RESEARCH ARTICLE

Identification of CHI3L1 and MASP2 as a biomarker pair for liver cancer through integrative secretome and transcriptome analysis

Jun Wang[1], Feng Gao[2]*, Fan Mo[1], Xu Hong[1], Hongyang Wang[3]**, Shusen Zheng[2]** and Biaoyang Lin[1]*

[1] Systems Biology Division, Zhejiang-California International Nanosystems Institute (ZCNI) of Zhejiang University, Hangzhou, China
[2] Department of Hepatobiliary Pancreatic Surgery, Key Lab of combined Multi-Organ Transplantation of Ministry of Public Health, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China
[3] International Co-operation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, Shanghai, China

Hepatocellular carcinoma (HCC) is the fifth most frequent neoplasm with more than 500 000 new cases diagnosed yearly. Novel liver cancer biomarkers are needed. By tandem mass spectrometry, we analyzed the secretomes of 12 individual paired samples of liver cancer and adjacent normal tissues and identified 1528 proteins with >2 unique peptide hits. The false discovery rate was 3.4%. Using spectral counting, we found 87 proteins in the HCC group and 86 proteins in the normal group that showed fivefold overexpression. These proteins provided a rich source of biomarker candidates. We presented a novel paradigm in combining biomarkers that include an up-regulated cancer biomarker and a down-regulated organ-enriched marker, and identified chitinase-3-like protein 1(CHI3L1) and mannan-binding lectin serine peptidase 2 (MASP2) as the top biomarker pair for HCC diagnosis using integrative transcriptomics and proteomics analysis. Using ELISA assays, we further evaluated this biomarker pair in a separate cohort of 25 serum samples of liver cancer patients and 15 age-matched normal controls. The combined marker pair (YKL40/MASP2 ratio) performed better than either marker alone with an AUC of 0.97 for liver cancer diagnosis. Further validation of the biomarker pair in HCC patients *versus* disease controls and independent cohorts is warranted.

Received: June 27, 2008
Revised: September 29, 2008
Accepted: October 17, 2008

Keywords:
Biomarker / Chitinase-3-like protein 1 / Hepatocellular carcinoma / Secretome

1 Introduction

Primary hepatocellular carcinoma (HCC) is the fifth most frequent neoplasm and the third most common cause of

Correspondence: Dr. Biaoyang Lin, Systems Biology, Zhejiang University Zhejiang-California Nanosystems Institute (ZCNI), 268 Kaixuan Road, Hangzhou 310029, China
E-mail: biaoylin@zju.edu.cn
Fax: +86-571-869-71851

Abbreviations: CHI3L1, chitinase-3-like protein 1; **GO**, Gene Ontology; **HCC**, hepatocellular carcinoma; **MASP2**, mannan-binding lectin serine peptidase 2; **MW**, molecular weight; **ROC**, receiver operating characteristic; **tpm**, transcripts *per* million

* These authors contributed equally to this work.
** Additional corresponding authors: Dr. Shusen Zheng, E-mail: shusenzheng@zju.edu.cn; Dr. Hongyang Wang, E-mail: hywangk@vip.sina.com

cancer-related death, with more than 500 000 new cases diagnosed yearly [1]. HCC is curable by surgery if it is identified early enough and patients with liver cirrhosis were screened with biomarker alpha-fetoprotein (AFP) and by ultrasound every 6 months to detect HCC at earlier stages [2]. Unfortunately, although the AFP is widely used for diagnosis and monitoring of HCC [3, 4], its false negative or false positive rate is as high as 40% [5, 6].

Analyzing serum samples directly for disease diagnosis and prognosis offers several key advantages, including low invasiveness, minimum cost, easy sample collection and processing. However, due to the complexity and an extraordinary huge dynamic range of at least 10^9–10^{10} [7] of the serum proteome, direct proteomics analysis is inherently challenging. We adopted an approach to (i) analyze secretome (secreted proteins) from HCC and uninvolved surrounding tissue culture media, (ii) analyze the transcriptome of HCC and uninvolved surrounding tissues to identify differentially expressed genes, (iii) identify top candidate serum biomarkers by integrative transcriptome and secretome analysis, (iv) validate candidate biomarkers using ELISA assays, and (v) combine cancer biomarker with organ-enriched expression biomarker to increase accuracy in diagnosis. Using the above approach, we identified a biomarker pair CHI3L1 and MASP2 that has the AUC of 0.97 for liver cancer diagnosis.

2 Materials and methods

2.1 Clinical sample collection

HCC and benign adjacent paired tissues (at least 2 cm away from the edge of HCC tissues) were collected from 12 HCC patients who underwent hepatectomy or liver transplantation at the First Affiliated Hospital, Zhejiang University with IRB approval. Clinical and pathologic data of the 12 cases are summarized in Table 1. None of these patients received antineoplastic therapy prior to surgery. Additional serum samples were obtained from the serum bank at the First Affiliated Hospital, Zhejiang University.

2.2 Tissue culture

The paired tissues were transferred to a Petri dish containing 20 mL of PBS and were finely minced into 2–3-cubic millimeter pieces using scissors. Thereafter the tissue pieces were re-suspended in 50 mL PBS and were poured over the stainless steel filter (200-μm diameter) to discard single cells and cell debris. The collected tissue pieces were washed three times with PBS and were re-suspended in 20 mL serum-free DMEM (Sigma, St Louis, MO) in a Petri dish. The tissue pieces were cultured at 37°C in a cell culture incubator (Thermo Scientific, Milford, MA) with 5% CO_2.

2.3 Proteomics sample preparation

Supernatants from tissue culture were collected at 24 h after tissue culture. The supernatants were centrifuged at $2000 \times g$ for 10 min to remove any cells or cell debris that might be contained in the supernatants. The samples were concentrated about 20-fold by a Speedvac (Labconco Centrivap Concentrator, Kansas City, MO) and were re-suspended in 25 mM ammonium bicarbonate (NH_4HCO_3, Sigma, St Louis, MO). Of each sample, 60 μg was separated on 12% SDS polyacrylamide gels. Gels were stained with Colloidal CBB. Proteins in the gel were digested with trypsin using the Pierce In-Gel Tryptic Digestion Kit protocol (Pierce Biotechnology, Rockford, IL).

2.4 MS analysis

Tryptic peptide mixture was separated by the Ettan MDLC nanoflow/capillary LC system (GE Healthcare, Pittsburgh, PA) equipped with a trapping column (Dionex/LC Packings μ-Precolumn Cartridge P/N 160454 C18 PepMap 100, 5 μm, 100 Å, 300-μm id x 5 mm, Sunnyvale, CA) and a nanocolumn (Dionex/LC Packings P/N 160321 150×0.075-mm id, C18 PepMap, 3 μm, 100 Å), and then analyzed using LTQ-Orbitrap (Thermo Finnigan, Bremen, Germany) with a nanospray configuration. The precursor ion scan MS spectra (m/z 300–1600) were acquired in the orbitrap with the resolution R = 60 000 at m/z 400 with the number of accumulated ions being 1×10^6. The five most intense ions were isolated and fragmented in linear IT (number of accumulated ions: 3×10^4). The resulting fragment ions were recorded with the resolution R = 15 000 at m/z 400.

