosed method yields more reliable or robust results for those subjects with Asian genetic makeup or Asian descent.

In some embodiments, the predictive survival outcome portrays or only portrays survival percentage of the subject over a predetermined period which preferably spans for 6 to 48 months, or more preferably 12 to 36 months.

Pursuant to another aspect of the present disclosure, an apparatus for generating a predictive survival outcome of a subject affected of liver cancer is disclosed. Particularly, the disclosed apparatus comprises a first module capable of identifying or detecting transcriptome and/or protein expression of PLK1 and ECT2 in a tissue specimen of the subject; and a second module being configured to derive a ratio of PLK1 expression over ECT2 expression within a defined area on the provided specimen and generating the predictive survival outcome of the subject according to the derived ratio, the second module being configured to regard the subject to have a first predictive survival outcome when the derived ratio is less than a predetermined value or regard the subject to have a second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value. Preferably, the first predictive survival outcome corresponds to more than 40%-60%of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40%-60%or lower of overall survival percentage within the predetermined period in the survival analysis.

The first module described in the present disclosure can be any known platform operable under the principles of genetic, proteomic, immunochemistry and/or histochemistry to selectively render PLK1 and/or ECT2 expression detectable by human or machine reading. Therefore, in some embodiments, the disclosed apparatus may further comprises a test module capable of performing at least one nucleic acid-based or proteomic-based assay to couple expressed transcriptome or protein of PLK1 and ECT2 with a signaling moiety fashioned to continuously or periodically emit a signal detachable by the first module. Particularly, the first module is configured to attach or couple the transcriptomes and/or proteins expressed for PLK1 and/or ECT2 with a signaling

moiety capable of releasing detectable signal to highlight presence of the expressed transcriptomes or protein which the signaling moiety coupled to. The first module of the present disclosure may be composed of multiple working stations arranged to run sequential processes to prepare and treat the tissue specimen to finally lead to identification of the PLK1 and/or ECT2 expression. These multiple working stations can be interconnected or operate separately one another according to the preference of the disclosed embodiments. Preferably, in more embodiments, the first module comprises an illumination device for the first module to highlight the expression of PLK1 and/or ECT2, and an image capture device designed to create a digital image of the illuminated specimen; the first module uploads the taken image for the second module to analyze and derive the predictive survival outcome. Most preferably, the first module stores the created digital image in a database in communication such that the stored image can be recalled and subjected to further analysis. For instance, the first module can be gene expression microarray or tissue microarray system for detecting and quantifying expression of the PLK1 and/or ECT2 in tissue sample affected of hepatocellular carcinoma or cholangiocarcinoma.

According to a number of embodiments, the second module is adapted to derive the ratio of PLK1 expression over ECT2 expression or vice versa within the defined area of the provided specimen and generating the predictive survival outcome of the subject correlating to the derived ratio. The second module can be a computing means installed with a set of commands in the form of computing program to run the analysis on the captured image, properties of the captured image may be correspondingly manipulated in line with the analysis performed. Preferably, the second module analyzes the treated tissue specimen, or more preferably image of the treated tissue specimen, and provides score for various image parameters or properties such as percentage of tumor cells and intensity of the indicating signal or staining shown. The second module may further manipulate scores of different image properties generated such as multiplying the scores to each other or input the scores into one or more algorithm to yield more insightful finding with respect to the subject affected by hepatocellular carcinoma or cholangiocarcinoma. Still, in more embodiments, the second module defines expression level of PLK1 or ECT2 in a numeral computed score resulted from different scores individually associated with a plurality of properties in the image captured for the treated tissue specimen. For example, the second module quantifies the expression level by way of multiplying score of percentage of tumor cells expressing PLK1 or ECT2 as detected with score of intensity of the signal or staining of the expressed PLK1 or ECT2, yielding the computed score substantially representing the expression level of each targeted tumor marker. For such embodiments, the expression level of the interested tumor marker is regarded by

the disclosed apparatus corresponding to the computed score. The scale of the scores for each image property can range from 1-15, more preferably 1-10, and most preferably 1-4. The second module generates individual score for multiple areas of the treated specimen in some embodiments that the second module computes a mean of these independent scores to attain better representation of the PLK1 and/or ECT2 expression in the captured image.

Furthermore, the second module computes or derives the ratio of the PLK1 expression over ECT2 expression, or vice versa, within a defined area of the provided specimen. Preferably, the area has a dimension of 0.1 to 1.0mm in length, 0.1 to 1.0mm and 1 to 4um in thickness. To simplify the process in deriving the predictive survival outcome of the subject, the disclosed second module is regards the subject to have the first predictive survival outcome when the derived ratio configured to regard the subject to have a first predictive survival outcome when the derived ratio is less than a predetermined value or regard the subject to have a second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value. The predetermined value can be any number, but not limited to, within the range of 0 to 16 depending on the scale of score implemented. As abovementioned, the first predictive survival outcome corresponds to more than 40%-60% of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40%-60% or lower of overall survival percentage within the predetermined period in the survival analysis. Preferably, the predetermined period preferably spans for 6 to 48 months, or more preferably 12 to 36 months.

For some embodiments, the nucleic-acid based assay is any one or combination of microarray-based technologies for DNA and/or RNA polynucleotides detection, and polymerase chain reaction based technology including digital PCR.

