

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 May 2007 (24.05.2007)

PCT

(10) International Publication Number
WO 2007/058623 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(74) Agent: **SCHIWECK, Wolfram**; Viering, Jentschura & Partner, P.O. Box 1088, Rochor Post Office, Rochor Road, Singapore 911833 (SG).

(21) International Application Number:
PCT/SG2006/000340

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date:
8 November 2006 (08.11.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/738,697 21 November 2005 (21.11.2005) US

(71) Applicant (for all designated States except US): **SINGAPORE HEALTH SERVICES PTE LTD** [SG/SG]; 11 Third Hospital Avenue #07-00, SNEC Building, Singapore 168751 (SG).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

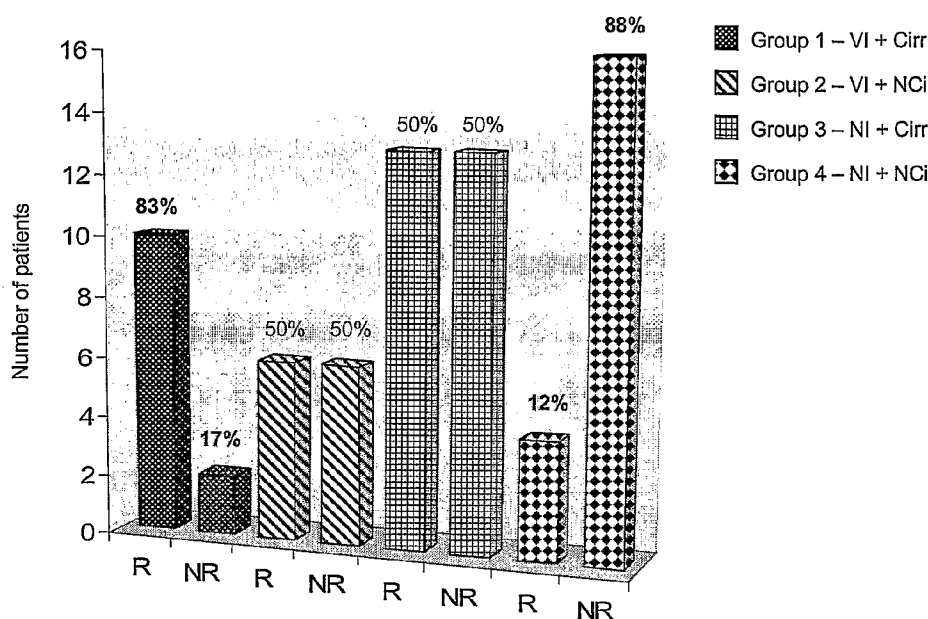
(75) Inventors/Applicants (for US only): **HUI, Kam Man** [SG/SG]; 66 Trevoe Crescent #04-08, Trevoe Park, Singapore 298067 (SG). **WANG, Suk Mei** [MY/SG]; 16 Cantonment Close, #15-41, Singapore 080016 (SG).

Published:

— with international search report

[Continued on next page]

(54) Title: METHODS OF PREDICTING HEPATOCELLULAR CARCINOMA RECURRENCE BY THE DETERMINATION OF HEPATOCELLULAR CARCINOMA RECURRENCE-ASSOCIATED MOLECULAR BIOMARKERS



(57) Abstract: Disclosed is a method to provide an accurate and reliable clinical prediction of the recurrence of human hepatocellular carcinoma (HCC), comprising the determination of a combination of clinicopathological information and novel molecular markers. The method relies on the evaluation of two important clinicopathologic features, namely vascular invasion and cirrhosis, and gene expression profiling of a predictive gene signature. Utilizing this method HCC patients with a high risk of disease recurrence can be identified more reliably and accurately. Based on the newly identified gene signature, disease recurrence could be predicted with up to 88% accuracy in HCC patients.

WO 2007/058623 A1



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Methods of predicting hepatocellular carcinoma recurrence by the determination of hepatocellular carcinoma recurrence-associated molecular biomarkers

Field of the invention

[0001] The invention relates to methods of predicting hepatocellular carcinoma recurrence in subjects utilizing the determination of clinicopathological factors as well as gene expression profiling of molecular biomarkers associated with recurrent human hepatocellular carcinoma.

Background of the invention

[0002] Hepatocellular carcinoma (HCC), the predominant histological subtype of primary liver cancer, is one of the most prevalent cancer types worldwide, accounting for an estimated 500,000 deaths annually. Although the highest HCC incidence rates occur in Southeast Asia and sub-Saharan Africa, recent epidemiological studies have projected an overwhelming increase in the incidence and mortality rates of HCC for the next decade in North America, Europe, and Japan (Bonn D., Hepatocellular carcinoma on the increase in USA, *Lancet* 353 (1999), 989). Over the past 25 years the incidence of HCC has doubled in the United States, and incidence and mortality rates are likely to double again over the next 10-20 years.

[0003] HCC is commonly fatal due to its high incidence of metastasis and recurrence and its frequent association with cirrhosis. Although early HCC is potentially curable by surgery (partial hepatectomy), only a minority of patients is amenable to surgical resection due to the asymptomatic feature of HCC progression (Llovet, J.M. et al., Hepatocellular carcinoma, *Lancet* 362 (2003), 1907-17). Although great advancements have been made in the surgical treatment of HCC, recurrence after surgery remains a key clinical challenge and intrahepatic metastasis and *de novo* tumor formation from residual primary tumor often lead to clinical complications (Llovet, J.M. et al., Hepatocellular carcinoma, *Lancet* 362 (2003), 1907-17; Thomas, M.B. & Zhu, A.X., Hepatocellular carcinoma: the need for progress, *J. Clin. Oncol.* 23(13) (2005), 2892-9; Okuda, K., Hepatocellular

carcinoma., J. Hepatol. 32 Suppl 1 (2000), 225-37). Due to the high recurrence rate in the remnant liver following resection, the overall five-year survival rate after resection has remained as poor as 35-50%.

[0004] To date, the molecular pathogenesis of HCC is not well understood, as the carcinogenesis of HCC is associated with multiple risk factors and is a long-term, multi-step process that is believed to require many contributing factors. However, the knowledge about both the cellular changes and the etiological factors leading to HCC, including chronic liver disease, viral hepatitis (hepatitis B virus and hepatitis C virus) infections, hemochromatosis, abuse of alcohol, and exposure to hepatic carcinogens and aflatoxins (Schafer, D.F. & Sorrell, M.F., Lancet 353 (1999), 1253-7; Evans, A.A. et al., Cancer Epidemiol. Biomarkers Prev. 11 (2002), 369-76; Block, T.M. et al., Oncogene 22 (2003), 5093-107), has increased in recent years. Besides HBV and HCV, aflatoxins, which are produced by a mold that is a contaminant of nuts (most commonly peanuts), grains, and beans, have been implicated as a major risk factor for causing hepatocellular carcinoma. Although virtually non-existent in the United States, aflatoxins are common in other parts of the world and often contaminate food.

[0005] Considerable efforts have been devoted to establishing a prognostic model for HCC by using clinical information and pathological classification to provide information at diagnosis on both survival and treatment options. Liver cirrhosis accompanies at least 70% of hepatocellular carcinomas worldwide (Schafer, D.F. & Sorrell, M.F., Lancet 353 (1999), 1253-7; Llovet J.M. et al., Lancet 362 (2003), 1907-17) and cirrhosis, together with vascular invasion, are the clinical prognostic markers being presently used to predict recurrent disease.

[0006] Although the prognostic and treatment-predictive markers presently in use in HCC management provide valuable information, they are not fully adequate in identifying patients that require additional therapy or in determining the most optimal therapy for the individual patient. Hence, the ability to predict the biological behavior of HCC would enable selection of the optimum treatment and follow-up strategies.

[0007] Extensive studies have been made to increase the understanding of the clinicopathological features as well as the molecular aspects of HCC in order to

improve the management of patients, as the prognosis of HCC continues to be dismal. However, the heterogeneous nature of human HCC has limited the usefulness of conventional clinicopathological features at diagnosis such as the status of cirrhosis and vascular invasion for both treatment and prediction of disease outcome. Moreover, patients with similar tumor stage disease (tumor-node-metastasis (TNM) classification) can often have very different disease outcomes. This uncertainty inherent in the conventional classification methods can be attributed to the incomplete exploration of the biological differences of tumors in patients belonging to identical classification subgroups. Thus, it is still not possible to provide an accurate and reliable clinical prediction for disease recurrence associated with patients with poor prognosis. As the phenotype of a tumor reflects only the underlying genetic alterations, direct studies of these alterations on a molecular level may be beneficial to assess the biological potential of a tumor.

[0008] Recently, microarray technologies have been successfully used to predict clinical outcome and survival as well as classify different types of cancer. These microarray technologies have also been applied in many studies to define global gene expression patterns in primary human HCC as well as HCC-derived cell lines in an attempt to gain insight into the mechanisms of hepatocarcinogenesis. These studies have identified subgroups of HCC that differ according to etiological factors, mutations of tumor suppressor genes, rate of recurrence, and intrahepatic metastasis, as well as novel molecular markers for HCC diagnosis. However, most of these studies identified genes that are associated with limited aspects of tumor pathogenesis, and thus failed to create molecular prognostic indices that could be applied to the HCC patient population in general.

[0009] Kurokawa et al. (Journal of Hepatology Vol. 41 (2004) 284-291) investigated genes useful to predict early intrahepatic recurrence of HCC. 92 significant genes were identified on the basis of a gene expression profile analysis of 60 patients using a PCR-based array system.

[0010] Iizuka et al. (The Lancet Vol. 361 (2003) 923-929) investigated mRNA expression profiles in tissue specimens from a training set, comprising 33 patients with HCC, with high-density oligonucleotide microarrays representing about 6,000

genes. On the basis of this training set, Iizuka et al. developed a predictive system consisting of 12 genes with the Fisher linear classifier. The predictive performance of the system thus constructed predicted early intrahepatic recurrence or non-recurrence for patients with HCC much more accurately than a support vector machine (SVM)-based system on a blinded set of samples from newly enrolled patients.

[0011] Ye et al. (Nature Medicine Vol. 9 (4) (2003), 416-423) generated a molecular signature that can classify metastatic HCC patients and identified genes that are relevant for metastasis and patient survival using a supervised machine-learning algorithm. In this study, Osteopontin was identified as a lead gene in the signature.

[0012] Lee et al. (Hepatology Vol. 40 (3) (2004), 667-676) analysed the global gene expression pattern of 91 human HCCs to define the molecular characteristics of the tumors and to test the prognostic value of the expression profiles. Using unsupervised classification methods, they revealed two distinctive subclasses of HCC that are associated with patient survival. The genes most strongly associated with survival were identified by using the Cox proportional hazards survival analysis. This approach identified a limited number of genes, the expression profile of which provided a method to predict the length of survival of the HCC-afflicted patients.