2.5 MS data analysis

The extract_msn of the BioWorks program V3.2 (Thermo Electron, Waltham, MA) was used to generate the MS peak list with the default parameters. The ICIS peak-detection algorithm peaks of the Xcalibur (Thermo Electron) was used for automated detection of mass spectrum. The SEQUEST algorithm (Thermo Fisher) was used for the SEQUEST database search, the spectra were searched against the ipi.HUMAN.v3.29.fasta protein database (with 70 757 entries) (http://www.ebi.ac.uk/IPI/IPIhu man.html) using the BioWorks program V3.2 (Thermo Electron). In the TurboSEQUEST search parameter setting, the threshold for Dta generation was 10 000, and precursor mass tolerance for Dta generation was 1.4. For the SEQUEST search, peptide tolerance was set at 3 Da and fragment ions tolerance was set at 0.01 Da. PeptideProphet™ [8] was used to assess the MS/MS spectra quality and a threshold score for accepting individual MS/MS spectra was set at p value of 0.9, which corresponds to a 0.5% error rate in our dataset. One missed tryptic cleavage was permitted. Carboxyamidomethyl cysteine (Cys_CAM) (+ 57) was included as a fixed modification for iodoacetamide reduction and alkylation. As the proteins were prepared by PAGE, the cysteines might react with free

Table 1. Pathologic data from 12 HCC patients used in LC/MS/MS analysis

Patient no.	Gender	Age	Size (cm)	TNM grade	Edmondson grade	Pathologic data	HBV	HCV	Fibrosis stage	Child-pugh grade
P1	M	45	5.1 × 3.5	II	III	Hepatocellular carcinoma (right lobe), grade III, mixed nodular type hepatocirrhosis	+		4	A
P2	M	68	3 × 3	II	III	Hepatocellular carcinoma (right lobe), grade III	+		4	A
P3	F	71	4.9 × 4.0	II	III	Hepatocellular carcinoma (right lobe), grade III, mixed nodular type hepatocirrhosis	+		4	A
P4	M	82	3 × 2	II	II	Hepatocellular carcinoma (left lobe), grade	-		4	A
P5	M	64	5.9 × 7.7	IVa	IV	Carcinoma sarcomatodes (right lobe), grade a, small nodular type hepatocirrhosis	+		4	A
P6	M	50	7.1 × 7.8	IVa	III	Hepatocellular carcinoma (left lobe), grade III, mixed nodular type hepatocirrhosis	+		4	A
P7	M	57	4.5 × 3.7	IVa	III	Hepatocellular carcinoma (right lobe), grade III, mixed nodular type hepatocirrhosis	+		4	B
P8	M	65	4.2 × 4	II	III	Mixed hepatocellular and cholangio-cellular carcinoma (right lobe) grade III, small nodular type hepatocirrhosis		+	4	B
P9	M	45	2.4 × 2.5	II	I	Hepatocellular carcinoma (right lobe), grade I, mixed nodular type hepatocirrhosis	+		4	A
P10	F	49	4 × 3	IIIa	III	hepatocellular carcinoma (right lobe), grade III, mixed nodular type hepatocirrhosis	+		4	A
P11	M	70	2.7 × 2.3	II	III	Hepatocellular carcinoma (right lobe), grade III, mixed nodular type hepatocirrhosis	+		4	A
P12	M	39	3.3 × 3.2	IIIa	IV	Carcinoma sarcomatodes (right lobe), grade, mixed nodular type hepatocirrhosis, mixed nodular type hepatocirrhosis	+		4	A

acrylamide monomers to form propionamide cysteine (Cys_PAM). We included an optional 14 Da in the search to account for potential propionamide cysteine (the mass difference between Cys-PAM and Cys-CAM is 14). Methionine oxidation (+16 Da) was chosen as another optional modification for the database search. Proteins with ProteinProphet *p* value greater than 0.9 and with more than two unique peptide hits were considered as true hits. A randomized database of the ipi.HUMAN.v3.29.fasta was used as a decoy database to calculate the false discovery rate of protein identification. The perl script used for randomization was from www.matrixscience.com/downloads/decoy.pl.gz. The false discovery rate (FDR) was calculated by the ratio of the number of matches to the randomized database to that to the ipi.HUMAN.v3.29.fasta database.

2.6 Spectral counting and Gene Ontology analyses

We summed up the total spectrum numbers in the HCC group (12 samples) and the control group (12 samples). The spectrum numbers were normalized to the total number of spectra of all proteins identified. Ratios of spectrum numbers between the HCC and the control group were calculated. GoMiner [9] was used to find statistically represented Gene Ontology (GO) categories. The 1528 proteins with more than two hits were used as the total input and the differentially expressed genes were analyzed using evidence level 3.

2.7 Western blot analysis

Proteins from the HCC and the uninvolved surrounding tissues were separated on 12% polyacrylamide gels and trans-

ferred to PVDF membranes (Amersham Pharmacia. Biotech, Uppsala, Sweden). These blots were incubated for 2 h at room temperature in the TBST buffer (20 mM Tris-Cl, 140 mM NaCl, pH 7.5, 0.05% Tween-20) containing 5% skim milk, then incubated with the primary antibody anti-AAT (Alpha-1-antitrypsin, IPI00553177) (Santa Cruz Biotechnology, CA) overnight at 4°C. After washing three times in TBST, blots were incubated with HRP-conjugated secondary antibody (diluted 1:10 000, Santa Cruz Biotechnology) for 1 h at room temperature. ECL reagents were used for visualization (Pierce Biotechnology).

2.8 ELISA assay

The ELISA kit for CHI3L1 (YKL40) and MASP2 were purchased from Quidel Corp. (San Diego, CA) and Hycult biotechnology bv (Uden, The Netherlands) and ELISA was performed according to the manufacturer's instruction. Serum samples were diluted three times with PBS buffer before analysis. ROC (receiver operating characteristic) curve analysis was performed using GB STAT V10.0 (Dynamic Microsystems, Silver Spring, MD).

3 Results

3.1 Identification of differentially expressed proteins between the secretomes of HCC and uninvolved surrounding tissues

We compared the proteome of the culture media (secretome) of 12 paired HCC and uninvolved surrounding tissues in serum-free media. We identified 1107 and 977 proteins with ProteinProphet *p* value greater than 0.9 and with more than two unique peptide hits in HCC and normal secretomes, respectively (Supporting Information Tables 1 and 2). Additional 70 proteins have one unique peptide in the HCC or the normal secretome, and the unique peptide in the HCC is different from that in the normal secretome. When the data from the HCC and the normal secretome were combined, these proteins had two unique peptide hits. Therefore, the final total number of peptides with >2 unique peptide hits is 1528. Using a randomized database of the ipi.HUMAN.v3.29.fasta as a decoy database, we calculated that these 1528 proteins have an FDR of 3.4%.