In other embodiments, the proteomic-based assay is tissue microarrays tissue microarrays and/or immunohistochemistry staining of human tissues.

The following example is intended to further illustrate the invention, without any intent for the invention to be limited to the specific embodiments described therein.

Example 1

Cohorts and Clinical Specimens

A set of 398 surgical paired tumor and nontumor specimens derived from 199 sequential patients of the TIGER-LC cohort (130 CCA patients and 69 HCC patients) were used in this study. Tumor

diagnosis was confirmed by pathological assessment. Clinical, demographic, socioeconomic and morbidity data were abstracted from comprehensive questionnaires and medical chart records. A list of clinical variables assessed in this study is provided in Suppl Table S1. The characteristics of 156 Asian HCC patients and 163 Caucasian HCC patients from TCGA were also used in this study (TCGA Research Network: http://cancergenome.nih.gov/.). The characteristics of 104 Caucasian CCA patients and 182 Japanese patients from independent cohorts were described recently [14, 25]. The HCC cohort of 247 Chinese patients from LCI was previously described [35]. Informed consent was obtained from all patients included in this study and approved by the Institutional Review Boards of the respective institutions.

RNA Isolation and Transcriptomics

Total RNA was extracted from frozen tissue using TRIzol (Invitrogen) according to the manufacturer's protocol. Only RNA samples with good RNA quality as confirmed with the Agilent 2100 Bioanalyzer (Agilent Technologies) were included for array studies. The Affymetrix Human Transcriptome Array 2.0 was used to measure transcripts among paired tumor and nontumor tissue specimens. Raw gene expression data were normalized using the Robust Multi-array Average (RMA) method and sketch quantile normalization method. For genes with more than one probe set, the mean gene expression was calculated. The microarray platform and data were submitted to the Gene Expression Omnibus (GEO) public database at NCBI following MIAME guidelines (GEO Series GSE76297). Gene expression data of the three independent cohorts are accessible through GEO Series (GSE14520 and GSE26566), TCGA and European Genome-phenome Archive (EGA) database (EGA00001000950). All expression data were $\log_2$ transformed. Consensus clustering (cCluster; hierarchical clustering; Pearson distance; complete linkage; 1000 resampling iteration) was used to define subtypes among HCC or CCA using a 2615 most variable gene set that is common in both TIGER-LC HCC and CCA cohorts (Gene filter: 1.5-fold change from gene's median value, Log intensity variation p value>0.01). cClustering of validation cohorts was performed using genes that are common between the individual cohort and the aforementioned 2615 most variable genes. Hierarchical clustering (Partek Genomic Suite 6.6; Pearson distance; complete linkage) was used to determine and visualize the relationship among various subgroups defined by cCluster. An unsupervised subclass mapping method (SubMap) was used to identify common subgroups between independent cohorts using the default setting found in the SubMap module (http://genepattern.broadinstitute.org/) [11].

Previous studies have shown that HCC and CCA have an extensive inter-tumor, but to a much lesser degree, intra-tumor heterogeneity, as determined by transcriptome profiling [5-7]. To define molecularly homogenous tumor subgroups in Thai CCA and HCC patients, Affymetrix Human Transcriptome Array 2.0 was performed on 398 surgical specimens derived from 199 patients. Among them, 153 passed the quality control tests and were used for transcriptome analyses. Consensus clustering (cCluster) was used to define subtypes in HCC and CCA as previously described 8. cCluster revealed 4 major subgroups of CCA and 3 major subgroups of HCC based on consensus distributions and the corresponding consensus matrices (Figure 1). The relationship among various subgroups defined by cCluster can be visualized through unsupervised hierarchical clustering (Figure 2A-B). This analysis revealed unique subgroups within HCC or CCA cases with distinct gene expression patterns.

Several recent studies have described transcriptomic similarity between CCA and HCC [9,10]. The present disclosure used Subclass Mapping (SM) to better define the relationship among molecular subgroups of CCA and HCC [11]. This analysis revealed that the CCA-C1 and HCC-C1 subtypes have a similar gene expression matrix, while the CCA-C2 and HCC-C2/C3 subtypes are similar (Figure 2C). However, this relationship was not observed in non-tumor tissues, suggesting that subtype-related genes are tumor-specific. Analyses of gene expression concordance in C1 and C2 subtypes revealed that upregulated genes in gene cluster 1 (GC1) are similar between CCA-C1 and HCC-C1 (Figure 2A-B and Figure 3). Similar results were observed in GC2 gene clusters. Interestingly, cases defined as CCA-C1 and HCC-C1 had poor survival while those in CCA-C2 and HCC-C2 had better survival as revealed in Figure 4. Gene Set Enrichment Analysis (GSEA) of subtype-specific gene signatures, as shown in Figure 5, revealed that both CCA-C1 and HCC-C1 subtypes were enriched for mitotic checkpoint signaling pathways, suggesting that this subtype contains high chromosomal instability. In contrast, CCA-C2 and HCC-C2 subtypes were enriched for cell immunity-related pathways, suggesting that inflammatory responses are linked to the C2 subtype. These results indicate that despite the distinct histological differences between CCA and HCC, there were common subtypes with similar gene expression matrices and tumor biology.