[0013] However, all the afore-mentioned methods have the disadvantage that the set of gene markers used provides only for a certain degree of reliability and accuracy for the prediction of HCC recurrence and does not satisfy the need for a highly accurate method to assess the recurrence of HCC.

[0014] Thus, a method to more accurately assess and stratify patients with different risks of disease recurrence at diagnosis will be extremely beneficial for the clinical management of HCC.

Summary of the invention

[0015] The present invention provides a method of predicting the recurrence of hepatocellular carcinoma based on a combination of clinicopathological information and the determination of novel molecular markers.

[0016] In a first aspect, the invention relates to a method of diagnosing and/or determining the risk of HCC recurrence in a subject comprising determining at least two clinicopathological factors.

[0017] In another aspect, the invention relates to the identification and selection of novel molecular markers for recurrent hepatocellular carcinoma (HCC), which are differentially expressed in recurrent hepatocellular carcinoma and which are useful to predict the risk of HCC recurrence with high predictive accuracy and to choose an adequate therapy.

[0018] Thus a method is provided that allows predicting hepatocellular carcinoma recurrence in a subject comprising obtaining a biological sample from the subject and detecting the level of expression of two or more differentially expressed nucleic acid molecules in the sample.

[0019] In another aspect, the invention relates to a kit for the prediction of hepatocellular carcinoma recurrence. This kit provides reagents to determine the expression level of two or more nucleic acid molecules in a biological sample. These nucleic acids can be selected from the group of the nucleic acid molecules that are set forth in Table 3 or Table 4 or complements, fragments, variants, or analogs thereof.

[0020] In a further aspect, the invention provides a polynucleotide comprising the nucleotide sequence of SEQ ID NO: 2.

Brief description of the figures

[0021] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the drawings, in which:

[0022] Figure 1 illustrates a grouping of recurrence and recurrence-free cases according to four combinations of the two risk factors cirrhosis and vascular invasion. Abbreviations used: VI, vascular invasion; NI, non-invasion; Ci, cirrhosis; NCi, non-cirrhosis; R, recurrence; NR, non-recurrence.

[0023] Figure 2 shows a comparison of the expression pattern of 5 up-regulated genes and 2 down-regulated genes selected from 57-gene set by

quantitative real-time PCR and microarrays. The upper panels illustrates the expression pattern as analyzed by Affymetrix GeneChip analysis, whereas the lower panels relate to the expression pattern as determined by real-time PCR analysis.

[0024] Figure 3 illustrates a hierarchical clustering of the 57 genes that gave the best prediction for recurrence and non-recurrence cases within the group of HCC patients with cirrhosis and without vascular invasion and the group of HCC patients without cirrhosis and with vascular invasion. **A.** List of the 22 genes with the highest level of expression in recurrent HCC samples. **B.** List of subset of 7 genes with distinctive lower level of expression in non-recurrent HCC samples. **C.** List of 16 genes that are down-regulated in the recurrent HCC samples. **D.** List of 12 genes with distinctively higher level of gene expression in the non-recurrent HCC group. Abbreviations used: R, HCC patients with recurrence; NR, HCC patients without recurrence; ST, matched surrounding normal tissue; NN, normal liver of colon patients with metastases to liver; red color represents up-regulation; black color represents no change; green represents down-regulation.

Detailed description of the invention

[0025] The present invention provides a method that allows the combination of clinicopathological information and novel molecular markers to facilitate better prediction of the recurrence of hepatocellular carcinoma.

[0026] Thus, in a first aspect, the invention relates to a method of diagnosing and/or determining the risk of HCC recurrence in a subject comprising determining at least two clinicopathological factors. In one embodiment, the clinicopathological factors that are determined in a method of the invention are cirrhosis and vascular invasion.

[0027] In another embodiment (which can be carried out independently or together with using at least two clinicopathological factors as mentioned above), the invention relates to the identification and selection of novel molecular markers for recurrent hepatocellular carcinoma (HCC), which are differentially expressed in recurrent hepatocellular carcinoma and which are useful to predict the risk of HCC

recurrence with high predictive accuracy and to choose an adequate therapy. In some embodiments of the invention, a combination of these molecular markers is utilized in a method to predict the risk of HCC recurrence.

[0028] One aspect of the invention thus relates to nucleic acid molecules that are associated with hepatocellular carcinoma and more particularly to genes differentially expressed between recurrent and non-recurrent hepatocellular carcinoma. These nucleic acid molecules can in one embodiment be genomic DNA.

[0029] Thus, a method is provided that allows predicting hepatocellular carcinoma recurrence in a subject comprising obtaining a biological sample from the subject and detecting the level of expression of two or more genes or expression products thereof in the sample.

[0030] In one embodiment of the invention, these at least two genes are selected from the group of genes that are set forth in Table 3 or expression products, complements, fragments, variants, or analogues thereof. Such a combination of two or more of these molecular markers is utilized to predict the risk of HCC recurrence. Any combination of at least two genes can be used for this prediction. For example, in some embodiments, a combination of at least 5, 7, 10, 12, 16, 20, 30, 40, 50, 55, 56, or of all 57 of the genes set forth in Table 3 is used to predict the risk of HCC recurrence. The HCC recurrence-associated nucleic acid molecules of the invention thus include, but are not limited to, the following genes or the expression products thereof: PARD3, MPHOSPH9, CUL4B, CETN2, USH1C, PTPN11, ARL5, SMURF2, SH3GLB2, RACGAP1, DNAJC10, KCNK1, GSTM3, SSR3, FLJ13105, SYNGR2, PRL10A, FLJ23749, FLJ11196, R3HDM, FLJ35036, KIAA0924, LOC159090, FLJ11016, LOC285550, NARG2, FAM33A, FLJ22198 (clone HRC01218) (SEQ ID NO: 1), transcribed locus moderately similar to XP_517655.1 (similar to KIAA0825) (SEQ ID NO: 2), INSIG1, SKP2, ETS2, OASL, CDC42SE1, CSNK1G3, CYP17A1, CDO1, CNGA1, NUDT9, SLC16A5, SARA2, ID2, RNF130, MAFB, ELL2, PTP4A1, PGGT1B, HEXB, SLC27A2, AGPAT3, GCH1, SPACA3, AS3MT, TIGD2, KIAA0676, DKFZp434H2226, and MARVELD2. In one embodiment, the gene expression profile of each of the genes RACGAP1, KCNK1, SMURF2, USH1C, GSTM3,

CNGA1, and INSIG1 is determined. In this context it is noted that the nucleotide sequence of SEQ ID NO: 2 represents a gene that has not been known so far and as shown herein, this gene is involved in the development of HCC, including reoccurrence of HCC.

[0031] In another embodiment, the genes the expression of which is to be detected in a method of the invention are selected from the group of the genes set forth in Table 4 or expression products, complements, fragments, variants, or analogs thereof. Also in this case, any combination of at least two of these genes can be used for this prediction. In one embodiment of the invention, a combination of at least 5, 10, 15, 20, 30, 40, 50, 60, or 65 of the genes set forth in Table 4 is utilized. Further HCC recurrence-associated nucleic acid molecules of the invention include, but are not limited to, the genes set forth in Table 4 or the expression products thereof. Thus HCC recurrence-associated nucleic acid molecules that are used in the invention include, but are not limited to: KIF2, MCM6, RRM1, APEX1, CD24, CD58, SOCS5, PIK4CA, TRAPPC4, RAN, AP3B1, MTF2, ZNF207, SFRS2, UBE2L3, RPS16, ADSL, PAPD1, CALU, PSF1, ANP32B, AR, LECT2, MASP2, TGFBR3, FLJ38991, CYP8B1, CYP2C9, CYP3A43, CYP2C19, MGC25181, COX17, SFXN5, RAB33B, PROZ, NR1I2, SMARCA2, HRSP12, HXMA, LOC123876, ALDH6A1, DHRS4, ALAS1, DCXR, GSTZ1, MTHFD1, MTCBP-1, ADH4, CPN2, DPYS, GLYAT, RCI1, NUDT6, MGC15875, C1orf168, LOC283666, C8orf40, and C10orf65.

[0032] As mentioned earlier, the methods that use the detection of the level of expression of two or more genes can be used in combination with the determination of at least two clinicopathological factors. In one embodiment, the determined clinicopathological factors are cirrhosis and vascular invasion. In one embodiment of the invention, the risk of recurrence of hepatocellular carcinoma is determined on the basis of cirrhosis and vascular invasion as clinicopathological factors. If desired, the expression level of two or more of the genes depicted in Table 3 or Table 4 is subsequently determined in order to increase the accuracy of the prediction. For example, if a patient that has undergone cancer treatment is found to have both clinicopathological cirrhosis and vascular invasion (which are included in group 1, cf. Experimental Section and Fig.1), this information alone is

usually sufficient to predict the probability of recurrence and for the group of patients to thus chose an appropriate treatment (given the more than 80 % certainty of reoccurrence of HCC). However, in order to further enhance the accuracy of the prediction or to have an independent verification the expression level of at least two of the genes shown in Table 4 can be tested. Likewise, if the patient shows neither cirrhosis and vascular invasion (see patient group 4 in Figure 1), the expression profile of two or more genes of Table 4 can be used to independently verify this result (that no recurrence is likely). If a patient has only either one of cirrhosis and vascular invasion (patient groups 2 and 3 in Figure 1) then the further medical treatment will typically only be decided after having determined the expression profile of two or more of the genes listed in Table 3. It should be noted in this conjunction, that when the expression profile of all 57 genes indicated in Table 3 was determined, the disease recurrence could be predicted with up to 88% accuracy in HCC patients.

[0033] The gene expression profile of the genes set forth in Table 3 and Table 4, respectively, or the expression level of any of their expression products, complements, fragments, variants, and analogs thereof, can be determined using any method known to the skilled person. In one embodiment of the invention, the expression is determined in a DNA array. In another embodiment of the invention, the expression is determined in a DNA microarray. In another aspect of the invention, the expression is determined using quantitative PCR analysis. In one embodiment, the PCR analysis is carried out using real time PCR.

[0034] The invention also provides a composition including an addressable collection of two or more nucleic acid molecules or polypeptides encoded by these nucleic acid molecules that are differentially expressed in recurrent hepatocellular carcinoma. These at least two nucleic acid molecules can essentially be the expression products of the genes set forth in Table 3, Table 4 or complements, fragments, variants, or analogs thereof, or any subset thereof. The nucleic acids or corresponding polypeptides may be attached to a solid support. The compositions may be used in the preparation of a medicament for the therapy or prevention of recurrent hepatocellular carcinoma. In one embodiment of the invention, the compositions are for diagnostic purpose.