We applied spectral counting method [10] for semi-quantitative comparative analysis of the secretomes between the HCC and the normal group. We compared the sum of the spectral counts of 12 samples of HCC to that of 12 normal controls. We identified 87 proteins as over-expressed ($\geq$5-fold) in the HCC group (Supporting Information Table 3) and 86 proteins as overexpressed ($\geq$5-fold) in the normal group (Supporting Information Table 4). By spectral counting, AAT (Alpha-1-antitrypsin precursor, IPI00553177) showed 3.8-fold (1249 in the HCC group *vs.* 325 in the normal group) overexpression in the HCC secretomes as compared to the normal secretomes. Western blot analysis confirmed that AAT is overexpressed in the HCC group compared to the normal group (Supporting Information Fig. 1).

Enrichment analysis in GO categories of the differentially expressed genes was performed by GoMiner [9]. By GO Cellular Component categories, there was enrichment in the extracellular region (GO:0005576) term for the proteins identified as over expressed in the HCC secretomes. Of the 73 proteins that can be mapped to the extracellular region GO term, 10 belong to the HCC overexpressed group, but none belongs to the control over expressed group, suggesting that the HCC might have increased secretion activities. By GO biological process terms, we found that the GO terms enriched in the HCC over expressed proteins include heparin-binding (GO:0008201), calmodulin-binding (GO:0005516), glycosaminoglycan-binding (GO:0005539), and I-kappaB kinase NF-kappaB cascade (GO:0007249) (Fig. 1). Interesting categories enriched in the HCC underexpressed proteins include androgen metabolic process (GO:0008209) (Fig. 1). AK1C4 and AK1D1 (Aldo-keto reductase family 1 member C4, and member D1; also named as 3 alpha-hydroxysteroid dehydrogenase and steroid 5-beta-reductase, respectively) are two proteins in the GO term 0008209 and they are underexpressed in the HCC group. These two proteins are involved in androgen and estrogen metabolism (hsa00150 of the KEGG pathway). AK1D1 (E.C. 1.3.99.6) is involved in converting testosterone to more potent 5-beta-dihydrotestosterone (has00150 of the KEGG pathway).

3.2 Integrating proteomics data with transcriptomics data

Spectral counting method is only semi-quantitative. We tried to integrate our semi-quantitative proteomics data with the public data and our own transcriptomics data to help us identify and prioritize candidate genes to be validated. Using the Illumina's next-generation sequencing technology, we generated digital transcriptomics data for an HCC and its uninvolved surrounding tissue. In digital expression profiling, the abundance of transcripts is represented in transcripts *per* million (tpm) [11]. We integrated the transcriptomics data with the proteomics data. To prioritize our list, we set a filter so that the ratio of spectral counting for proteomics data would be >2 and the ratio in transcriptomics data would be >3 and absolute difference between cancer and normal in tpm would be >30 (to increase confidence in comparing lowly expressed genes with low tpm values). Table 2 lists 103 top candidate genes derived from this integrative analysis. When selecting the top candidates for further validation, we also compared our data with the liver cancer microarray data [12–14] in the Oncomine database (www.oncomine.org) (data not shown).

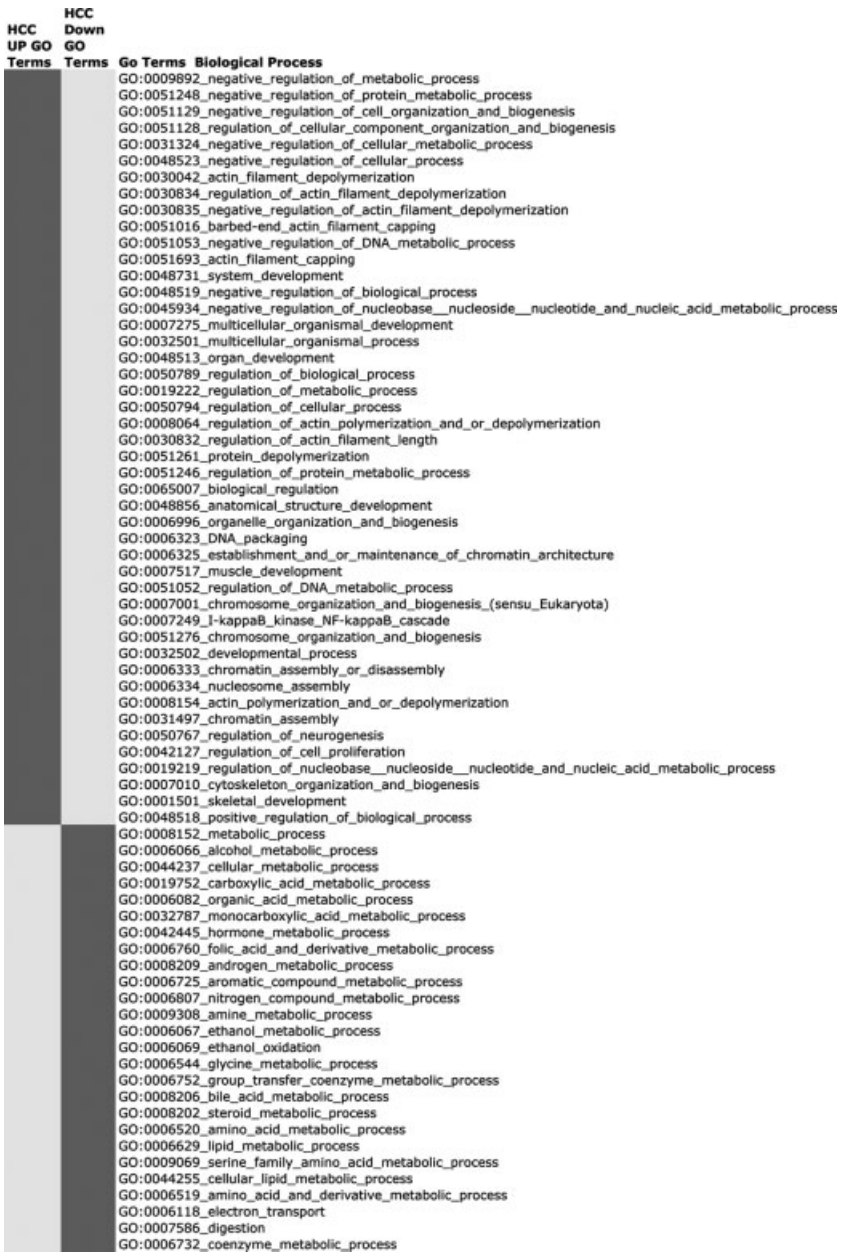

Figure 1. Enriched GO biological process terms for the proteins identified as overexpressed more than fivefold in the HCC group and the normal group. The enriched GO terms with p values <0.05 are shown in dark grey, and those with p values >0.05 are shown in light grey.

3.3 Identification of chitinase-3-like protein 1 (CHI3L1) as a serum biomarker for hepatocellular carcinoma

We searched for differentially expressed proteins that have commercial ELISA kits or good antibodies for easy confirmation analysis. Chitinase-3-like protein 1 (CHI3L1), one of the top-ranking genes in Table 2, has an ELISA kit available. CHI3L1 is a 39-kDa secretory glycoprotein and a member of the chitinase protein family and it plays role in macrophage differentiation and tissue remodeling [15, 16]]. We found that CHI3L1 was only identified in the HCC secretome, but never in the normal secretome (Table 2). In the transcriptomics analysis, it is one of the top genes that showed differential expression (3335 tpm in HCC *vs.* 15 tpm in the uninvolved surrounding tissue). In the liver cancer array data published by Chen [12], CHI3L1 was shown to be expressed higher in the cancer tissues compared to the normal tissues (*t*-test: 3.742; *p*-value: 2.5E-4; data from www.oncomine.org).