Several gene signatures have been linked to HCC/CCA prognostic subtypes, cancer stem cell features and tumor metastasis [5,6,12-15]. We thus examined the relationship between the Thai HCC subtypes and known signatures using a nearest template prediction algorithm [16]. We found that the both CCA-C1 and HCC-C1 subtype is enriched for S1-2-related genes – based upon Hoshidas' studies [6,11,16], stem cell genes – based upon Lees' studies [12, 34] and EpCAM genes – based upon

Yamashitas' study [13]. Similarly, the CCA-C1 subtype is enriched for S1-2 genes, stem cell genes and EpCAM genes (Figure 6). In addition, both CCA-C2 and HCC-C2 are enriched for cases that are negative for the above signatures while C3/4 subtypes contain mixed cases. In contrast, the class 2 signature [14] and the proliferation signature [15] did not separate common subtypes well.

A major hallmark of liver cancer is its association with various types of etiological factors and its high heterogeneity in clinical presentation and underlying tumor biology. Consequently, most patients with liver cancer are refractory to treatment and have a dismal outcome. One of the essential requirements needed to improving their outcome is to provide the diagnostic tool kit that is capable of accurately defining homogenous molecular subtypes, each displaying unique tumor biology and potentially druggable driver genes in order to implement rational treatment choices based on molecular subtypes. Accordingly, the development of well-annotated biobanks of cancer patients, such as the efforts by TCGA and ICGC, are key resources to further develop the goals of precision medicine. Currently, TCGA and ICGC mainly include HCC specimens largely derived from North America, Europe and Japan. Considering that HCC and CCA are much more prevalent in Asian populations, the present disclosure established a TIGER-LC consortium to conduct a case-control study in Thailand where liver cancer, especially CCA, is endemic.

Hepatocarcinogenesis is a complex process resulting from an accumulation of genetic and epigenetic alterations of various cancer drivers over many decades. Tumor evolution is expected to vary significantly between different tumor types due to the fact that each tumor cell needs to adapt to the microenvironment's stress induced by different etiological factors. Consequently, liver cancer is especially prone to genetic heterogeneity. While whole genome/exome sequencing approaches are powerful in identifying potential cancer drivers, candidate HCC or CCA-related driver mutations are extremely heterogeneous, as each tumor carries a large number of low frequency mutated genes in various combinations [25,31,32]. Furthermore, most of these genetic abnormalities characterized by whole exome sequencing have been considered as either passenger or histological mutations that do not have any functional impact on the tumors being diagnosed [33]. For those genes with a relatively higher mutation frequency that have been characterized as candidate drivers, vast inter-tumor heterogeneity among different tumor lesions that carry varying combinations of these candidate drivers is clearly evident [31,32]. It is conceivable that a combination of different cancer drivers may emerge as new convergent adaptive pathways rewired for cancer cell survival that are unique to different subtypes. It should be noted that cancer drivers not only can be acquired by somatic mutations, but also via epigenetic mechanisms, such as DNA

methylation of oncogenes or tumor suppressor genes. This may explain why finding stable tumor subtypes with unique tumor biology by whole exome sequencing is very challenging as it only captures a fraction of tumor characteristics. In contrast, transcriptomic profiling has been successful in defining stable HCC subtypes that may reflect their unique tumor biology [6,13,34-36], although defining key driver genes among different tumor subtypes by transcriptome-based approaches has been challenging. An integrated omics approaching correlating genomics, epigenomics, transcriptomics, proteomics, and metabolomics data may be key to address tumor heterogeneity and cancer drivers.

Example 2

DNA Isolation and Somatic Copy Number Alterations (SCNA)

Total DNA was extracted from frozen tissue using a Phenol/Chloroform extraction protocol. Samples with sufficient amount of double-stranded DNA as confirmed by Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific) were included for array studies. The Affymetrix Genome-Wide Human SNP Array 6.0 was used to determined somatic copy number alterations (SCNA) among paired tumor and nontumor tissue specimens. The raw SCNA data is accessible through GEO Series GSE76213. SCNA of CCA and HCC was analyzed using the Partek Genomics Suite 7.5 using the paired non-tumor tissue as the reference for each patient. The segmented regions in CCA and HCC were found using the genomic segmentation algorithm in Partek. To identify genes that are concordantly regulated with SCNA, the Pearson correlation value was calculated between the copy number segmented region and the transcriptome data from either CCA or HCC tissues. To define the C1 and C2 subtype-specific concordant genes, the genes located in segmented regions with a positive correlation value and p-values $\leqslant 0.005$ were considered. To define the copy number concordant genes for the C1 and C2 subtype in CCA and HCC, the p-value from a Student's t-test and fold change of the $\log_2$ transformed expression value among the most variable genes used for the class prediction between C1 and C2 CCA or HCC subtypes were used. To identify C1 subtype-specific copy number concordant genes in CCA and HCC, inventors of the present disclosure selected the genes with a Student's t-test p-value $\leqslant 0.005$ when C1 and C2 subtypes were compared.

The inventors of the present disclosure next determined somatic copy number alterations (SCNA) among Thai tumor specimens and paired non-tumor tissues as controls using Affymetrix Genome-Wide Human SNP Array 6.0. Consistent with previously published studies [17], a typical SCNA

profile with recurrent gains and losses on 1q, 6p, 8q and 4q, 8p, 13q, 16, 17p, respectively, was evident in HCC specimens as presented in Figure 7A. It was found that SCNA profiles between HCC and CCA differ considerably (Figure 7B). However, when the inventors analyzed SCNA profiles based on C1 and C2 subtypes, there was a significantly higher degree of recurrent gains and losses found in both CCA-C1 and HCC-C1, when compared to CCA-C2 and HCC-C2 Figure 8A-B. This result is consistent with the transcriptome and GSEA analyses and suggests that CCA-C1 and HCC-C1 contain mitotic checkpoint defects, which may result in higher degrees of aneuploidy.