[0035] In other aspects, the invention provides a solid support including two or more nucleic acid molecules or polypeptides encoded by these nucleic acid molecules that are differentially expressed in recurrent hepatocellular carcinoma tissue, where the nucleic acid molecules consist essentially of the expression products of the genes set forth in Table 3, Table 4 or complements, fragments, variants, or analogs thereof. The nucleic acid molecules may consist essentially of all the expression products of the genes set forth in Table 3 or any subset thereof. Or the nucleic acid molecules may consist essentially of all the expression products of the genes set forth in Table 4 or any subset thereof. The polypeptides may consist essentially of all the polypeptides encoded by the nucleic acid molecules expressed by the genes set forth in Table 3, Table 4 or any subset thereof. The nucleic acid molecules or the polypeptides may be covalently or non-covalently attached to the solid support, e.g. in a microarray.

[0036] In another aspect, the invention relates to a kit for the prediction of hepatocellular carcinoma recurrence, wherein said kit provides the reagents or means to determine the expression level of two or more genes set forth in Table 3 or expression products, complements, fragments, variants, or analogues thereof in a biological sample. Alternatively, the kit can provide reagents to determine the expression level of two or more genes set forth in Table 4 or expression products, complements, fragments, variants, or analogues thereof in a biological sample. The kit can of course also include reagent for determination of the expression level of at least two genes that are shown in Table 3 and Table 4.

[0037] In one embodiment of the invention, the kit provides means/reagents to determine a gene expression profile of all the genes set forth in Table 3 and/or of all the genes set forth in Table 4, or expression products, complements, fragments, variants, or analogues thereof in a biological sample.

[0038] A kit of the invention comprises one or more oligonucleotides. These oligonucleotides can be either used as probes or as amplification primer for at least two of the nucleic acids given in Table 3 and/or Table 4. The kit may contain multiple oligonucleotides to facilitate detection of all or any subset of the nucleic acids set forth in Table 3 and/or Table 4. If these oligonucleotides are oligonucleotide probes, the probes may be complementary to one of the nucleic

acids set forth in Table 3 and Table 4, respectively. In this context, "complementary" refers to the ability to hybridize to the desired target nucleic acid under high stringency conditions. The oligonucleotide probes may be labeled. The label can, for example, be a radioactive label, a fluorescence label, a chemoluminescence label, an affinity label, an enzymatic label or any other suitable label. The affinity label may be reagent that is commonly used in the detection of nucleic acids. Examples of such as reagent include, but are not limited to biotin or digoxigenin. The label serves to facilitate detection of the target nucleic acid. The probes may also be immobilized on a substrate. The substrate may be a DNA microchip. If the kit contains amplification oligonucleotides, the amplification oligonucleotides are suitable to amplify the desired target nucleic acid. Thus, the amplification oligonucleotides are complementary to one or more of the nucleic acids set forth in Table 3 and Table 4, respectively. If the kit is an amplification kit, the amplification primers can be designed for any desired amplification method known to the skilled person. The kit oligonucleotides can be of any length, for example, are up to about 100 nucleotides in length, up to about 60 nucleotides or up to about 30 nucleotides in length.

[0039] The sample analyzed in the present invention may be liver tissue, and may be suspected of being cancerous, or may be non-cancerous. More particularly, the tissue may include a liver cancer cell, more specifically a hepatocellular carcinoma cell.

[0040] The methods of the present invention may further include comparing the level of expression of two or more or any subset of the genes set forth in Table 3, Table 4 or expression products thereof in a sample with control samples that may be derived from non-cancerous tissue, recurrent or non-recurrent hepatocellular carcinomas. Differential expression of the genes or expression products thereof may be indicative for the risk of HCC recurrence, or of the efficacy of the HCC therapy. The methods of the invention may be high throughput methods.

[0041] The control samples used in the invention may be derived from normal liver tissue or non-cancerous tissue surrounding the tumour. The control samples may be obtained from the same subject the test sample is obtained from or from

another subject. The subject having hepatocellular carcinoma may be a mammal, for example a human, an ape, a rat, a horse, a dog or a cat.

[0042] The invention also relates to a method to identify compounds that modulate the expression of any of the genes indicated in Table 3 and Table 4, respectively, and/or interacting with the expression products thereof or the polypeptides encoded by these expression products. The compound the ability to modulate the expression of any of these genes is investigated can be, but is not limited to, a siRNA, a ribozyme, a protein, a peptide, a small organic molecule or the like. The term "organic molecule" as used herein preferably denotes an organic molecule comprising at least two carbon atoms, but preferably not more than seven rotatable carbon bonds, having a molecular weight in the range between 100 and 2000 Dalton, preferably 1000 Dalton, and optionally including one or two metal atoms. In line with the above, the genes that are identified as markers for the prediction of the risk of reoccurrence of HCC or the encoded proteins can also be used as targets for the development of new anti-HCC drugs. In one embodiment, the gene that is used in such an identification methods is one or more of the genes RACGAP1, KCNK1, SMURF2, USH1C, GSTM3, CNGA1, and INSIG1. These identification methods or screening assay can be carried out as using any known method that is suitable for this purpose and can for example include, determining the expression level of a gene described herein when being brought into contact with the compound to be tested. The method can also comprise contacting the encoded polypeptide with a compound of interest and determining whether complex formation takes place, i.e. whether the compound of interest and the chosen protein interact with each other. Any of these identification methods can be implemented as a high-throughput screening method, if desired.

[0043] In the context of this invention, the term "prediction" or "predicting" refers to a highly accurate and reliable diagnosis method for evaluating the risk of HCC recurrence.

[0044] The term "HCC" refers to hepatocellular carcinoma and "HCC recurrence associated genes" refers to genes differentially expressed in HCC recurrence. Said genes include, but are not limited to, those which are listed in Table 3 of the present invention.

[0045] The term “differentially expressed” refers to gene expression that is altered between a sample of a recurrent HCC and a non-recurrent HCC. Said alteration can be an up-regulation as well as a down-regulation.

[0046] A “composition” as used herein includes a plurality of the nucleic acid molecules described herein, including complements, fragments, analogues, and variants thereof. A composition as used herein may also refer to a plurality of polypeptides encoded by the nucleic acid molecules described herein, and complements, analogues, variants, and fragments thereof. The composition may include any combination of the nucleic acid molecules set forth in Table 3.

[0047] “Complements” refers to nucleic acid molecules that contain a sufficient number of nucleotides capable of forming Watson-Crick base pairs to produce a region of double-strandedness between two nucleic acids. Thus, a nucleic acid is a complement of another nucleic acid molecule, if it hybridizes to it under high stringency conditions. In some embodiments of the invention, a complement of the nucleic acid molecules set forth in Table 3 is complementary over the full length of the nucleic acid molecule.

[0048] A “fragment” may be any portion of a nucleic acid molecule or protein disclosed herein that is capable of being differentially expressed or detected in an assay or screening method according to the invention.

[0049] A “variant” is a nucleic acid molecule that is a variation of a nucleic acid molecule or an expression product thereof, for example a splice variant. Allelic variants have a very high sequence similarity and often only deviate in a small number of bases. A change of a single base as a result of substitution, insertion, or deletion (single nucleotide polymorphism) is encompassed by the term “variant”.

[0050] An “analog” is a nucleic acid molecule or polypeptide that is chemically modified. Analogs may contain non-traditional bases or base analogs and non-traditionally amino acids and amino acid analogs respectively. Analogs generally retain the biological activity of the natural occurring molecule but may confer further advantages, such as a higher stability, enhanced activity and the like.

[0051] The term “nucleic acid” or “nucleic acid molecule” refers to both RNA (plus and minus strands) and DNA, including cDNA, genomic DNA, and synthetic DNA. The nucleic acid may be single or double stranded, and, if single-stranded,

sense or antisense. A nucleic acid may be any chain of two or more nucleotides, including naturally occurring nucleotides as well as synthetic nucleotides, nucleotide analogs or derivatives.

[0052] A "protein" or "polypeptide" is a chain of two or more amino acids, including naturally occurring as well as synthetic amino acids, amino acids analogs and derivatives. A protein or polypeptide of the invention may be post-translationally modified.

[0053] A "gene" is a part of the genomic DNA, encoding for a polypeptide.

[0054] An "expression product" is a product of gene expression and comprises the transcription product, i.e. a mRNA, as well as the translation product, i.e. a polypeptide.

[0055] "Expression profile" or "expression profiling" refers to the determination of the amount of different mRNAs in a sample.

[0056] The inventions illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

[0057] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter

from the genus, regardless of whether or not the excised material is specifically recited herein.

[0058] Other features and advantages of the invention will become apparent from the following examples and the claims. It is to be understood that the examples are of illustrative purpose only and should not be construed to be limiting the scope of the present invention.

Examples

[0059] There have been several recent reports on applying molecular gene profiling to segregate HCC from normal tissues (Xu, X.R. et al., Natl Acad Sci U.S.A 2001, 98(26), 15089-94; Kim, B.Y. et al., Biochim Biophys Acta 2004, 1739(1), 50-61) and to discriminate patients according to etiological factors (Okabe, H. et al., Cancer Res. 2001, 61, 2129-37) and disease state (Chen, X. et al., Mol. Biol. Cell 2002, 13, 929-39, Chuma, M. et al., Hepatology 2003, 37(1), 198-207). Lately, molecular signatures that are associated with pre-malignant lesions including cirrhosis (Kim, J.W. et al., Hepatology 2004, 39(2), 518-27, Zindy, P. et al., FEBS Letters 2005, 579, 95-9), and are capable of evaluating metastatic or recurrent potentials (Lee et al., Hepatology Vol. 40 (3) (2004), 667-676, Ye et al., Nature Medicine Vol. 9 (4) (2003), 416-423, Iizuka et al., The Lancet Vol. 361 (2003) 923-929, Kurokawa et al., Journal of Hepatology 41 (2004) 284-291) have been uncovered. However, the molecular information that has accumulated is still far from sufficient to accurately predict disease recurrence after curative hepatectomy.

[0060] When the HCC patients studied were arbitrarily divided into four sub-groups by considering jointly the presence or absence of cirrhosis and vascular invasion, it was observed that HCC patients who had vascular invasion and cirrhosis at diagnosis (group 1) are most likely to recur (83%), while only a minority (12%) of HCC patients who had no vascular invasion and were non-cirrhotic at diagnosis (group 4) had recurrence (Figure 1). Therefore, the combined clinicopathological features of vascular invasion and the status of cirrhosis were apparently adequate to predict disease recurrence for HCC patients in group 1 (83% accuracy) and group 4 (88% accuracy) (Figure 1). After establishing that

vascular invasion ($p = 0.018$) and cirrhosis ($p = 0.015$) are the only two clinical parameters that are significantly associated with the recurrence of HCC, it was attempted to employ these two clinicopathological features to predict recurrence disease in the HCC patients studied. However, the combined clinicopathological features of vascular invasion and the status of cirrhosis were not sufficient to predict recurrence for HCC patients in groups 2 and 3. Group 2 comprised of 12 patients with invasion but without cirrhosis, and group 3 had 26 cirrhotic patients without vascular invasion (Figure 1). The rate of recurrence for these two groups of HCC patients, using the combined clinicopathological features of vascular invasion and the status of cirrhosis, yielded an unsatisfactory 50% (Figure 1). We then focused our subsequent experiments by studying genes that were differentially expressed between recurrent and non-recurrent tissues of HCC patients in conjunction either with vascular invasion or cirrhosis using Affymetrix arrays in order to identify a molecular gene signature that could improve the accuracy to predict disease recurrence.