We purchased CHI3L1 (alias YKL-40) ELISA kits and used them for evaluation of this biomarker directly in 25 serum samples of liver cancer patients and 15 age-matched normal controls. The mean and range of serum expression levels are shown as box-and-whisker plots (Fig. 2, top panel). The detailed clinical information of the cohort is provided in

Table 2. Top candidates identified by integrative proteomics and transcriptomics analysis

ID	GeneID	Symbol	Description	Trans-criptomics HCC (tpm)	Trans-criptomics normal (tpm)	Proteomics Spectral counting (HCC)[a)]	Proteomics Spectral counting (normal)
IPI00002147	1116.0	CHI3L1	Chitinase-3-like protein 1 precursor	3335	15	4.3	0
IPI00220827	9168.0	TMSB10	Thymosin beta-10	2038	217	7.2	0
IPI00029737	2182.0	ACSL4	Isoform Long of Long-chain-fatty-acid–CoA ligase 4	1337	14	5.8	0
IPI00008527	6176.0	RPLP1	60S acidic ribosomal protein P1	1691	551	15.9	5
IPI00032292	7076.0	TIMP1	Metalloproteinase inhibitor 1 precursor	1277	194	53.6	0
IPI00221093	6218.0	RPS17	40S ribosomal protein S17	1587	506	8.7	0
IPI00012011	1072.0	CFL1	Cofilin-1	1165	275	173.9	45
IPI00007221	5104.0	SERPINA5	Plasma serine protease inhibitor precursor	1015	257	7.2	0
IPI00418169	302.0	ANXA2	Annexin A2 isoform 1	871	122	68.1	10
IPI00221224	290.0	ANPEP	Aminopeptidase N	980	287	142.0	14
IPI00006114	5176.0	SERPINF1	Pigment epithelium-derived factor precursor	826	229	137.7	5
IPI00219038	3020.0	H3F3A	Histone H3.3	699	140	97.1	20
IPI00465439	226.0	ALDOA	Fructose-bisphosphate aldolase A	640	83	323.2	64
IPI00014572	6678.0	SPARC	SPARC precursor	573	97	4.3	0
IPI00024320	5935.0	RBM3	Putative RNA-binding protein 3	462	49	7.2	0
IPI00221222	10923.0	SUB1	Activated RNA polymerase II transcriptional coactivator p15	452	81	85.5	27
IPI00215914	375.0	ARF1	ADP-ribosylation factor 1	495	129	30.4	11
IPI00216308	7416.0	VDAC1	Voltage-dependent anion-selective channel protein 1	517	153	4.3	0
IPI00020956	3068.0	HDGF	Hepatoma-derived growth factor	476	149	144.9	32
IPI00514127	3068.0	HDGF	Hepatoma-derived growth factor	476	149	124.6	29
IPI00022443	174.0	AFP	Alpha-fetoprotein precursor	329	2	2.9	0
IPI00014587	1211.0	CLTA	Isoform Brain of Clathrin light chain A	387	73	58.0	15
IPI00023673	3959.0	LGALS3BP	Galectin-3-binding protein precursor	274	40	17.4	7
IPI00302592	2316.0	FLNA	Filamin A, alpha	251	32	1153.6	214
IPI00029266	6635.0	SNRPE	Small nuclear ribonucleoprotein E	247	47	40.6	9
IPI00010790	633.0	BGN	Biglycan precursor	259	71	58.0	5
IPI00304612	23521.0	RPL13A	60S ribosomal protein L13a	244	55	4.3	0
IPI00027285	6628.0	SNRPB	Isoform SM-B′ of Small nuclear ribonucleoprotein-associated proteins B and B′	200	17	37.7	17
IPI00009750	3960.0	LGALS4	Galectin-4	196	21	21.7	10
IPI00304962	1278.0	COL1A2	Collagen alpha-2(I) chain precursor	240	65	2.9	0
IPI00643041	5901.0	RAN	GTP-binding nuclear protein Ran	238	66	15.9	0
IPI00025491	1973.0	EIF4A1	Eukaryotic initiation factor 4A-I	239	68	68.1	5
IPI00010896	1192.0	CLIC1	Chloride intracellular channel protein 1	203	34	40.6	2
IPI00296099	7057.0	THBS1	Thrombospondin-1 precursor	177	22	333.3	7
IPI00030910	4076.0	CAPRIN1	GPI-anchored membrane protein 1	212	62	15.9	3
IPI00550363	8407.0	TAGLN2	Transgelin-2	197	51	194.2	68
IPI00413778	2280.0	FKBP1A	FKBP1A protein	200	57	50.7	17
IPI00292452	80760.0	ITIH5	Inter-alpha (Globulin) inhibitor H5	141	2	4.3	0
IPI00298971	7448.0	VTN	Vitronectin precursor	158	21	55.1	11
IPI00296922	3913.0	LAMB2	Laminin subunit beta-2 precursor	177	40	24.6	0
IPI00297646	1277.0	COL1A1	Collagen alpha-1(I) chain precursor	152	17	5.8	0
IPI00021263	7534.0	YWHAZ	14-3-3 protein zeta/delta	173	52	344.9	160
IPI00000760	23564.0	DDAH2	NG,NG-dimethylarginine dimethylaminohydrolase 2	151	33	23.2	3
IPI00026944	4811.0	NID1	Isoform 1 of Nidogen-1 precursor	140	23	37.7	0
IPI00554737	5518.0	PPP2R1A	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	145	31	26.1	0
IPI00007118	5054.0	SERPINE1	Plasminogen activator inhibitor 1 precursor	115	5	30.4	0
IPI00217296	5524.0	PPP2R4	Isoform 3 of Serine/threonine-protein phosphatase 2A regulatory subunit B′	121	14	7.2	3
IPI00021700	5111.0	PCNA	Proliferating cell nuclear antigen	139	33	24.6	0
IPI00034319	51596.0	CUTA	Isoform A of Protein CutA precursor	154	51	36.2	16
IPI00306322	1284.0	COL4A2	Collagen alpha-2(IV) chain precursor	126	25	5.8	0
IPI00003966	768.0	CA9	Carbonic anhydrase 9 precursor	100	1	10.1	0