To determine potential subtype-related driver genes based on the notion that a driver gene should have high concordant alteration among SCNA and gene expression [17, 18], the inventors of the present disclosure first performed Pearson's correlation between SCNA and the transcriptome. This analysis revealed that there is a significant positive correlation among a subset of genes (Suppl Fig S5A-B), i.e., 239 genes for CCA-C1 and 89 genes for HCC-C1 tumor specimens. Among them, 51 genes overlapped between CCA-C1 and HCC-C1, suggesting the existence of common subtype-specific drivers in HCC and CCA. Consistently, more copy number gain and elevated expression were associated with the C1 subtype than C2 subtype as indicated in Figure 9. Consistent with the findings above, a network analysis of these 51 genes revealed enrichment of mitotic checkpoint signaling pathways linked to PLK1 signaling. In addition, ECT2 was the top ranking differentially expressed gene between the C1 and C2 subtypes. Similar results were observed using the TCGA HCC dataset presented in Figure 10.

Example 3

Immunohistochemistry

Tissue microarrays (TMAs) were constructed with 1.0 mm cores from formalin fixed, paraffin embedded tissue that had been reviewed by SH 44 as separate TMAs for cholangiocarcinoma and hepatocellular carcinoma. Matched TMAs of normal tissue from the patients were constructed as a reference. All TMAs contained internal control tissues. Immunohistochemistry was performed on 5μm TMA sections. First slides were deparaffinized in xylene, and rehydrated in graded alcohol. Antigen retrieval was performed in a pressure cooker for 20 minutes with a pH6 citrate buffer. Anti-PLK1 (mouse monoclonal, clone CN05-844, Millipore) at a 1:1000 dilution was applied at a 1:1000 dilution in a 2% non-fat milk solution at room temperature for 2hrs. Anti-ECT2 (mouse monoclonal, "E-1" cat SC-514769, Santa Cruz Biotechnology) was applied at a dilution of 1:500 in a 2% non-fat milk solution. Antigen-Antibody complexes were detected with Envision+ (Dako)

secondary, and DAB. Slides were counterstained with hematoxylin, dehydrated, cleared and coverslipped. Positive and negative controls were performed. A TMA of matched normal liver was stained concurrent with the tumor TMAs. Interpretation of immunohistochemical staining was performed at 200X magnification. Tumors were scored for percentage of tumor cells stained (0-4 in quartiles) and intensity of staining (0-4). The two values were multiplied (range 0-16). Normal tissue TMAs failed to demonstrate staining of tumor markers. High or low levels of PLK1 or ECT2 were determined using a median score. ECT2/PLK1 ratios were calculated from raw data, and cut points were determined as ECT2/PLK1 ratio $\leqslant 1$ for low and >1 for high group, similar to previously described 24, and plotted as Kaplan-Meier plots. The Cox-Mantel log-rank test was used for statistical analysis.

Pathway analysis was performed using Gene Set Enrichment Analysis (GSEA) version 16 and Ingenuity Pathway Analysis (IPA) version 24718999. Kaplan-Meier survival analysis was used to compare patient survival using GraphPad Prism 6 and the statistical p-value was generated by the Cox-Mantel log-rank test. All p-values are two-sided and the statistical significance was defined as $p<0.05$ unless otherwise noted. The relationship between previously reported signatures and the Thai CCA and HCC cohorts was determined using a nearest template prediction algorithm implemented in GenePattern based on a prediction confidence false discovery rate (FDR) cut-off of 0.05. An estimation of the relative fractions of immune/inflammatory cell subsets from tissue expression profiles of CCA or HCC was conducted using CIBERSORT 30. The gene expression data were converted by quantile normalization of log2 scaled expression matrix and relative fractions of leucocytes in C1 and C2 subtype of CCA and HCC were quantified according to the website (cibersort.stanford.edu/) with implemented analyses using the built-in LM22 signature matrix (LM22). Welch Two Sample t-test was used to compare each of the 22 relative leucocyte fractions between C1 and C2 subtype. Pearson's correlation analysis was used to determine correlation among leucocytes and BMI in C1 and C2 subtypes.

The data acquired suggested that ECT2 and PLK1 could be clinically relevant functional biomarkers useful to detect HCC and CCA subtypes since both genes have been previously linked to tumor progression [19-21]. Inventors of the present disclosure thus evaluated ECT2 and PLK1 by immunohistochemistry (IHC) on tissue microarrays (TMAs) of 199 Thai patients. TMAs of CCA and HCC were constructed. PLK1 is detected in the cytoplasm, but is not expressed in the normal hepatocytes (Figure 11). ECT2 is expressed in the nucleus, but is absent in normal hepatocytes. The expressions detected by IHC were correlated with mRNA data provided in Figure 12A.