[0061] Firstly, statistical analyses were conducted using the SVM classifier to separately compare the molecular profiling data obtained for patients with cirrhosis, vascular invasion, or recurrence. However, we were unable to obtain a satisfactory gene signature that could accurately predict the recurrence of HCC disease. The accuracy of prediction obtained for all the groups were around 70% (data not shown). Next, genes were studied that were differentially expressed between patients that had recurrence ($n=11$) to patients that did not have recurrence ($n=12$) in both groups 2 and 3 using Affymetrix arrays. An optimized gene cluster, derived from statistical analyses, consisted of 61 probe sets (57 genes) was identified from the 130 top ranked probe sets with median fold change greater than 1.5 (see Materials and Methods). Using three different classifier algorithms including Support Vector Machine (SVM), Sparse Linear Discriminant Analyst (SLD), and K Nearest Neighbour (KNN) ($K=3$), the accuracy obtained for estimating the training set of HCC patients ranged from 96% to 100% while the accuracy obtained using this gene cluster to predict recurrence disease for an independent set of 8 HCC patients with recurrence and 7 HCC patients without recurrence approached 87% (sensitivity=88%; specificity=86%), (Table 2). This

61 probe sets (57 genes) is tentatively designated as the HCC recurrence-associated molecular signature. This HCC recurrence-predictive molecular signature thus provides a method to predict the recurrence rate for HCC patients with an overall accuracy of more than 80%.

[0062] The HCC recurrence-associated molecular signature comprises of 57 genes that are associated with a wide variety of cellular functions, including cell growth and maintenance, DNA replication and cellular metabolism, transcription and protein processing, cellular signaling, transport, immune regulators, and apoptosis (Table 3). These 57 genes could be conveniently clustered into four distinct groups (Figure 3). Figure 3A consists of 23 probe sets (22 genes) that are up-regulated in the recurrent HCC patients of groups 2 and 3 and not in the non-recurrent HCC patients and normal controls. Figure 3B shows 7 genes that are not only up-regulated in recurrent HCC but also with distinct decreased expression in the non-recurrent HCC patients of groups 2 and 3. Figures 3C and 3D, on the other hand, show genes that are specifically down-regulated in the groups 2 and 3 HCC patients with recurrence (19 probe sets and 16 genes) and distinctly up-regulated in HCC without recurrence (12 genes) respectively. Expression in the distal normal surrounding tissues of cancer (ST), and normal control liver tissues (NN) were also included for comparison.

[0063] To further validate the differential expression pattern of the genes in the HCC recurrence-predictive molecular signature, quantitative real-time PCR experiments were conducted to confirm the level of gene expression of those genes detected by Affymetrix gene chips. A total of 17 selected genes (11 up-regulated and 6 down-regulated) chosen from the HCC recurrence-predictive molecular signature with the highest median fold changes (ascertained from expression studies with the Affymetrix HG-U133 probe arrays) were tested. Half of the total samples used in quantitative real-time PCR (7 NN, 7 ST, 10 NR and 11 R) were also used in Affymetrix gene chips analysis. Results obtained by real-time PCR analyses correlated well with those of the Affymetrix gene chips for all the 17 genes tested. Similar pattern of up-regulation and down-regulation for all the genes tested could be validated by real-time PCR. Figure 2 showed the results obtained for the 7 genes validated by quantitative real-time PCR and that

were shown to be statistically significant ($p \leq 0.05$) differed between recurrent and non-recurrent HCC liver samples. Similar to results obtained with the Affymetrix gene chips, RACGAP1, KCNK1, SMURF2, USH1C and GSTM3 were up-regulated while CNGA1 and INSIG1 were down-regulated in the recurrent HCC liver samples (Figure 2).

[0064] Within the HCC recurrence-associated gene signature of the above-mentioned 57 genes, there are 22 genes with expression levels, which are specifically elevated in the recurrent HCC samples (Figure 3).

[0065] The top most up-regulated gene in recurrent HCC samples, as determined by both Affymetrix gene chips and real-time PCR assays, was USH1C. USH1C encodes the gene product harmonin, a protein containing a PDZ domain. PDZ domains are modular protein interaction domains that play a role in protein targeting and the assembly of large protein complexes involved in signaling or subcellular transport. In these protein-protein complexes harmonin acts as a scaffold protein binding these USH1 molecules via its PDZ domains. Mutations in USH1C cause Usher syndrome type 1C, an autosomal recessive disorder characterized by congenital sensorineural deafness, vestibular dysfunction, and blindness. This was the first mutation in a PDZ-encoding gene linked to a human disease. However, at present, it is not clear how PDZ domain proteins and its ligands contribute to HCC.

[0066] Another gene which is also consistently up-regulated in the recurrent HCC samples is the Rac GTPase activating protein 1. Rho GTPases, which include Rho, Rac and CDC42 as the most prominent members, play pivotal roles in regulating the organization of the actin cytoskeleton, which is crucial for cell motility, cell-cell and cell-matrix adhesion, cell migration, chemotaxis, and malignant transformation in many cell types. The Rho GTPases are responsible for the regulation of many downstream kinases, such as the PAK kinases, and Rho and Rac have been reported to be cooperatively involved in both the invasion and the related morphological changes of MM1 cells.

[0067] Moreover, casein kinase that regulates the cytoskeletal organization through small GTPases via the Wnt signaling pathway is down-regulated in the recurrent HCC samples.

[0068] The expression of the detoxification enzyme cytochrome P450 and enzymes involved in metabolism, such as AS3MT and HEXB, all of which are predominantly expressed in differentiated hepatocytes, is also down-regulated in recurrent HCC samples. The suppression of these genes could reflect tumor de-differentiation following the progression of malignancy.

[0069] On the other hand, another gene that is noticeably up-regulated in recurrent HCC samples is Acheron (FLJ11196), a DNA-binding protein.

[0070] There are 7 genes with distinctive lower expression in non-recurrent HCC patients as compared to recurrence groups and these include FLJ11196, FLJ23749, 229926_at, FLJ13105, ARL5, LOC285550, and NARG2 (Figure 3B).

[0071] There are 12 genes that are markedly up-regulated in the non-recurrent HCC patients and these include CYP17A1, DKFZp434H2226, SKP2, OASL, AS3MT, MARVELD2, AGPAT3, KIAA0676, HEXB, CSNK1G3, SLC16A5, and RNF130 (Figure 3D).

[0072] Many of these genes are uncharacterized genes of unknown functions (FLJ23749, 229926_at, FLJ13105, ARL5, LOC285550, DKFZp434H2226, and KIAA0676).

[0073] There have been several HCC gene signatures reported in the literature, mostly based on single individual clinical parameter, to predict survival, intrahepatic metastasis and early recurrence of HCC (Lee *et al.*, Hepatology Vol. 40 (3) (2004), 667-676; Ye *et al.*, Nature Medicine Vol. 9 (4) (2003), 416-423; Iizuka *et al.*, The Lancet Vol. 361 (2003) 923-929; Kurokawa *et al.*, Journal of Hepatology 41 (2004) 284-291). To compare those reported gene signatures with the identified HCC recurrence-predictive molecular signature, the genes of those gene signatures reported in the literature have been transformed to be compatible with the used software platform. Specifically, there were 621, 231, 29 and 37 reported probe sets representing 406, 153, 12 and 20 genes from the studies of Lee *et al.* (Hepatology Vol. 40 (3) (2004), 667-676), Ye *et al.* (Nature Medicine Vol. 9 (4) (2003), 416-423), Iizuka *et al.* (The Lancet Vol. 361 (2003) 923-929) and Kurokawa *et al.* (Journal of Hepatology 41 (2004) 284-291), respectively. When these gene sets were employed to predict the same cohort of 70 HCC patients in our study using similar reported criteria to include early recurrence of less than

one year for Lizuka *et al.* and recurrence of less than 2 years as specified by Kurokawa *et al.*, we noted that the overall accuracy scores were 65% for Lee *et al.*'s gene set, 57% for Ye *et al.*'s gene set, 61% for Lizuka *et al.*'s gene set, and 48% for Kurokawa *et al.*'s gene set.

[0074] It appears therefore that the current gene set (as set forth in Table 3), derived from two combined clinicopathological features of cirrhosis and vascular invasion, gave the best accuracy in predicting recurrent disease.

Example 1

Sample preparation

[0075] Cancerous and corresponding distal non-cancerous tissues were obtained from 70 patients who underwent partial hepatectomy as treatment for HCC. Tissues were immediately snap-frozen and stored in liquid nitrogen. Frozen sections of the tumor samples were stained with hematoxylin and eosin and evaluated by qualified pathologists.

[0076] Linked clinical and histopathological data were collected from medical records for all the patients who contributed tumor specimens and were rendered anonymous to protect patient confidentiality. Biopsies of histologically normal liver tissues of ten colorectal cancer patients who have liver metastasis were used as reference normal liver tissues.

[0077] The clinicopathological information of the human HCC samples employed in this study is summarized in Table 1. Male patients were predominant with its ratio to female was 4.8:1 (82.9% males compared to 17.1% females). Of all the patients studied, 78.6% were HBV carriers, 4.3% were HCV carriers and 17.1% were negative for HBV and HCV infection. The tumor of thirty-seven patients (52.9%) was either partially or completely encapsulated, 54.3% (n=38) of patients had cirrhosis, and the majority of the patients (84.3%, n=59) had single nodule. The median size of all resected tumor was 4.9 cm (range: 1.2 – 17cm) and the median α -fetoprotein (AFP) level was 25.3 ng/ml (range: 1.2 – 70700 ng/ml). According to the Edmondson grading system, more than half of the cases (58.6%; n=41) were classified as grade 2, 14.3% were classified as grade 1 and 22.9%

were classified as grade 3. Vascular invasion occurred in 34.3% (n=25) of the cases and the median time of recurrence was 4 months (range: <1 – 30 months).

[0078] Based on these commonly used clinical prognostic markers, it was determined which of them are significantly associated with the recurrence of HCC. Table 1 summarized the results of these analyses and ascertained that vascular invasion ($p = 0.018$) and cirrhosis ($p = 0.015$) are the only two observable clinical parameters that are significantly associated with the recurrence of HCC. Although patients with recurrence apparently had larger tumor size (median = 6 cm) and elevated serum concentration of AFP (median = 40.75 ng/ml) compared to those without recurrence (Table 1), no other clinicopathological factors being significantly associated with the recurrence of HCC were identified.

Table 1. Clinicopathological factors in recurrence and non-recurrence cases of 70 HCC patients.