Table 2. Continued

ID	GeneID	Symbol	Description	Trans-criptomics HCC (tpm)	Trans-criptomics normal (tpm)	Proteomics Spectral counting (HCC)[a)]	Proteomics Spectral counting (normal)
IPI00215911	328.0	APEX1	DNA-(apurinic or apyrimidinic site) lyase	138	42	42.0	11
IPI00399319	11316.0	COPE	Epsilon subunit of coatomer protein complex isoform c	115	20	5.8	0
IPI00217465	3006.0	HIST1H1C	Histone H1.2	133	42	92.7	0
IPI00010133	11151.0	CORO1A	Coronin-1A	109	17	10.1	0
IPI00396321	55379.0	LRRC59	Leucine-rich repeat-containing protein 59	123	35	33.3	13
IPI00028946	10313.0	RTN3	Isoform 3 of Reticulon-3	101	14	2.9	0
IPI00168479	128240.0	APOA1BP	Apolipoprotein A-I binding protein precursor	112	26	84.1	31
IPI00743696	1282.0	COL4A1	161 kDa protein	103	20	11.6	0
IPI00418262	230.0	ALDOC	Fructose-bisphosphate aldolase C	96	14	30.4	14
IPI00297160	960.0	CD44	Isoform 12 of CD44 antigen precursor	108	26	5.8	0
IPI00032957	7329.0	UBE2I	SUMO-conjugating enzyme UBC9	110	32	11.6	0
IPI00443909	10330.0	CNPY2	Isoform 1 of MIR-interacting saposin-like protein precursor	115	38	21.7	5
IPI00059366	9555.0	H2AFY	H2A histone family, member Y isoform 2	100	23	87.0	0
IPI00018931	55737.0	VPS35	Vacuolar protein sorting-associated protein 35	104	27	8.7	0
IPI00293464	1642.0	DDB1	DNA damage-binding protein 1	112	37	21.7	7
IPI00479997	3925.0	STMN1	Stathmin	91	17	33.3	0
IPI00374563	375790.0	AGRN	Agrin precursor	81	10	5.8	0
IPI00290770	7203.0	CCT3	Chaperonin containing TCP1, subunit 3 isoform b	92	21	15.9	0
IPI00438229	10155.0	TRIM28	Isoform 1 of Transcription intermediary factor 1-beta	94	24	14.5	0
IPI00032325	1475.0	CSTA	Cystatin-A	81	11	2.9	0
IPI00029468	10121.0	ACTR1A	Alpha-centractin	102	32	4.3	0
IPI00216153	6209.0	RPS15	40S ribosomal protein S15	83	18	5.8	0
IPI00026833	159.0	ADSS	Adenylosuccinate synthetase isozyme 2	76	13	2.9	0
IPI00014361	7264.0	TSTA3	GDP-L-fucose synthetase	81	20	15.9	7
IPI00018769	7058.0	THBS2	Thrombospondin-2 precursor	75	14	113.0	0
IPI00012535	3301.0	DNAJA1	DnaJ homolog subfamily A member 1	80	18	26.1	0
IPI00012382	6626.0	SNRPA	U1 small nuclear ribonucleoprotein A	70	16	18.8	8
IPI00005198	3608.0	ILF2	Interleukin enhancer-binding factor 2	67	14	95.6	38
IPI00016862	2936.0	GSR	Isoform Mitochondrial of Glutathione reductase, mitochondrial precursor	77	25	63.8	27
IPI00018768	7247.0	TSN	Translin	78	26	46.4	20
IPI00013894	10963.0	STIP1	Stress-induced-phosphoprotein 1	66	16	133.3	7
IPI00105407	57016.0	AKR1B10	Aldo-keto reductase family 1 member B10	61	12	466.6	30
IPI00455033	10801.0	39334.0	Isoform 3 of Septin-9	58	9	4.3	0
IPI00018236	2760.0	GM2A	Ganglioside GM2 activator precursor	65	15	2.9	0
IPI00006052	5202.0	PFDN2	Prefoldin subunit 2	67	21	18.8	0
IPI00014589	1212.0	CLTB	Isoform Brain of Clathrin light chain B	65	21	14.5	3
IPI00306825	7165.0	TPD52L2	Isoform 1 of Tumor protein D54	55	12	8.7	0
IPI00007138	22839.0	DLGAP4	Isoform 1 of Disks large-associated protein 4	50	8	4.3	0
IPI00032516	8907.0	AP1M1	AP-1 complex subunit mu-1	54	12	4.3	0
IPI00030968	286257.0	C9orf142	Uncharacterized protein C9orf142	44	5	2.9	0
IPI00478390	1029.0	CDKN2A	Cyclin-dependent kinase inhibitor 2A, isoform 4	40	1	2.9	0
IPI00001560	1029.0	CDKN2A	Isoform 1 of Cyclin-dependent kinase inhibitor 2A, isoforms 1/2/3	40	1	2.9	0
IPI00014537	813.0	CALU	Isoform 1 of Calumenin precursor	53	15	11.6	0
IPI00456750	64855.0	FAM129B	Niban-like protein	45	8	5.8	0
IPI00010154	2664.0	GDI1	Rab GDP dissociation inhibitor alpha	51	17	4.3	2
IPI00033025	989.0	39332.0	Isoform 1 of Septin-7	46	13	27.5	4
IPI00032460	57819.0	LSM2	U6 snRNA-associated Sm-like protein LSm2	40	7	8.7	3
IPI00411426	9559.0	VPS26A	Vacuolar protein sorting-associated protein 26A	47	15	4.3	0

Table 2. Continued

ID	GeneID	Symbol	Description	Trans-criptomics HCC (tpm)	Trans-criptomics normal (tpm)	Proteomics Spectral counting (HCC)[a)]	Proteomics Spectral counting (normal)
IPI00294158	644.0	BLVRA	Biliverdin reductase A precursor	45	12	2.9	0
IPI00022078	10397.0	NDRG1	Protein NDRG1	45	14	7.2	0
IPI00020672	10072.0	DPP3	Isoform 1 of Dipeptidyl-peptidase 3	34	4	20.3	8
IPI00013454	6001.0	RGS10	Isoform 1 of Regulator of G-protein signaling 10	40	10	2.9	0

a) Normalized by total spectral numbers.

Supporting Information Table 5. ROC curve analysis revealed that CHI3L1 has an AUC (area under the ROC curve) of 0.92. With the cut-off value of 57.3, it has a sensitivity of 0.92, a specificity of 0.87, and an accuracy of 0.9 (Fig. 3A).

3.4 Combining liver-enriched expression biomarker MASP2 with CHI3L1 for liver cancer diagnosis

As YKL-40 seems to be a generic marker for multiple types of cancers including glioma [17], endometrial cancer [18] and ovarian cancer [19], we sought to add a tissue-enriched expression gene, in this case, a liver-enriched expression gene, that is secreted and changed with diseases. We have previously proposed to identify tissue-enriched expression genes as biomarkers [20]. We found that MASP2 (mannan-binding lectin serine peptidase 2) is a liver tissue-enriched expression gene and it is predicted to be a secreted protein, as it contains a signal peptide (Signal peptide probability: 1.000 by SignalP3.0 prediction server http://www.cbs.dtu.dk/services/SignalP/) and additionally has an odd score of 2.52 as a secreted protein predicted by the SecretomeP program (http://www.cbs.dtu.dk/services/SecretomeP/). MASP2 is the 2nd member of the serum proteases identified to play an important role in the activation of the complement system *via* mannose-binding lectin [21]. Our MPSS data (data not shown) as well as Chen's liver cancer array [12] data indicate this gene is down-regulated in normal liver compared to liver cancer tissues. We hypothesize that liver cancer may affect normal liver functions and therefore down-regulate the expression of liver-enriched expression genes. We purchased MASP2 ELISA kits and used them for evaluation of this biomarker directly in 25 serum samples of liver cancer patients and 15 age-matched normal controls (Supporting Information Table 5). The mean and range of serum expression levels are shown as box-and-whisker plots (Fig. 2, bottom panel). ROC curve analysis revealed MASP2 has an AUC of 0.83. Using a cut-off value of 292.9, the sensitivity is 0.87, specificity is 0.77 and accuracy is 0.8 (Fig. 3B). We then calculated the ratio of YKL40 and MASP2 as a combination marker for liver cancer diagnosis. We found that using a cut-off value of 14.36 increased detecting sensitivity over using either markers alone (sensitivity of 0.96 for the combined marker *versus* 0.92 and 0.87 for individual markers), the accuracy remaining at 0.9 (Fig. 3C) and the specificity at 0.8 (higher than using MASP2 alone but lower than using CHI3L1 alone with their respective cut-off values described previously). The AUC of YKL40/MASP2 ratio for liver cancer diagnosis is 0.97, which is greater than the AUC of MASP2 (0.83) and that of YKL-40 (0.92).