Further analysis, Figure 12B, demonstrated correlation of PLK1 and ECT2 expression at the RNA level and protein expression level. PLK1 and ECT2 were subjected to survival analysis individually showing that, Figure 13, patients with high TMA score of PLK1 or ECT2 had a trend of poor outcome. Given the substantial correlation in expression between PLK1 and ECT2 and the fact that PLK1 has been demonstrated to phosphorylate ECT2 in vitro [22], the inventors evaluated the combination of the two biomarkers to predict outcome. Based on prior examples of ratiometric combination of related biomarkers [23,24], the present disclosure generated a PLK1/ECT2 ratio from the TMA scores, and performed survival analysis, shown in Figure 14A-B, which demonstrated that the PLK1/ECT2 ratio was robust to discriminate outcomes (CCA, p=0.01, HCC, p=0.012).

The experiments conducted revealed that the C1 subtype is linked to mitotic checkpoint defects and that PLK1 and ECT2 are two clinically relevant key genes for C1. It was found that both PLK1 and ECT2 expressions are highly expressed in the C1 subtype and are robust in defining tumor subtypes using IHC. This is clinically meaningful since IHC is a preferable method for pathological diagnosis. The ECT2 gene is also preferentially amplified or mutated in the C1 subtype, consistent with the hypothesis that it is a driver gene for this subtype. Consistently, both PLK1 and ECT2 have been functionally linked to cancer. PLK1 is one of the mitotic serine/threonine protein kinases and is required to initiate mitosis and to regulate spindle assembly. Over reactive PLK1 signaling has been found in human tumors, including HCC, and has been proposed to serve as a potential target for the treatment of cancer because of its crucial functional node in the oncogenic network [19]. Many PLK1 inhibitors are currently being tested in solid tumors [37]. ECT2 belongs to the Ras superfamily GEFs and GAPs [20] and is a classical oncogene originally identified in 1991 based on its ability to transform NIH 3T3 cells [21]. Its role as a tractable target for cancer therapy, including HCC, has been suggested since it is responsible for promoting early recurrence of HCC [20, 38]. The results acquired suggest that PLK1 and ECT2 expressions could serve as biomarkers for patient stratification and molecular targets for the C1 subtype of Asian HCC and CCA.

Example 4

Race/ethnicity-related common tumor subtypes

To determine if the common molecular subtypes of CCA and HCC we observed in the Thai samples are universal, inventors of the present disclosure examined several independent cohorts with available transcriptome data from patients who resided in Asia, Europe and North America. These include 247 HCC patients from China, 378 HCC patients from the U.S. (TCGA data), 104

CCA patients from Europe and 128 CCA patients from Japan. SM was used to determine similarities among various subtypes identified by the transcriptome. We found that the C1 or C2 prognostic molecular subtypes were observed in Chinese HCC patients, Asian American (AsA) HCC patients, Japanese CCA patients, but not in European American (EA) HCC patients or in European CCA patients (Figure 15). Similar results showing an association of 51 common subtype-related genes and race/ethnicity-related prognosis were also observed in HCC patients from the U.S. Taken together, these results suggest that common prognostic molecular subtypes of CCA and HCC are also related to race/ethnicity.

Since gene mutation data were available for 128 Japanese CCA tumors, the present disclosure tried to determine whether any relationships existed between the common subtypes identified in this study and the top 32 mutated genes described by Nakamura et al [25]. The present disclosure found that while mutation data do not clearly discriminate common subtypes, and mutation frequencies range from 1-43% for each subtype, mutations in TP53, KRAS, MYC, and GNAS (>10% mutations) showed enrichment in the C1 subtype with poor prognosis, and mutations in BAP1 and IDH1 were more frequent in the X subtype, a subgroup which did not match the C1/C2 subtypes, but was associated with a good prognosis. Noticeably, 15% of the X subtype, but none of C1 and C2 subtypes, carries IDH1 mutations, suggesting that this unique subtype of CCA, independent of C1 and C2, has a distinct gene pattern and better prognosis. Interestingly, it was found that a statistically significant number of genes (3 of 51 driver genes, i.e., NRAS, PRKCI and ECT2), based on SCNA and the transcriptome (Figure 9) overlap with these 32 mutated genes (p=0.0004; hypergeometric test). Noticeably, 6% of the Japanese CCA-C1 subtype also showed mutations in ECT2, consistent with our finding that ECT2 is a functional driver for the common C1 subtype.

Example 5

Obesity, metabolomics and tumor subtypes

To determine whether any of the etiological/demographic/clinical features are linked to the identified tumor subtypes, the present disclosure compared C1 and C2 subtypes based on available clinical variables that include age, gender, tobacco and alcohol consumption, body mass index (BMI), and tumor characteristics. Only age, BMI status and tumor size appeared to be different between the C1 and C2 subtypes of Thai HCC. It should be noted that alcohol consumption, HBV and HCV status, cirrhosis by Child-Pugh score, levels of alkaline phosphatase (ALP), CA19-9, alpha-fetoprotein (AFP), and tumor staging differed significantly between CCA and HCC. It is

interesting that these etiological factors are not associated with common molecular subtypes, but they are linked to histological subtypes.

Metabolon's Discover HD4 Platform was employed to measure small biochemical species among paired tumor and nontumor tissue specimens. Both liquid chromatography/mass spectrometry in positive and negative modes (LC+/LC−) and gas chromatography/mass spectrometry (GC/MS) were employed. A total of 718 metabolites were measured. The missing values were imputed using the minimum value of each metabolite. The 178 most variable metabolites common in both HCC and CCA cohorts were selected within each of the cohorts (Filter: 1.5-fold change from metabolite's median value, Log intensity variation p value>0.01). Pearson correlation values between the selected metabolites and the most variable genes were calculated for each cohort separately. Only significantly correlated metabolites and genes within the same samples (P <0 .05) were included in further analysis. Furthermore, metabolites that were significantly associated with at least 20% of the genes were selected.