	Recurrence (n=33)	Non-recurrence (n=37)	<i>P</i> value ^a
Sex (men/women)	30/3	28/9	0.091
Age (median) (years)	61	61	0.684
Viral infection (HBV / HCV / non-B non-C)	25/2/6	30/1/6	0.755
Capsulation (yes / no)	16/17	21/16	0.689
Tumor size (median) (cm)	6	4.2	0.386
AFP level (median) (ng/ml)	40.75	15.1	0.333
{low / medium / high}	{10/13/9}	{17/13/7}	
Lesion (single / multiple nodules)	27/6	32/5	0.592
Histological grading (G1/G2/G3/G2 to G4)	3/18/11/1	7/23/5/2	0.205
Cirrhosis (yes / no)	23/10	15/22	0.015
Invasion (yes / no)	16 /17	8/29	0.018

Abbreviations used: HBV, hepatitis B virus; HCV, hepatitis C virus; non-B non-C, negative for both HBV and HCV antigen; low, less than 10 ng/ml; medium, 10 to 300 ng/ml; high, more than 300 ng/ml; G1, well differentiated; G2, moderately differentiated; G3, poorly differentiated; G4, extremely poor differentiation; ^a, *P* value were calculated using univariate analysis.

Isolation of total RNA and Affymetrix gene chips experiments

[0079] Total RNA was extracted from frozen HCC biopsies using Trizol as described in the manufacturer protocol. All purified RNA samples were stored in RNAssecure™ (Ambion Inc., Austin, TX, USA) at -80°C.

[0080] Total RNA (5 µg) was reversibly transcribed to synthesize first-strand cDNA with superscript II RNase H-reverse transcriptase (Invitrogen Life Technologies, Carlsbad, CA, USA). The cDNAs was purified by phase-lock gel (Eppendorf AG, Hamburg, Germany) and employed as a template for *in vitro* transcription with RNA transcript labeling kit (Enzo Diagnostics, Farmingdale, NY, USA) to produce amplified biotin-labeled antisense RNA (cRNA), which was subsequently purified with Qiagen RNeasy kit (Qiagen GmbH, Hilden, Germany). The purified cRNA was fragmented and 15 µg was used to hybridize to human HG-U133A and HG-U133B oligonucleotide probe arrays (Affymetrix, Santa Clara, CA, USA) as described previously (Linn, Y.C. et al., Comparative gene expression profiling of cytokine-induced killer cells in response to acute myloid leukemic and acute lymphoblastic leukemic stimulators using oligonucleotide arrays, *Exp. Hematol.* 33(6) (2005), 671-81 ; Tan, M.G. et al., Cloning and identification of hepatocellular carcinoma down-regulated mitochondrial carrier protein, a novel liver-specific uncoupling protein, *J. Biol. Chem.* 279(43) (2004), 45235-44). The probe arrays were washed according to the EU-GE-WS2v4 fluidics protocol with Affymetrix GeneChip fluidics station 400 and the arrays were eventually scanned using Gene Array scanner G2500A (Agilent technologies, Palo Alto, CA, USA).

Data preprocessing

[0081] Quality of each array was assessed by the ratio of 3'/5' of two housekeeping genes, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β-actin, and the percentage of genes with "present" calls. All the arrays were normalized against all probesets prior to subsequent analyses. The data generated in chp file format were exported to Affymetrix® Data Mining Tool 3.0 and were filtered with absent/present calls. Only genes that have at least 70% present calls in all samples grouped under a particular clinicopathological factor were kept for further analyses. All samples generated by Affymetrix® Microarray Suite version

5.0 in cel file format were exported to Genedata Expressionist Refiner software package using protocol of GD Affx diagnosis with reference. Refining of all HG-U133A arrays consisted of both tumor and normal tissues were done separately from HG-U133B arrays. The signal intensity of the entire probe sets in both arrays of all samples was normalized to 500.

Statistical analysis

[0082] The correlation between the clinicopathological features of HCC patients with their recurrence status was assessed using the STASTICAL PACKAGE FOR THE SOCIAL SCIENCES (SPSS) for WINDOWS (version 11.0). Univariate analysis was used to investigate the relationship between recurrence and the various clinical parameters including sex, age, presence of HBV or HCV, tumor size, capsulation, blood level of alpha-fetoprotein (AFP), single or multiple lesions, histological grading, presence or absence of cirrhosis and vascular tumor invasion. Recurrence was entered as dependent factor and the other variables studied were entered as covariate factors into a forward stepwise logistic regression model. A p value of less than 0.05 was denoted as statistically significant association between two factors.

[0083] Before the selection of molecular signatures that can predict recurrence, all HCC samples were divided into groups using clinicopathological characteristics that have significant correlation with recurrence. Each group was segregated randomly into training and test sets. Parametric (Student T-test) and non-parametric (Wilcoxon test) statistical tests of the Genedata Expressionist Analyst (version 1.0) were initially employed and followed by median fold change filter (F) to select genes that could discriminate between two subgroups of each predefined clinical parameter. The predefined clinical parameters included patients with and without vascular invasion, patients with and without cirrhosis, and patients with and without recurrence. Genes were ranked using p values calculated from both tests with the top ranking genes having the smallest p values from both tests.

[0084] To minimize the error rate of misclassification during the leave-one-out cross-validation of the training sets, varying number of the top-ranking genes selected above based on p values as well as fold changes progressively increased

from 1.25 to 5 were employed to generate different gene sets for cross-validation. During cross-validation with each of the tested gene set, one HCC sample was omitted at a time from the training set, and the remaining samples were employed to form the prediction model. The cross-validation step was repeated until each of the samples in the training set had been removed once. Gene sets that gave the lowest error rate for the cross-validation test were then employed to estimate the probability whether recurrence will occur for an independent group of HCC samples. The two gene sets that subsequently gave the best prediction were designated as HCC recurrence-predictive molecular signatures (Table 3 and 4).

[0085] Support Vector Machines (SVM), Sparse Linear Discriminant Analyst (SPLDA), and K Nearest Neighbour (KNN) were used for both cross-validation of the training sets and prediction of the independent test sets. Unsupervised hierarchical clustering algorithm was performed with the CLUSTER and TREEVIEW software (M. Eisen, <http://rana.Stanford.EDU/software/>) using mean centered correlation as measurements of similarity and average linkage (9). Genes were annotated using the Affymetrix NetAffx™ GeneOntology analysis system (<http://www.affymetrix.com/site/login/login.affx>) and FATIGO (<http://fatigo.bioinfo.cnio.es/>).

Table 2. Performance of 57-gene set selected by comparing recurrence and non-recurrence of group 2 and 3 on prediction using different classifiers.

Training set	Filter criteria	Probe set (genes)	Accuracy of estimating training set	Test set	Accuracy of predicting test set
11R,	Top	61 (57)	SVM-100%	8R, 7NR	SVM-87% (88%, 86%)
12 NR	130,		SLD-100%		SLD-87% (88%, 86%)
(Gp2 and Gp3)	F>1.5		KNN-96% (100%, 92%)		KNN-87% (88%, 86%)

Abbreviations used: SVM, Support Vector machine; SLD, Sparse Linear Discriminant Analyst; KNN, K Nearest Neighbour; Gp1, patients with invasion and cirrhosis; Gp2, patients with invasion;

but without cirrhosis; Gp3, patients with cirrhosis but without invasion; Gp4, patients negative for both invasion and cirrhosis; R, recurrence; NR, non-recurrence; F, median fold change.

Quantitative real-time PCR

[0086] Real-time PCR was performed using Rotor-Gene 2000 Real Time Cycler (Corbett Research, Mortlake, Sydney, Australia). The forward and reverse primers flanking a region of approximately 200bp were designed with the Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). All primers were designed to have melting temperatures ranging from 58°C to 60°C. A total of 17 genes (11 up-regulated and 6 down-regulated) chosen from the HCC recurrence-predictive molecular signature with the highest median fold changes (ascertained from expression studies with the Affymetrix HG-U133 probe arrays) were tested. A total of 7 normal liver from colorectal patients with liver metastasis (NN), 7 distal normal surrounding tissues of HCC patients (ST), 10 non-recurrent (NR) and 11 recurrent HCC liver samples (R) were studied for each gene.

[0087] Before real-time PCR reaction, each gene was amplified by normal PCR reaction using the melting temperature of 55°C for 30 cycles. Five microgram of DNase1-treated total RNA was reversibly transcribed to first strand cDNA using SuperScript™ II RNase H⁻ Reverse Transcriptase (Invitrogen Life Technologies, Carlsbad, CA, USA) as described in manufacturer's manual. Random hexamer primer and OligodT primer (Invitrogen, Life Technologies, Carlsbad, CA, USA) were used for the reverse transcription (RT) reaction in a total volume of 25 µl and the first strand cDNA was then diluted 3.2 times. The amplicons were then purified by Qiagen PCR purification kit (Qiagen, Hilden, Germany) for later use in constructing the standard curves containing several dilutions.

[0088] The real-time PCR amplification was performed in a total reaction volume of 20 µl containing 2x QuantiTect™ SYBR® Green RT-PCR Master Mix (Qiagen, Hilden, Germany), 1µM mixture of each forward and reverse primers and 1 µl of diluted cDNA. All reactions were carried out with 45 cycles (94°C, 15 sec; 55°C, 30 sec; 72°C, 30 sec) using Rotor-Gene RG 2000.

[0089] Each sample was amplified in duplicates. The 18S ribosomal RNA was employed as a control for normalization to adjust any difference in the amount of

RNA samples added to the reactions. For each gene, the averaged copy concentration of every single sample was normalized against its corresponding averaged concentration of 18S ribosomal RNA to obtain relative expression for comparison among HCC subgroups and normal liver tissues. Statistical analysis was performed with parametric test (unpaired T test) and non-parametric test (Mann-Whitney test) using GraphPad Prism 3.0 software. Median relative signal of each subgroup was used to calculate fold difference between groups.

Table 3. Functional annotation of the 57 genes in the HCC recurrence-predictive molecular signature gene set.