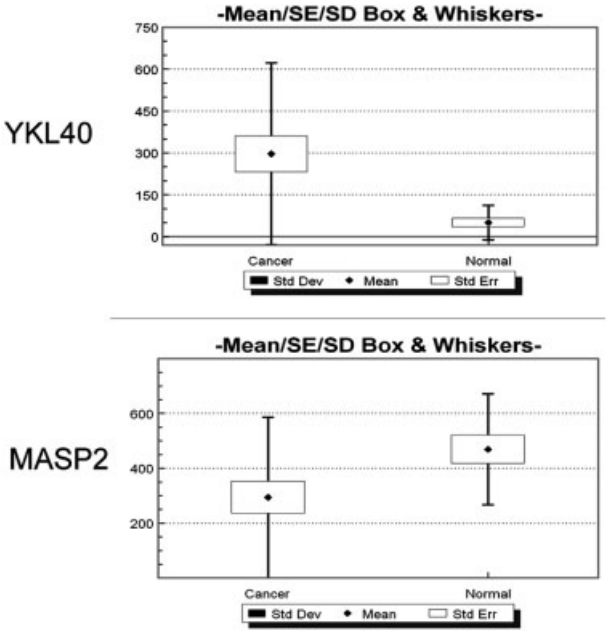

Figure 2. Box-and-whisker plots showing the mean, standard deviation (Std Dev) and standard error (Std Err) of CHI3L1 and MASP2 in 25 serum samples of liver cancer patients and 15 age-matched normal controls.

4 Discussion

Identification of proteins from tissue interstitial fluids or conditioned cell culture media as biomarker candidates and therapeutic targets has been proposed previously [22, 23]. Here, we apply this similar approach in identifying potential secreted proteins in *ex vivo* liver cancer tissue culture as biomarker candidates. In GO analysis, the proteins in the *ex vivo* tissue culture media also include many intracellular pro-

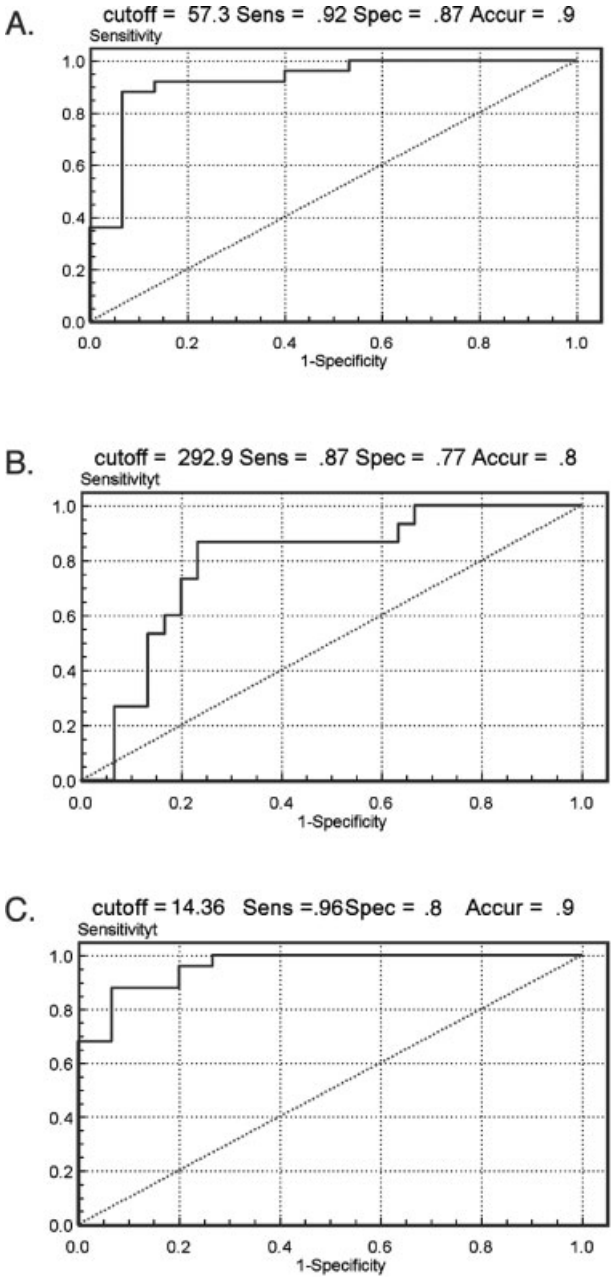

Figure 3. ROC curve of CHI3L1 (A), MASP2 (B) and CHI3L1/MASP2 ratio (C) indicating its AUC, specificity, and sensitivity in detecting HCC.

teins. This could be because some proteins have multiple cellular localizations. In addition, it is likely that during the tissue culture, some cells died, resulting in the release of intracellular proteins into the media. This scenario has been observed in previous publications of cell culture media proteomics analyses [23, 24]. We have tried to minimize these effects by limiting the length of the time of tissue culture. Nonetheless, we found that about 42% of the identified proteins were predicted as secreted proteins, which is much higher than the percentage predicted as secreted proteins from cell lysate proteomics data and much higher than that from a randomly selected gene list (data not shown).

We employed a simple spectral counting approach [10] to determine the relative abundance and succeeded in identifying around 190 differentially expressed proteins with at least fivefold over- or underexpression in the HCC secretome compared to the normal secretome (Supporting Information Tables 3 and 4). Spectral counting was shown to be a valid method for quantitative proteomic analysis [25]. As we analyzed 12 pairs of samples, spectral counting method was clearly much easier to implement compared with stable isotopic labeling and quantification. In spectral counting, the higher the spectrum numbers are obtained for comparison, the more accurate the comparison will be. The spectral numbers for individual cancer and normal pairs are often small (<10) except for those abundant proteins, and comparison between them by spectral counting method would not be reliable. To increase our sensitivity and confidence in detecting the differences, we first summed up the total spectrum numbers in the HCC group and the control group, and then performed the comparison.