It was further determined BMI profiles in Thai, AsA and EA patients with available BMI data. The present disclosure found that EA patients tend to have higher BMI than Asian patients regardless of whether they live in Asia or in the U.S. C2 subtypes tend to have a higher BMI than C1 subtypes and the difference is statistically significant (Figure 16). To determine if the difference in BMI is associated with changes in tumor metabolism, we performed untargeted metabolomic profiling among 199 HCC and CCA tumor specimens and paired non-tumor tissues from Thai patients. Since integrating metabolite and gene expression profiles are a powerful method of reducing false positivity and increasing the chances of defining functional metabolites [26], the present disclosure performed a global Pearson correlation analysis between metabolite and gene expression profiles. The result obtained is revealed in Figure 17. A total of 77 metabolites for CCA and 81 metabolites for HCC from a total of 178 most variable metabolites showed a high correlation with gene expression (-0.5<R<0.5; P <0 .05 and associated with at least 20% of the genes). It was found a statistically significant number of overlapping metabolites (n=46) between CCA and HCC (hypergeometric p=0.0007). Moreover, metabolites that showed a high concordance with gene expression discriminated the C1 and C2 subtypes as illustrated in Figure 18A-B. The top networks of most significant metabolites from HCC and CCA were strikingly similar (Figure 19A-B). The present disclosure found that bile acid-related metabolites, such as taurochenodeoxycholate and tauroursodeoxychoate (TUDCA), are significantly more abundant in both HCC-C2 and CCA-C2 than in C1 subtypes (Figure 20A-B).

The C2 subtype is associated with an increased BMI, which is known to be linked to metabolic disease and cellular inflammation [27,28]. In addition, the C2 subtype-related bile acid metabolites are known to be linked to inflammation and immunity [29]. Inventors of the present disclosure thus examined infiltrating immune cells in Thai HCC and CCA using CIBERSORT [30]. It was found that the activity of leukocyte infiltrates is much higher in the C2 than in the C1 subtype. The results are presented in Figure 21A-B. Noticeably, elevated CD4+ memory T cells, along with γδT cells, but reduced Treg cells, are associated with the C2 subtype (Figure 22A-B). Taken together, the results indicate that while the C1 subtype contains mitotic checkpoint defects with altered PLK1 and ECT2, the C2 subtype has elevated BMI, immune cell abnormality and abnormal bile acid metabolism.

The present disclosure found that both CCA and HCC, regardless of their differences in histology and associated etiologies, consist of several common molecular subtypes shared only among Asian patients. Interestingly, common CCA and HCC subtypes that share distinct gene expression matrices have similar prognostic outcomes, suggesting the presence of common molecular types beyond traditionally viewed histological tumor subtypes. Systematic integration of the cancer transcriptome, SCNA and metabolome revealed key oncogenic drivers linked to an aggressive molecular subtype. Specifically, the C1 subtype contains mitotic checkpoint defects, while the C2 subtype is linked to inflammation, obesity and bile acid biogenesis. These results suggest that treatment stratification should not be based on histological types, but rather on molecular types.

Metabolic liver diseases and obesity have been linked to liver inflammation and cancer [27,28,39,40]. However, molecular mechanisms underlying these associations are unclear. We found that a common clinical feature linked to the C2 subtype is an increased BMI, suggesting a possible association with liver-related metabolic diseases. As BMI is linked to inflammatory responses and metabolic disorders, we also examined tumor-associated leukocytes among the molecularly defined subtypes. It was found that lymphoid cells such as CD4+ memory T cells and γδT cells, but not myeloid cells, are significantly elevated in the C2 subtype. These results are consistent with the gene expression data of the C2 subtype, which was enriched for cell immunity-related pathways, reaffirming the idea that an inflammatory response is linked to the C2 subtype. It is interesting that several bile acid metabolites such as TUDCA, taurocholic acid and glycochenodeoxycholate are also consistently much more abundant in the C2 subtype than the C1 subtype in both HCC and CCA. Bile acids have recently emerged as versatile signaling molecules

to regulate cholesterol metabolism, energy and glucose homeostasis, which forms the basis for developing novel drug targets to treat common metabolic and hepatic diseases [41]. It is conceivable that some of these drugs may be applicable to treat the C2 subtype of liver cancer. The acquired results are also consistent with recent studies indicating that the obesity-induced gut microbial metabolite deoxycholic acid promotes liver carcinogenesis through senescence secretome and that diet can alter the human gut microbiome, facilitating diet-related diseases, such as obesity [29,42]. Moreover, dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in IL1-deficient mice, revealing a plausible mechanistic basis for diets with certain saturated fats for induction of immune-mediated diseases in genetically susceptible hosts [43].