Upregulated genes

Gene function	Gene Name	Gene symbol
Cell cycle	Par-3 partitioning defective 3 homolog (C. elegans)	PARD3
	M-phase phosphoprotein 9	MPHOSPH9
	Cullin 4B	CUL4B
	centrin, EF-hand protein, 2	CETN2
Signaling	Usher syndrome 1C (autosomal recessive, severe)	USH1C
	Protein tyrosine phosphatase, non-receptor type 11	PTPN11
	ADP-ribosylation factor-like 5	ARL 5
	SMAD specific E3 ubiquitin protein ligase 2	SMURF2
Transport	SH3-domain GRB2-like endophilin B2 The	SH3GLB2
	Rac GTPase activating protein 1	RACGAP1
	DnaJ (Hsp40) homolog, subfamily C, member 10	DNAJC10
	Potassium channel, subfamily K, member 1	KCNK1
Development	Glutathione S-transferase M3 (brain)	GSTM3
Translation	Signal sequence receptor, gamma (translocon-associated protein gamma)	SSR3
Metabolism/Biosynthesis	Hypothetical protein FLJ13105	FLJ13105
	Synaptogyrin 2	SYNGR2
	Ribosomal protein L10a	PRL10A
Ubiquitination	Hypothetical protein FLJ23749	FLJ23749
Nucleic acid binding	Acheron	FLJ11196
	R3H domain (binds single-stranded nucleic acids)	R3HDM
	Hypothetical protein FLJ35036	FLJ35036
ATP/ion/protein binding	KIAA0924 protein	KIAA0924
No annotation	Similar to hypothetical protein MGC17347	LOC159090
	Hypothetical protein FLJ11016	FLJ11016
	Hypothetical protein LOC285550	LOC285550
	NMDA receptor-regulated gene 2	NARG2
	Family with sequence similarity 33, member A	FAM33A
	cDNA: FLJ22198 fis, clone HRC01218	
	Transcribed locus, moderately similar to XP_517655.1, similar to KIAA0825 protein [Pan troglodytes]	

Downregulated genes

Gene function	Gene name	Gene symbol
Cell proliferation	Insulin induced gene 1	INSIG1
Cell cycle	S-phase kinase-associated protein 2 (p45)	SKP2
Cell growth & maintenance	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)	ETS2
Immune defense	2'-5'-oligoadenylate synthetase-like	OASL

Signaling	CDC42 small effector 1	CDC42SE1
	Casein kinase 1, gamma 3	CSNK1G3
Transport	Cytochrome P450, family 17, subfamily A, polypeptide 1	CYP17A1
	Cysteine dioxygenase, type I	CDO1
	Cyclic nucleotide gated channel alpha 1	CNGA1
	Nudix-hydrolase type motif 9	NUDT9
	Solute carrier family 16, member 5	SLC16A5
	SAR1a gene homolog 2	SARA2
	Inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	ID2
Proteolysis/Peptidolysis	Ring finger protein 130	RNF130
Transcription	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	MAFB
	Elongation factor, RNA polymerase II, 2	ELL2
	Protein tyrosine phosphatase type IV A, member 1	PTP4A1
Protein processing	Protein geranylgeranyltransferase type I, beta subunit	PGGT1B
	Hexoseaminidase B (beta polypeptide)	HEXB
Metabolism/Biosynthesis	Solute carrier family 27 (fatty acid transporter), member 2	SLC27A2
	1-Acylglycerol-3-phosphate O-acyltransferase 3	AGPAT3
	GTP-Cyclohydrolase 1 (dopa-responsive dystonia)	GCH1
	Sperm acrosome associated 3	SPACA3
	arsenic (+3 oxidation state) methyltransferase	AS3MT
Nucleic acid binding	Tigger transposable element derived 2	TIGD2
ATP/ion/protein binding	KIAA0676 protein	KIAA0676
No annotation	Hypothetical protein DKFZp434H2226	DKFZp434H2226
	MARVEL domain containing 2	MARVELD2

Table 4. Functional annotation of the 65 genes in the HCC recurrence-predictive molecular signature gene set.

Upregulated genes

Gene function	Gene Name	Gene symbol
Cytoskeleton	203087_s_at, Kinesin heavy chain member 2	KIF2
DNA replication	201930_at, Minichromosome maintenance deficient 6	MCM6
	201476_s_at, Ribonucleotide reductase M1 polypeptide	RRM1
DNA repair	210027_s_at, APEX nuclease (multifunctional DNA repair enzyme) 1	APEX1
Immune defense	209772_s_at, CD24 antigen	CD24
	211744_s_at, CD58 antigen	CD58
	209647_s_at, Suppressor of cytokine signaling 5	SOCS5
Signaling	213408_s_at, Phosphatidylinositol 4-kinase, catalytic, alpha polypeptide	PIK4CA/LOC220686
Transport	217959_s_at, Trafficking protein particle complex 4	TRAPPC4
	200750_s_at, RAN, member RAS oncogene family	RAN
	203142_s_at, Adaptor-related protein complex 3, beta 1 subunit	AP3B1
Transcription	203345_s_at, Metal response element binding transcription factor 2	MTF2
	200828_s_at, Zinc finger protein 207	ZNF207
RNA processing	200754_x_at, Splicing factor, arginine/serine-rich 2	SFRS2
Protein processing	200682_s_at, Ubiquitin-conjugating enzyme E2L 3	UBE2L3
Metabolism/Biosynthesis	213890_x_at, Ribosomal protein S16	RPS16
	202144_s_at, Adenylosuccinate lyase	ADSL
Nucleic acid binding	233873_x_at, PAP associated domain containing 1	PAPD1
ATP/ion/protein binding	200757_s_at, Calumenin	CALU
	214845_s_at, Calumenin	CALU
Miscellaneous	206102_at, DNA replication complex GINS protein PSF1	PSF1
	201306_s_at, Acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	ANP32B

Downregulated genes

Gene function	Gene name	Gene symbol
Cell proliferation	211621_at, Androgen receptor	AR
Chemotaxis	207409_at, Leukocyte cell-derived chemotaxin 2	LECT2
Immune defense	207041_at, Mannan-binding lectin serine peptidase 2	MASP2
Signaling	204731_at, Transforming growth factor, beta receptor III	TGFBR3
	227442_at, Mitochondrial COX18	FLJ38991
Transport	232494_at, Cytochrome P450, family 8, subfamily B, polypeptide 1	CYP8B1
	214420_s_at, Cytochrome P450, family 2, subfamily C,	CYP2C9

	polypeptide 9	
	214421_x_at, Cytochrome P450, family 2, subfamily C,	CYP2C9
	polypeptide 9	
	207773_x_at, Cytochrome P450, family 3, subfamily A,	CYP3A43
	polypeptide 43	
	220017_x_at, Cytochrome P450, family 2, subfamily C,	CYP2C9
	polypeptide 9	
	216661_x_at, Cytochrome P450, family 2, subfamily C,	CYP2C19//CYP2C9
	polypeptide 19	
	216025_x_at, Cytochrome P450, family 2, subfamily C,	CYP2C19//CYP2C9
	polypeptide 19	
	228738_at, Hypothetical protein MGC25181	MGC25181
	203880_at, COX17	COX17
	226373_at, Sideroflexin 5	SFXN5
	221014_s_at, RAB33B	RAB33B
Coagulation factor	208034_s_at, Protein Z, vitamin K-dependent plasma glycoprotein	PROZ
Transcription	207202_s_at, Nuclear receptor subfamily 1, group I, member 2	NR1I2
	217707_x_at, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	SMARCA2
Translation	203790_s_at, Heat-responsive protein 12	HRSP12
Metabolism/Biosynthesis	214069_at, Homo sapiens xenobiotic/medium-chain fatty acid:CoA ligase	HXMA
	241914_s_at, Hypothetical protein LOC123876	LOC123876/HXMA
	230554_at, Hypothetical protein LOC123876	LOC123876
	204290_s_at, Aldehyde dehydrogenase 6 family, member A1	ALDH6A1
	221588_x_at	---
	218021_at, Dehydrogenase/reductase (SDR family) member 4	DHRS4
	205633_s_at, Aminolevulinate, delta-, synthase 1	ALAS1
	217973_at, Dicarboxyl/L-xylulose reductase	DCXR
	209531_at, Glutathione transferase zeta 1 (maleylacetoacetate isomerase)	GSTZ1
	202309_at, Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase	MTHFD1
	222400_s_at, Membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1	MTCBP-1
	217761_at, Membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1	MTCBP-1
	223781_x_at, Alcohol dehydrogenase 4 (class II), pi	ADH4

	polypeptide	
Protein processing	216223_at, Carboxypeptidase N precursor	CPN2
Response to stimuli	206065_s_at, Dihydropyrimidinase	DPYS
No annotation	222083_at, Glycine-N-acyltransferase	GLYAT
	206930_at, Glycine-N-acyltransferase	GLYAT
	218544_s_at	RCI1
	230329_s_at, Nudix (nucleoside diphosphate linked moiety X)-type motif 1	NUDT6
	226519_s_at, Hypothetical protein MGC15875	MGC15875
	230577_at	---
	220878_at	---
	238625_at, Chromosome 1 open reading frame 168	C1orf168
	226192_at	---
	226197_at	---
	226682_at, Hypothetical protein LOC283666	LOC283666
	225534_at, Chromosome 8 open reading frame 40	C8orf40
	239093_at, Chromosome 10 open reading frame 65	C10orf65

What is claimed is:

1. A method of clinically predicting recurrent hepatocellular carcinoma in a subject, comprising determining at least two clinicopathological factors.
2. The method of claim 1, wherein the at least two clinicopathological factors are cirrhosis and vascular invasion.
3. A method of clinically predicting recurrent hepatocellular carcinoma in a subject, comprising: (a) obtaining a sample from the subject; and (b) determining the expression level of two or more nucleic acid molecules in the sample.
4. The method of claim 3, wherein the two or more nucleic acid molecules the gene expression profile of which is determined are selected from the group of genes consisting of PARD3, MPHOSPH9, CUL4B, CETN2, USH1C, PTPN11, ARL5, SMURF2, SH3GLB2, RACGAP1, DNAJC10, KCNK1, GSTM3, SSR3, FLJ13105, SYNGR2, PRL10A, FLJ23749, FLJ11196, R3HDM, FLJ35036, KIAA0924, LOC159090, FLJ11016, LOC285550, NARG2, FAM33A, FLJ22198 (clone HRC01218) (SEQ ID NO: 1), transcribed locus moderately similar to XP_517655.1 (similar to KIAA0825) (SEQ ID NO: 2), INSIG1, SKP2, ETS2, OASL, CDC42SE1, CSNK1G3, CYP17A1, CDO1, CNGA1, NUDT9, SLC16A5, SARA2, ID2, RNF130, MAFB, ELL2, PTP4A1, PGGT1B, HEXB, SLC27A2, AGPAT3, GCH1, SPACA3, AS3MT, TIGD2, KIAA0676, DKFZp434H2226, and MARVELD2, or expression products, complements, fragments, variants, or analogs thereof.
5. The method of claim 4, wherein the gene expression profile of 5, 7, 10, 12, 16, 20, 30, 40, 50, 55, 56 or 57 of said genes is determined.
6. The method of claim 5, wherein the gene expression profile of all 57 genes is determined.