Many proteins we identified as over expressed in the HCC secretome have been reported previously as HCC biomarkers [4, 26–29]. We showed that many members of the heat shock protein family were overexpressed (Table 2) including heat shock 70-kDa protein (HSPA5) (spectrum counting: 1023/102; HCC/normal), heat shock 70-kDa protein 4 (HSPA4) (spectrum counting: 29/6; HCC/normal), isoform 1 of heat shock 71-kDa protein (spectrum counting: 848/445), and heat shock protein HSP 90-alpha 2 (spectrum counting: 355/161). Sun *et al.* [26] analyzed HCC tissue by 2D-DIGE and showed that several HSP were up-regulated in the HCC tissues. We showed that apolipoprotein E (APOE) was over expressed in the HCC group (359 counts) compared to the normal group (2 counts). It was identified as up regulated in HCC by 2-DE and validated by Western blot [29]. We also identified other previously reported HCC-associated proteins, such as hepatoma-derived growth factor (HDGF) [30, 31], GST [32–34], and aldolase A [35, 36].

Some of the secretome proteins we identified have proved to be good serum biomarkers for liver cancer. We identified calreticulin and protein disulfide-isomerase A3 (PDIA3) as overexpressed in the HCC secretome compared to the control secretome (spectral counting: 758 in HCC *vs.* 223 in the control for calreticulin; 699 in HCC *vs.* 227 in the control for PDIA3). Chignard *et al.* [37] showed a statistically highly significant difference in calreticulin and PDIA3 fragment serum levels between patients with HCC and healthy individuals. Interestingly, we showed that, in addition to PDIA3, PDIA2 (115 in the HCC *vs.* 0 in the control), PDIA4 (383 in the HCC *vs.* 113 in the control), and PDIA6 (226 in the HCC *vs.* 76 in the control) were expressed significantly higher in the HCC compared to the control (Supporting Information Tables 1 and 2).

From our secretome analysis, we identified CHI3L1 (chitinase 3-like 1, also named YKL-40) as a good diagnostic marker for HCC with an AUC (area under the ROC curve) of 0.92. (Fig. 3A). CHI3L1 is a member of the chitinase protein family [15]. The function of CHI3L1 is unknown. It was suggested that it might play roles in tissue remodeling [16]. In addition, YKL-40 were shown to be a prognostic marker for predicting survival time for colorectal and breast cancers [38, 39] and a diagnostic marker for ovarian and endometrial cancers [18, 19].

Johansen *et al.* [40] showed that serum YKL-40 (CHI3L1) is increased in patients with hepatic fibrosis, including alcoholic cirrhosis, post-hepatitic cirrhosis, and non-cirrhotic fibrosis. More recently, YKL-40 expression level was shown to be associated with HCV-related fibrosis [41, 42]. As our patient cohorts have severe fibrosis (fibrosis stage 4, Table 1 and Supporting Information Table 5), the increase in YKL-40 levels might be related to liver fibrosis caused by liver cancer or HBV infections in our patient cohort. Liver fibrosis can be caused by viral infections (*e.g.* HBV or HCV), chemicals, or cancer cell growth [43]. Further studies comparing patients with fibrosis but without cancer, to those patients with fibrosis and liver cancer, and to those patients with liver cancer but without fibrosis will be necessary to determine the role of YKL-40 in liver cancer diagnosis.

To increase our ability to differentiate whether the increase of YKL-40 is due to liver cancer disease *versus* other type of cancers, we included a liver-enriched expression gene MASP2. Combining these two markers increased the AUC for liver cancer diagnosis. The AUC of YKL-40/MASP2 ratio for liver cancer diagnosis is 0.97, which is greater than the AUC of MASP2 (0.83) and that of YKL-40 (0.92). However, as we compared HCC to normal controls, instead of comparing HCC to those patients with liver fibrosis but without cancer, the AUC for liver cancer diagnosis might be overestimated. Furthermore, the sample size in this study is small, further evaluation using a separate cohort and larger sample size will be necessary to determine the true utility of this biomarker pair. In addition, a follow-up study will also be necessary to study HCC patients *versus* disease controls with the same type and severity of liver disease (*e.g.* fibrosis) to determine whether the combination of YKL40 and MASP2 marker pair can help to differentiate HCC from other liver diseases.

In summary, we present a novel paradigm in combining biomarkers that include (i) a cancer biomarker that is increased in cancer patients compared to normal individual, and (ii) an organ-enriched expression markers that is down regulated in cancer due to loss of organ function which may help to identify the tissue/organ origin of cancer even when the first cancer biomarker is not cancer-type specific. We have applied this approach in identifying a novel liver cancer biomarker pair CHI3L1 and MASP2. Other candidate biomarkers we identified (Supporting Information Tables 3 and 4) can be further evaluated using stable isotopic labeling and quantification using MS even when ELISA assays are not available or too expensive to be developed. Our approach might be a useful general approach to find cancer biomarkers that may eventually contribute to a panel consisting of multiple markers for assessing normal function or disease states of an organ.

This work was supported by grants from the National Infrastructure of Natural Resources for Science and Technology, Ministry of Science and Technology, China (2005DKA21301, 2006DFA32950, 2006AA02Z4A2), the National Basic Research Program of China (2003CB515501), the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT 0753), and from R&D Special Fund for the health industry (200802006) from China. It was also supported by the grant U54DA021519 from NIH, USA.

The authors have declared no conflict of interest.

5 References

[1] Parkin, D. M., Bray, F., Ferlay, J., Pisani, P., Estimating the world cancer burden: Globocan 2000. *Int. J. Cancer* 2001, *94*, 153–156.

[2] Spangenberg, H. C., Thimme, R., Blum, H. E., Serum markers of hepatocellular carcinoma. *Semin. Liver Dis.* 2006, *26*, 385–390.

[3] Fujiyama, S., Tanaka, M., Maeda, S., Ashihara, H. *et al.*, Tumor markers in early diagnosis, follow-up and management of patients with hepatocellular carcinoma. *Oncology* 2002, *62* Suppl. 1, 57–63.

[4] Li, C., Tan, Y. X., Zhou, H., Ding, S. J. *et al.*, Proteomic analysis of hepatitis B virus-associated hepatocellular carcinoma: Identification of potential tumor markers. *Proteomics* 2005, *5*, 1125–1139.

[5] Tu, D. G., Wang, S. T., Chang, T. T., Chiu, N. T., Yao, W. J., The value of serum tissue polypeptide specific antigen in the diagnosis of hepatocellular carcinoma. *Cancer* 1999, *85*, 1039–1043.

[6] Buscarini, L., Sbolli, G., Cavanna, L., Civardi, G. *et al.*, Clinical and diagnostic features of 67 cases of hepatocellular carcinoma. *Oncology* 1987, *44*, 93–97.

[7] Anderson, L., Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J. Physiol.* 2005, *563*, 23–60.

[8] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.

[9] Zeeberg, B. R., Feng, W., Wang, G., Wang, M. D. *et al.*, GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* 2003, *4*, R28.

[10] Liu, H., Sadygov, R. G., Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 2004, *76*, 4193–4201.

[11] Lin, B., White, J. T., Lu, W., Xie, T. *et al.*, Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. *Cancer Res.* 2005, *65*, 3081–3091.

[12] Chen, X., Cheung, S. T., So, S., Fan, S. T. *et al.*, Gene expression patterns in human liver cancers. *Mol. Biol. Cell* 2002, *13*, 1929–1939.