While the acquired results indicate that the common C1-C2 subtypes are mainly associated with Asian patients, the reason for this race/ethnicity-related association is unclear. It is plausible that Western-type diets may induce dysbiosis differently between Asians and Caucasians, and that race/ethnicity-related gut microbiome difference may cooperate with bile acid metabolism to induce distinct carcinogenesis processes observed in the Asian-related common C2 subtype. Increased infiltrating T cells in the C2 subtype also suggest a plausible scenario where these tumors may be sensitive to immune checkpoint inhibitors. In summary, integrated omics approach adopted in the present disclosure has defined common molecular subtypes of CCA and HCC across several Asian populations and identified potential driver genes and metabolic process linked to the specific subtypes.

It is to be understood that the present invention may be embodied in other specific forms and is not limited to the sole embodiment described above. However modification and equivalents of the disclosed concepts such as those which readily occur to one skilled in the art are intended to be included within the scope of the claims which are appended thereto.

REFERENCES

1 Theise, N. D. Liver cancer. (International Agency for Research on Cancer, 2014).

2 Wang, X. W. & Thorrgeirsson, S. S. The Biological and clinical challenge of liver cancer heterogeneity. Hepatic Oncology 1, 5, doi:10.2217/HEP.14.18 (2014).

3 El-Serag, H. B. Hepatocellular carcinoma. The New England journal of medicine 365, 1118-1127, doi:10.1056/NEJMra1001683 (2011).

4 Sripa, B. et al. Liver fluke induces cholangiocarcinoma. PLoS.Med. 4, e201 (2007).

5 Ye, Q. H. et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat.Med. 9, 416-423 (2003).

6 Hoshida, Y. et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer research 69, 7385-7392 (2009).

7 Roessler, S. et al. Integrative genomic and transcriptomic characterization of matched primary and metastatic liver and colorectal carcinoma. International journal of biological sciences 11, 88-98, doi:10.7150/ijbs.10583 (2015).

8 Monti, S. et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. Blood 105, 1851-1861, doi:10.1182/blood-2004-07-2947 (2005).

9 Woo, H. G. et al. Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. Cancer research 70, 3034-3041 (2010).

10 Oishi, N. et al. Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of miR-200c and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma. Hepatology 56, 1792-1803 (2012).

11 Hoshida, Y., Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Subclass mapping: identifying common subtypes in independent disease data sets. PloS one 2, e1195, doi:10.1371/journal.pone.0001195 (2007).

12 Lee, J. S. et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology 40, 667-676 (2004).

13 Yamashita, T. et al. EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. Cancer Res. 68, 1451-1461 (2008).

14 Andersen, J. B. et al. Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. Gastroenterology 142, 1021-1031 e1015, doi:10.1053/j.gastro.2011.12.005 (2012).

15 Sia, D. et al. Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. Gastroenterology 144, 829-840, doi:10.1053/j.gastro.2013.01.001 (2013).

16 Hoshida, Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. PloS one 5, e15543, doi:10.1371/journal.pone.0015543 (2010).

17 Roessler, S. et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. Gastroenterology 142, 957-966 (2012).

18 Woo, H. G. et al. Identification of potential driver genes in human liver carcinoma by genomewide screening. Cancer research 69, 4059-4066 (2009).

19 Strebhardt, K. Multifaceted polo-like kinases: drug targets and antitargets for cancer therapy. Nat Rev Drug Discov 9, 643-660, doi:10.1038/nrd3184 (2010).

20 Vigil, D., Cherfils, J., Rossman, K. L. & Der, C. J. Ras superfamily GEFs and GAPs: validated and tractable targets for cancer therapy? Nature reviews. Cancer 10, 842-857, doi:10.1038/nrc2960 (2010).

21 Cook, D. R., Rossman, K. L. & Der, C. J. Rho guanine nucleotide exchange factors: regulators of Rho GTPase activity in development and disease. Oncogene 33, 4021-4035, doi:10.1038/onc.2013.362 (2014).

22 Niiya, F., Tatsumoto, T., Lee, K. S. & Miki, T. Phosphorylation of the cytokinesis regulator ECT2 at G2/M phase stimulates association of the mitotic kinase Plk1 and accumulation of GTP-bound RhoA. Oncogene 25, 827-837, doi:10.1038/sj.onc.1209124 (2006).

23 Chung, J. Y. et al. The expression of phospho-AKT, phospho-mTOR, and PTEN in extrahepatic cholangiocarcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research 15, 660-667, doi:10.1158/1078-0432.CCR-08-1084 (2009).

24 Kitano, H. et al. Profiling of phospho-AKT, phospho-mTOR, phospho-MAPK and EGFR in non-small cell lung cancer. J Histochem Cytochem 62, 335-346, doi:10.1369/0022155414523022 (2014).

25 Nakamura, H. et al. Genomic spectra of biliary tract cancer. Nat Genet 47, 1003-1010, doi:10.1038/ng.3375 (2015).

26 Budhu, A. et al. Integrated metabolite and gene expression profiles identify lipid biomarkers associated with progression of hepatocellular carcinoma and patient outcomes. Gastroenterology 144, 1066-1075 (2013).

27 Calle, E. E., Rodriguez, C., Walker-Thurmond, K. & Thun, M. J. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N.Engl.J Med. 348, 1625-1638 (2003).

28 Park, E. J. et al. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. Cell 140, 197-208 (2010).

29 Yoshimoto, S. et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature 499, 97-101, doi:10.1038/nature12347 (2013).

30 30 Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. Nature methods 12, 453-457, doi:10.1038/nmeth.3337 (2015).