7. The method of claim 4, wherein the gene expression profile of the each of the genes RACGAP1, KCNK1, SMURF2, USH1C, GSTM3, CNGA1, and INSIG1 is determined.
8. The method of claim 3, wherein the gene expression profile of two or more nucleic acid molecules in the sample is determined, and wherein the nucleic acid molecules are selected from the genes set forth in Table 4.
9. The method of claim 8, wherein the gene expression profile of two or more of the following genes is determined: KIF2, MCM6, RRM1, APEX1, CD24, CD58, SOCS5, PIK4CA, TRAPPC4, RAN, AP3B1, MTF2, ZNF207, SFRS2, UBE2L3, RPS16, ADSL, PAPD1, CALU, PSF1, ANP32B, AR, LECT2, MASP2, TGFBR3, FLJ38991, CYP8B1, CYP2C9, CYP3A43, CYP2C19, MGC25181, COX17, SFXN5, RAB33B, PROZ, NR1H2, SMARCA2, HRSP12, HXMA, LOC123876, ALDH6A1, DHRS4, ALAS1, DCXR, GSTZ1, MTHFD1, MTCBP-1, ADH4, CPN2, DPYS, GLYAT, RCI1, NUDT6, MGC15875, C1orf168, LOC283666, C8orf40, and C10orf65.
10. The method of claim 8 or 9, wherein the gene expression profile of 5, 10, 15, 20, 30, 40, 50, 60, 64, or 65 of said genes is determined.
11. The method of any of claims 3-10, wherein the determined expression level is compared to a control sample.
12. The method of any of claims 3-11, further comprising the determination of at least two clinicopathological factors in said subject.
13. The method of claim 12, wherein the at least two clinicopathological factors are cirrhosis and vascular invasion.
14. The method of any of the preceding claims, wherein the subject is human.

15. The method of any of the preceding claims, wherein the sample comprises liver tissue.

16. The method of claim 15, wherein the sample comprises liver cancer tissue.

17. The method of claim 16, wherein the sample comprises a hepatocellular carcinoma cell.

18. The method of any of claims 3-10, wherein said determining step is carried out using a DNA array, a quantitative PCR assay, or real time PCR.

19. A kit for the detection of the expression level of two or more target nucleic acid molecules in a sample, said target nucleic acid molecules being selected from the group consisting of PARD3, MPHOSPH9, CUL4B, CETN2, USH1C, PTPN11, ARL5, SMURF2, SH3GLB2, RACGAP1, DNAJC10, KCNK1, GSTM3, SSR3, FLJ13105, SYNGR2, PRL10A, FLJ23749, FLJ11196, R3HDM, FLJ35036, KIAA0924, LOC159090, FLJ11016, LOC285550, NARG2, FAM33A, FLJ22198 (clone HRC01218), transcribed locus moderately similar to XP_517655.1 (similar to KIAA0825), INSIG1, SKP2, ETS2, OASL, CDC42SE1, CSNK1G3, CYP17A1, CDO1, CNGA1, NUDT9, SLC16A5, SARA2, ID2, RNF130, MAFB, ELL2, PTP4A1, PGGT1B, HEXB, SLC27A2, AGPAT3, GCH1, SPACA3, AS3MT, TIGD2, KIAA0676, DKFZp434H2226, and MARVELD2 and complements, variants, fragments, and analogs thereof, comprising one or more oligonucleotides complementary to a target nucleic acid molecule.

20. A kit for the detection of the expression level of two or more target nucleic acid molecules in a sample, said target nucleic acid molecules being selected from the group consisting of: KIF2, MCM6, RRM1, APEX1, CD24, CD58, SOCS5, PIK4CA, TRAPPC4, RAN, AP3B1, MTF2, ZNF207, SFRS2, UBE2L3, RPS16, ADSL, PAPD1, CALU, PSF1, ANP32B, AR, LECT2, MASP2, TGFBR3, FLJ38991, CYP8B1, CYP2C9, CYP3A43, CYP2C19, MGC25181, COX17, SFXN5, RAB33B, PROZ, NR1I2, SMARCA2, HRSP12, HXMA, LOC123876, ALDH6A1, DHRS4,

ALAS1, DCXR, GSTZ1, MTHFD1, MTCBP-1, ADH4, CPN2, DPYS, GLYAT, RCI1, NUDT6, MGC15875, C1orf168, LOC283666, C8orf40, and C10orf65.

21. The kit of claim 19 or claim 20, wherein the oligonucleotides are oligonucleotide probes.

22. The kit of claim 21, wherein said probes are labelled.

23. The kit of claim 22, wherein the label is a radioactive, fluorescent, chemoluminescent, affinity, or enzymatic label.

24. The kit of any of claims 20-23, wherein one or more of the oligonucleotide probes are immobilized on a substrate.

25. The kit of claim 24, wherein the substrate is a DNA microchip.

26. The kit of claim 19 or claim 20, wherein the oligonucleotides are amplification primers.

27. The kit of claim 25, wherein said amplification primers are suitable to amplify the target nucleic acid in an amplification step.

28. The kit of any of claims 19-27, wherein the oligonucleotides are up to about 30, about 60, or about 100 nucleotides in length.

29. A polynucleotide comprising the nucleotide sequence of SEQ ID NO: 2.

Figure 1

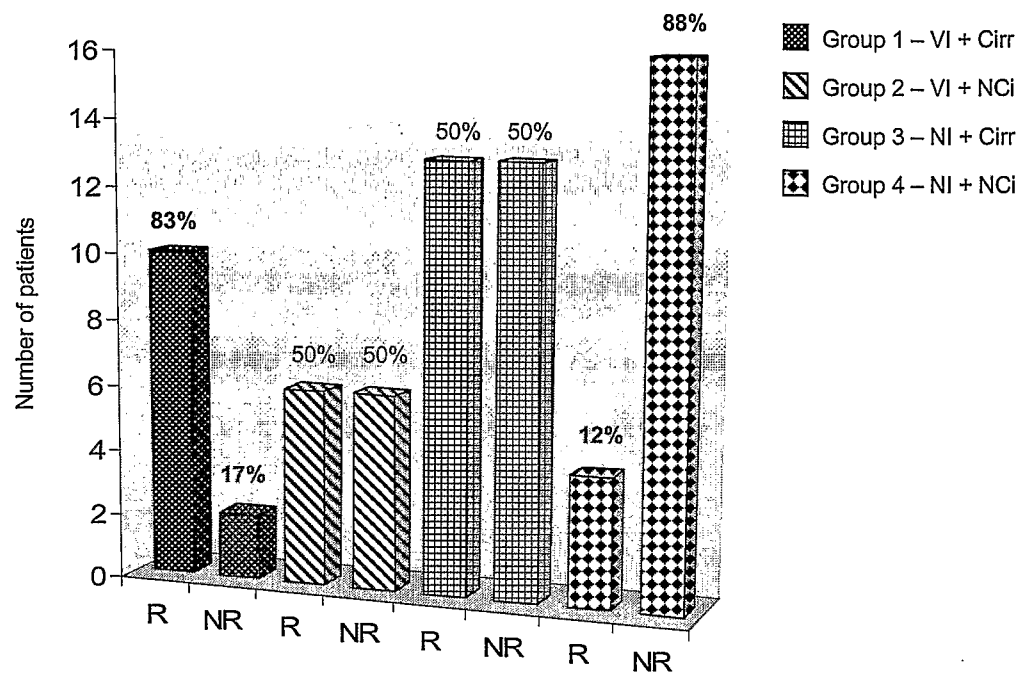


Figure 2

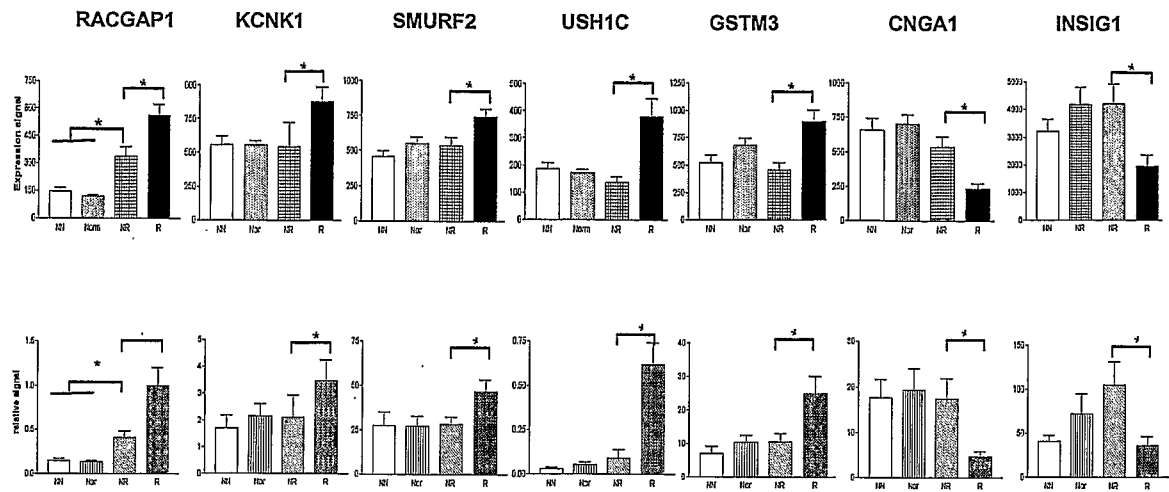
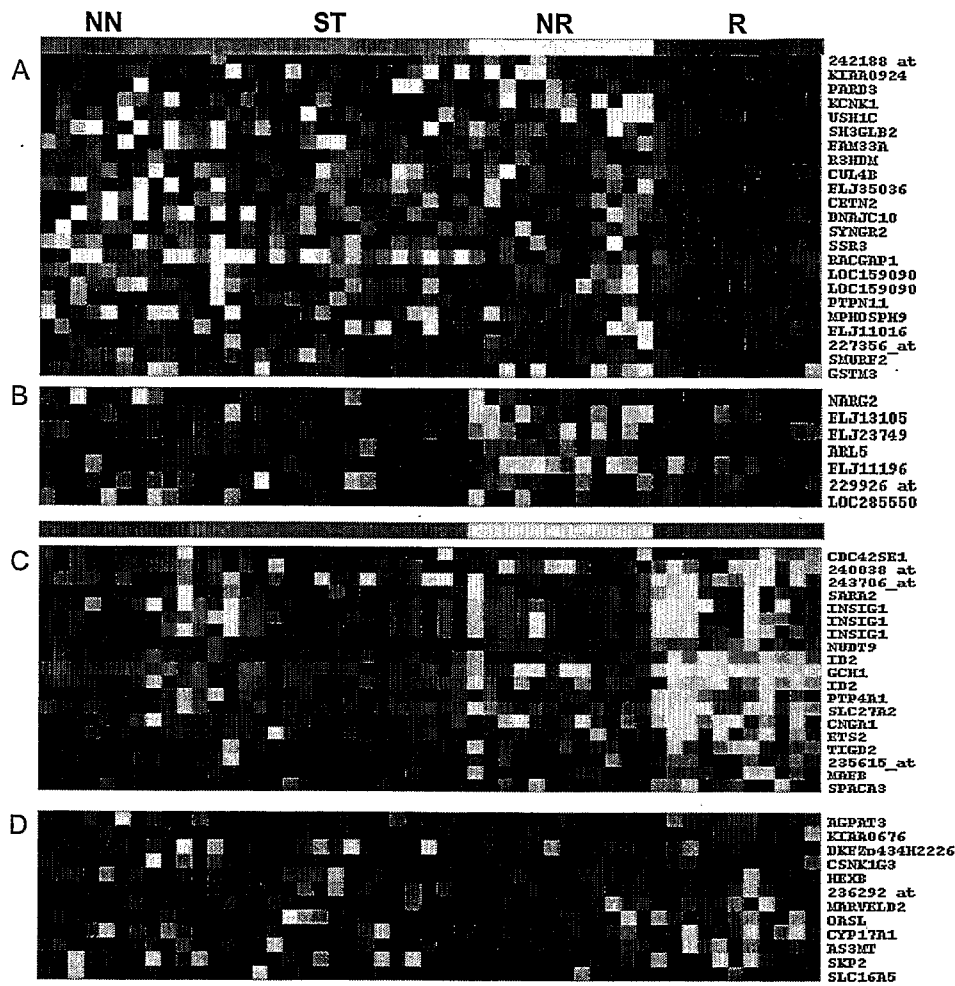