[13] Lee, J. S., Chu, I. S., Heo, J., Calvisi, D. F. *et al.*, Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology (Baltimore), Md* 2004, *40*, 667–676.

[14] Lee, J. S., Thorgeirsson, S. S., Genome-scale profiling of gene expression in hepatocellular carcinoma: classification, survival prediction, and identification of therapeutic targets. *Gastroenterology* 2004, *127*, S51–55.

[15] Rehli, M., Krause, S. W., Andreesen, R., Molecular characterization of the gene for human cartilage gp-39 (CHI3L1), a member of the chitinase protein family and marker for late stages of macrophage differentiation. *Genomics* 1997, *43*, 221–225.

[16] Bleau, G., Massicotte, F., Merlen, Y., Boisvert, C., Mammalian chitinase-like proteins. *Exs* 1999, *87*, 211–221.

[17] Hormigo, A., Gu, B., Karimi, S., Riedel, E. *et al.*, YKL-40 and matrix metalloproteinase-9 as potential serum biomarkers for patients with high-grade gliomas. *Clin. Cancer Res.* 2006, *12*, 5698–5704.

[18] Diefenbach, C. S., Shah, Z., Iasonos, A., Barakat, R. R. *et al.*, Preoperative serum YKL-40 is a marker for detection and prognosis of endometrial cancer. *Gynecol. Oncol.* 2007, *104*, 435–442.

[19] Dupont, J., Tanwar, M. K., Thaler, H. T., Fleisher, M. *et al.*, Early detection and prognosis of ovarian cancer using serum YKL-40. *J. Clin. Oncol.* 2004, *22*, 3330–3339.

[20] Hood, L., Heath, J. R., Phelps, M. E., Lin, B., Systems biology and new technologies enable predictive and preventative medicine. *Science* 2004, *306*, 640–643.

[21] Thiel, S., Vorup-Jensen, T., Stover, C. M., Schwaeble, W. *et al.*, A second serine protease associated with mannan-binding lectin that activates complement. *Nature* 1997, *386*, 506–510.

[22] Celis, J. E., Gromov, P., Cabezon, T., Moreira, J. M. *et al.*, Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol. Cell. Proteomics* 2004, *3*, 327–344.

[23] Kulasingam, V., Diamandis, E. P., Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Mol. Cell. Proteomics* 2007, *6*, 1997–2011.

[24] Martin, D. B., Gifford, D. R., Wright, M. E., Keller, A. *et al.*, Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. *Cancer Res.* 2004, *64*, 347–355.

[25] Zybailov, B., Coleman, M. K., Florens, L., Washburn, M. P., Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.* 2005, *77*, 6218–6224.

[26] Sun, W., Xing, B., Sun, Y., Du, X. *et al.*, Proteome analysis of hepatocellular carcinoma by two-dimensional difference gel electrophoresis: novel protein markers in hepatocellular carcinoma tissues. *Mol. Cell. Proteomics* 2007.

[27] Li, C., Hong, Y., Tan, Y. X., Zhou, H. *et al.*, Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* 2004, *3*, 399–409.

[28] Fujii, K., Kondo, T., Yokoo, H., Yamada, T. *et al.*, Proteomic study of human hepatocellular carcinoma using two-dimensional difference gel electrophoresis with saturation cysteine dye. *Proteomics* 2005, *5*, 1411–1422.

[29] Blanc, J. F., Lalanne, C., Plomion, C., Schmitter, J. M. *et al.*, Proteomic analysis of differentially expressed proteins in hepatocellular carcinoma developed in patients with chronic viral hepatitis C. *Proteomics* 2005, *5*, 3778–3789.

[30] Hu, T. H., Huang, C. C., Liu, L. F., Lin, P. R. *et al.*, Expression of hepatoma-derived growth factor in hepatocellular carcinoma. *Cancer* 2003, *98*, 1444–1456.

[31] Kishima, Y., Yoshida, K., Enomoto, H., Yamamoto, M. *et al.*, Antisense oligonucleotides of hepatoma-derived growth factor (HDGF) suppress the proliferation of hepatoma cells. *Hepato-Gastroenterol.* 2002, *49*, 1639–1644.

[32] Sakai, M., Muramatsu, M., Regulation of glutathione transferase P: a tumor marker of hepatocarcinogenesis. *Biochem. Biophys. Res. Commun.* 2007, *357*, 575–578.

[33] Takekoshi, H., Mizoguchi, T., Komasa, Y., Chubachi, H. *et al.*, Suppression of glutathione S-transferase placental form-positive foci development in rat hepatocarcinogenesis by *Chlorella pyrenoidosa*. *Oncol. Reports* 2005, *14*, 409–414.

[34] Zhou, T., Evans, A. A., London, W. T., Xia, X. *et al.*, Glutathione S-transferase expression in hepatitis B virus-associated human hepatocellular carcinogenesis. *Cancer Res.* 1997, *57*, 2749–2753.

[35] Castaldo, G., Calcagno, G., Sibillo, R., Cuomo, R. *et al.*, Quantitative analysis of aldolase A mRNA in liver discriminates between hepatocellular carcinoma and cirrhosis. *Clin. Chem.* 2000, *46*, 901–906.

[36] Zeindl-Eberhart, E., Jungblut, P., Rabes, H. M., Expression of tumor-associated protein variants in chemically induced rat hepatomas and transformed rat liver cell lines determined by two-dimensional electrophoresis. *Electrophoresis* 1994, *15*, 372–381.

[37] Chignard, N., Shang, S., Wang, H., Marrero, J. *et al.*, Cleavage of endoplasmic reticulum proteins in hepatocellular carcinoma: Detection of generated fragments in patient sera. *Gastroenterology* 2006, *130*, 2010–2022.

[38] Cintin, C., Johansen, J. S., Christensen, I. J., Price, P. A. *et al.*, Serum YKL-40 and colorectal cancer. *Brit. J. Cancer* 1999, *79*, 1494–1499.

[39] Johansen, J. S., Cintin, C., Jorgensen, M., Kamby, C., Price, P. A., Serum YKL-40: a new potential marker of prognosis and location of metastases of patients with recurrent breast cancer. *Eur. J. Cancer* 1995, *31A*, 1437–1442.

[40] Johansen, J. S., Christoffersen, P., Moller, S., Price, P. A. *et al.*, Serum YKL-40 is increased in patients with hepatic fibrosis. *J. Hepatol.* 2000, *32*, 911–920.

[41] Kamal, S. M., Turner, B., He, Q., Rasenack, J. *et al.*, Progression of fibrosis in hepatitis C with and without schistosomiasis: correlation with serum markers of fibrosis. *Hepatology (Baltimore), Md* 2006, *43*, 771–779.

[42] Fontana, R. J., Goodman, Z. D., Dienstag, J. L., Bonkovsky, H. L. *et al.*, Relationship of serum fibrosis markers with liver fibrosis stage and collagen content in patients with advanced chronic hepatitis C. *Hepatology (Baltimore), Md* 2008, *47*, 789–798.

[43] Alcolado, R., Arthur, M. J., Iredale, J. P., Pathogenesis of liver fibrosis. *Clin. Sci. (Lond)* 1997, *92*, 103–112.