31 Guichard, C. et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat.Genet. 44, 694-698 (2012).

32 Totoki, Y. et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet 46, 1267-1273, doi:10.1038/ng.3126 (2014).

33 Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet 15, 585-598, doi:Doi 10.1038/Nrg3729 (2014).

34 Lee, J. S. et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. Nat.Med. 12, 410-416 (2006).

35 Roessler, S. et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer research 70, 10202-10212 (2010).

36 Nault, J. C. et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. Gastroenterology 145, 176-187, doi:10.1053/j.gastro.2013.03.051 (2013).

37 Yim, H. Current clinical trials with polo-like kinase 1 inhibitors in solid tumors. Anticancer Drugs 24, 999-1006, doi:10.1097/CAD.0000000000000007 (2013).

38 Chen, J. et al. ECT2 regulates the Rho/ERK signalling axis to promote early recurrence in human hepatocellular carcinoma. Journal of hepatology 62, 1287-1295, doi:10.1016/j.jhep.2015.01.014 (2015).

39 Cohen, J. C., Horton, J. D. & Hobbs, H. H. Human fatty liver disease: old questions and new insights. Science (New York, N.Y.) 332, 1519-1523, doi:10.1126/science.1204265 (2011).

40 Welzel, T. M. et al. Metabolic syndrome increases the risk of primary liver cancer in the United States: a study in the SEER-Medicare database. Hepatology 54, 463-471, doi:10.1002/hep.24397 (2011).

41 Thomas, C., Pellicciari, R., Pruzanski, M., Auwerx, J. & Schoonjans, K. Targeting bile-acid signalling for metabolic diseases. Nat Rev Drug Discov 7, 678-693, doi:10.1038/nrd2619 (2008).

42 David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559-563, doi:10.1038/nature12820 (2014).

43 Devkota, S. et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature 487, 104-108, doi:10.1038/nature11225 (2012).

44 Kononen, J. et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nature medicine 4, 844-847 (1998).

45. Chen, J. et al. ECT2 REGULATES THE Rho/ERK signaling axis to promote in human hepatocellular carcinoma. J Hepatol. 62(6): 1287-95 (2015).

46. Sun, W. et al. High expression of Polo-Like Kinase 1 is associated with early development of hepatocellular carcinoma. International Journal of Genomics, volume 2014, Article ID 312130.

47. Takai, N. et al. Polo-like Kinases (PLKs) and cancer. Oncogene 24, 287-291 (2005).

Claims

1. A method for generating a predictive survival outcome of a subject affected of liver cancer comprising:

providing a tissue specimen of the subject, the tissue specimen being one of hepatocellular carcinoma or cholangiocarcinoma;

performing at least one nucleic acid-based and/or proteomic-based assay to quantify transcriptome and/or protein expression of polo-like kinase 1 (PLK1) and epithelial cell transforming 2 (ECT2) in the provided specimen;

deriving a ratio of PLK1 expression over ECT2 expression within a defined area on the provided specimen; and

generating the predictive survival outcome of the subject according to the derived ratio, the subject being regarded with a first predictive survival outcome when the derived ratio is less than a predetermined value or a second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value,

wherein the first predictive survival outcome corresponds to more than 40%-60% of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40%-60% or lower of overall survival percentage within the predetermined period in the survival analysis.

2. The method of claim 1, wherein the predetermined period is 12 to 36 months.

3. The method of claim 1, wherein the nucleic-acid based assay is any one or combination of microarray-based technologies and polymerase chain reaction based technology.

4. The method of claim 1, wherein the proteomic-based assay is tissue microarrays and/or immunohistochemistry staining.

5. The method of claim 1, wherein subject is of Asian descent.

6. An apparatus for generating a predictive survival outcome of a subject affected of liver cancer comprising

a first module capable of detecting transcriptome and/or protein expression of polo-like kinase 1 (PLK1) and epithelial cell transforming 2 (ECT2) in a tissue specimen of the subject; and

a second module being configured to derive a ratio of PLK1 expression over ECT2 expression within a defined area on the provided specimen and generating the predictive

survival outcome of the subject according to the derived ratio, the second module being configured to regard the subject to have a first predictive survival outcome when the derived ratio is less than a predetermined value or regard the subject to have a second predictive survival outcome when the derived ratio is equal to or greater than the predetermined value,

wherein the first predictive survival outcome corresponds to more than 40% to 60%of overall survival percentage within a predetermined period in a survival analysis, and the second predictive survival outcome corresponds to 40%-60%or lower of overall survival percentage within the predetermined period in the survival analysis.

7. The apparatus of claim 6 further comprising a test module capable of performing at least one nucleic acid-based or proteomic-based assay to couple expressed transcriptome or protein of PLK1 and ECT2 with a signaling moiety fashioned to continuously or periodically emit a signal detachable by the first module.

8. The apparatus of claim 6, wherein the predetermined period is 12 to 36 months.

9. The apparatus of claim 7, wherein the nucleic-acid based assay is any one or combination of microarray-based technologies and polymerase chain reaction based technology.

10. The apparatus of claim 7, wherein the proteomic-based assay is tissue microarrays, and/or immunohistochemistry staining.

11. The apparatus of claim 6, wherein the subject is of Asian descent.

12. The apparatus of claim 6, wherein the tissue specimen of the subject is carcinoma and being one of hepatocellular carcinoma or cholangiocarcinoma.