Figure 3



Sequence listing

<110> National Cancer Center

<120> Methods for the prediction of hepatocellular carcinoma recurrence
by the determination of hepatocellular carcinoma
recurrence-associated molecular biomarkers

<130> P101237

<160> 2

<170> PatentIn version 3.3

<210> 1

<211> 512

<212> DNA

<213> Homo sapiens

<400> 1

```
aaaactgtaa acagctccca tatccatcaa gtgtatggat aaacaaattt ggtgtattta      60
tacaatggaa tactactcgg caataaaaag aacagttgat actctcaaca acctagatgg      120
acctcaaaat aattcgggtt aatgaatgaa gccaaactta agaagagtac attgtatgta      180
cttggaagaac taacttcttg caatagattt ttaagcacta ttaggagcat atgacttaaa      240
cagtttttaa aagtcaggga gtaagtatgc ttaaataaaa tacaatctgt gaaacaaatc      300
tctgaattat tatcacttca ctggacactc taacttgacc atatttctga ctttaatgta      360
actcaactctt attccgtagt cacatgtttg cttgctcatt gggtcacatt acatttattc      420
agcatctgct tgagccaagg cactgtaact acatgttttt tttagttacc tacttttgta      480
aggtcctggt tctttggcta catctgatta ca                                     512
```

<210> 2

<211> 332

<212> DNA

<213> Homo sapiens

<400> 2

gaaatacata gtgggctcga ttcatagccccc cgcgccgtat ctatatctct atactctatag	60
agacataact tctttcttcg taagttatat atgtctactt ctacctgtgt tatcataata	120
aagggtgcat gatgatacag gtggaggtag aaatatataa cttatctctt cataagggct	180
aaaatgagaa ctaaaagtaa tgagaattca gttgaatatg attaatatc aggatatggt	240
aattagttct cttgatgctt actgtgatgc aaagtatgtg acgtatttta aacattttat	300
aataattggt cgtatgttaa ccattaggca ag	332

SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2006/000340

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

C12Q 1/68 (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Databases: MEDLINE, CAPLUS, WPIDS, Biotechabs

Keywords: Genes as listed in Table 3 of the application + synonyms; hepatocellular carcinoma/liver cancer/liver carcinoma/hepatocarcinoma

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CA 2536324 A (NIHON UNIVERSITY et al) 10 March 2005 Tables 1-8; Claims	3, 4, 11-19, 21-28
X, Y	WO 2005/024020 A (OLYMPUS CORPORATION) 17 March 2005 Claims 11-21	19, 21-28

☒ Further documents are listed in the continuation of Box C☒ See patent family annex

* "A"	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E"	earlier application or patent but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

22 February 2007

Date of mailing of the international search report

- 2 MAR 2007

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaustalia.gov.au
Facsimile No. (02) 6285 3929

Authorized officer

Sophina Calanni

Telephone No : (02) 6283 2038

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2006/000340

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Iizuka, N. et al., 2003, Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection, <i>Lancet</i> , 361 (9361):923-929. Whole document	19, 21-28
Y	Kurokawa Yet al., 2004, Molecular-based prediction of early recurrence in hepatocellular carcinoma, <i>Journal of Hepatology</i> , 41 (2):284-91. Whole document	19, 21-28
Y	He, H. et al., 2002, Identification and characterization of nine novel human small GTPases showing variable expressions in liver cancer tissues, <i>Gene Expression</i> , 10 : 231-242. Table 3	19, 21-28
Y	Ito, Y. et al., 2002, Linkage of Elevated Ets-2 Expression to Hepatocarcinogenesis, <i>Anticancer Research</i> , 22 : 2385-2390. Table 1	19, 21-28
Y	Stephens, B. J. et al., 2005(Nov), PRL phosphatases as potential molecular targets in cancer, <i>Molecular Cancer Therapy</i> , 4 (11): 1653-1661. Table 2	19, 21-28
Y	Severini, G. et al., 1995, A study of serum glycosidases in cancer, <i>Journal of Cancer Research and Clinical Oncology</i> , 121 : 61-63. Table 1	19, 21-28
Y	Xu, L. et al., 2001, Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma, <i>Cancer Research</i> , 61 : 3176-3181. Whole document	19, 21-28
Y	Li, Yao, et al., 2002, Discovery and analysis of hepatocellular carcinoma genes using cDNA microarrays, <i>Journal of Cancer Research and Clinical Oncology</i> , 128 : 369-379. Whole document	19, 21-28

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2006/000340

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: **3, 4, 19 and the claims appended thereto (partially)**
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

See Supplemental Box I

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See Supplemental box

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid; specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention covered by claims Nos. **4-7, 19 (in full); claims 3, 11-18, 21-28 (partially).**

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Supplemental Box

(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Box No: II

Claim 3 is not limited to the technical features of the invention described in the specification. The applicant's invention resides in the determination of a HCC recurrence associated molecular signature comprising 57 genes (as listed in Table 3 and claim 4) differentially expressed in recurrent HCC. As such, claim 3 has been searched to the extent that the claim relates to the specific sequences described in claim 4/Table 3.

Claim 4 and 19 relate to a method and kit for predicting recurrent hepatocellular carcinoma in a subject. The claims list 57 target nucleic acids for which the expression level can be determined, and also encompass methods and kits comprising fragments, analogues and variants of the genes of the genes listed. The full scope of the claims has not been searched because the scope of the terms fragments, analogues and variants is indeterminate and thus unsearchable. The claims have only been searched in so far as they relate to the specific genes listed, and complements thereof.

Continuation of Box No: III

PCT Rule 13.2, first sentence, states that unity of invention is only fulfilled when there is a technical relationship among the claimed inventions involving one or more of the same or corresponding special technical features. PCT Rule 13.2, second sentence, defines a special technical feature as a feature which makes a contribution over the prior art.

This International Application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

In assessing whether there is more than one invention claimed, I have given consideration to those features which can be considered to potentially distinguish the claimed combination of features from the prior art. Where different claims have different distinguishing features they define different inventions.

This International Searching Authority has found that there are different inventions as follows:

- Invention 1. Claims 1-2. It is considered that a method for clinically predicting recurrent hepatocellular carcinoma by determining at least two clinicopathological factors represents a first special technical feature.
- Inventions 2-57. Claims 3-7, 11-19, 21-28 (all partially). It is considered that each of the genes shown to be up- or down-regulated in the HCC recurrence predictive molecular signature gene set (See Table 3), and their use to predict the recurrence of hepatocellular carcinoma, represents a distinct special technical feature.
- Inventions 58-123. Claims 8-18, 20-28 (all partially). It is considered that each of the 65 genes listed in Table 4, and their use to predict the recurrence of hepatocellular carcinoma, represents a distinct special technical feature.
- Invention 124. Claims 3-7, 11-19, 21-28 (all partially) and claim 29 (completely). It is considered that the polynucleotide represented by SEQ ID NO: 2, and its use to predict the recurrence of hepatocellular carcinoma represents a 124th special technical feature.

Continued in Supplemental Box II

Supplemental Box II

(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Supplemental Box I

With respect to inventions 2-123, the claims are directed to methods and kits for predicting recurrent hepatocellular carcinoma (HCC) based on determining the expression level of two or more nucleic acid molecules that have been shown to be either up-or down-regulated in the HCC recurrence predictive molecular signature set. Although all of the genes identified share the feature that alterations in their expression may be indicative of a higher risk of recurrence of HCC, this does not represent a special technical feature.

This feature cannot be a special technical feature because numerous studies have determined the prognostic role of various biomarkers in predicting recurrent HCC (see documents listed below). For example, (1) discloses a correlation between low expression of connexin 26 and 32 mRNA and recurrence of HCC, while (2)-(4) describe increased expression of heparanase, granulin-epithelin and MMP-9 in association with recurrence of HCC.

(1) Sheen, I-S. et al., 2004, Are gap junction gene connexins 26, 32 and 43 of prognostic values in hepatocellular carcinoma? A prospective study, *World Journal of Gastroenterology*, 10(19): 2785-2790.

(2) Chen, X-P. et al., 2004, Heparanase mRNA expression and point mutation in hepatocellular carcinoma, *World Journal of Gastroenterology*, 10(19): 2795-2799.

(3) Cheung, S. T. et al., 2004, Granulin-Epithelin precursor overexpression promotes growth and invasion of hepatocellular carcinoma, *Clinical Cancer Research*, 10: 7629-7636.

(4) Jiang, YF et al., 2000, Recurrence or metastasis of HCC: predictors, early detection and experimental antiangiogenic therapy.

In addition, it is noted that the claims 3-28 relate to methods and/or kits for determining the expression of two or more nucleic acids in the sample. While the citations listed above do not necessarily describe methods of predicting HCC recurrence by detecting the expression level of two or more nucleic acid molecules in the sample, it is clear that alterations in the expression of specific nucleic acid molecules has previously been associated with an increased risk for HCC recurrence, and it would not be inventive to provide methods for predicting recurrent HCC based on determining the expression level of two or more such molecules. As such, methods of predicting recurrent HCC comprising determining the expression level of two or more nucleic acid molecules do not represent a special technical feature.

Thus, the common feature of the claims cannot constitute a special technical feature within the meaning of PCT Rule 13.2, second sentence, since it makes no contribution over the prior art. Because the common feature does not satisfy the requirement for being a special technical feature it follows that it cannot provide the necessary technical relationship between the identified inventions. Therefore the claims do not satisfy the requirement of unity of invention *a posteriori*.

Following consultation with the applicants, this International Searching Authority has agreed to search the method of claim 4 for all 57 genes specified in that claim.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2006/000340

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report				Patent Family Member			
CA	2536324	CN	1507106	EP	1429414	EP	1661991
		JP	2004208273	KR	2004005151	US	2004113710
		WO	2005021745				
WO	2005024020						
Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.							
END OF ANNEX							