BBA Clinical 3 (2015) 205–213

Contents lists available at ScienceDirect

BBA Clinical

journal homepage: http://www.journals.elsevier.com/bba-clinical/

Onco-proteogenomics identifies urinary S100A9 and GRN as potential combinatorial biomarkers for early diagnosis of hepatocellular carcinoma

Chun-Hao Huang [a,b,1], Chao-Jen Kuo [a,1], Shih-Shin Liang [c], Shu-Wen Chi [a], Edward Hsi [d], Chi-Chao Chen [e], King-Teh Lee [f,*], Shyh-Horng Chiou [a,b,g,**]

[a] Quantitative Proteomics Center and Graduate Institute of Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan
[b] Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan
[c] Department of Biotechnology, College of Life Science, Kaohsiung Medical University, Kaohsiung 807, Taiwan
[d] Department of Genome Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan
[e] Physiology Biophysics and Systems Biology Program, Weill Graduate School of Medical Sciences, Cornell University, NY 10065, USA
[f] Division of Hepatobiliary Surgery, Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 807, Taiwan
[g] Center for Infectious Disease and Cancer Research, Kaohsiung Medical University, Kaohsiung 807, Taiwan

ARTICLE INFO

Article history:
Received 2 January 2015
Received in revised form 23 February 2015
Accepted 24 February 2015
Available online 3 March 2015

Keywords:
Hepatocellular carcinoma
Proteogenomics
Quantitative proteomics
Urinary biomarkers
Protein S100-A9
Granulin

ABSTRACT

Hepatocellular carcinoma (HCC), the major type of liver cancer, is among the most lethal cancers owing to its aggressive nature and frequently late detection. Therefore, the possibility to identify early diagnostic markers could be of significant benefit. Urine has especially become one of the most attractive body fluids in biomarker discovery as it can be obtained non-invasively in large quantities and is stable as compared with other body fluids. To identify potential protein biomarker for early diagnosis of HCC, we explored protein expression profiles in urine from HCC patients and normal controls ($n = 44$) by shotgun proteomics using nano-liquid chromatography coupled tandem mass spectrometry (nanoLC–MS/MS) and stable isotope dimethyl labeling. We have systematically mapped 91 proteins with differential expressions ($p < 0.05$), which included 8 down-regulated microtubule proteins and 83 up-regulated proteins involved in signal and inflammation response. Further integrated proteogenomic approach composed of proteomic, genomic and transcriptomic analysis identified that S100A9 and GRN were co-amplified ($p < 0.001$) and co-expressed ($p < 0.01$) in HCC tumors and urine samples. In addition, the amplifications of *S100A9* or *GRN* were found to be associated with poor survival in HCC patients, and their co-amplification was also prognosed worse overall survival than individual ones. Our results suggest that urinary S100A9 and GRN as potential combinatorial biomarkers can be applied to early diagnosis of hepatocellular carcinoma and highlight the utility of onco-proteogenomics for identifying protein markers that can be applied to disease-oriented translational medicine.

1. Introduction

Hepatocellular carcinoma (HCC), the most common type of liver cancer, is the third leading cause of cancer related mortality worldwide [1]. A major etiological factor for HCC is cirrhosis, frequently caused by chronic infection with hepatitis B or C virus (HBV, HCV), nonalcoholic fatty liver disease, and alcohol abuse [2]. Many patients detected with HCC were initially found to have chronic liver disease and cirrhosis, leading to an increase in the replacement of normal tissue with fibrous tissue leading to the development of HCC [3]. HCC encompasses different pathological manifestations and etiology coupled with multiple genomic aberrations leading to high heterogeneity and intractable treatment. Although the introduction of the multikinase inhibitor sorafenib represents the biggest therapeutic advancement in the past decade, extending life expectancy from 8 to 11 months [4], its limited therapeutic efficacy emphasizes an urgent need for improved targeted therapies, such as CDK9 [5] or MET [6] inhibition. Moreover, these hurdles in developing therapies against HCC highlight the importance of the early detection and biomarker development for non-invasive diagnosis and prognosis.

An abdominal ultrasonography and measurement of serum alpha-fetoprotein (AFP) are two of major tools to detect HCC at an early stage [7]. However, the need of an operator's expertise generally

Abbreviations: HCC, hepatocellular carcinoma; S100A9, protein S100-A9; GRN, granulins; nanoLC–MS/MS, nano-liquid chromatography coupled tandem mass spectrometry; D/H, deuterium/hydrogen labeling; HILIC, hydrophilic interaction chromatography

* Corresponding author. Tel.: +886 7 3121101x7651; fax: +886 7 3216992.
** Correspondence to: S.-H. Chiou, Quantitative Proteomics Center and Graduate Institute of Medicine, Kaohsiung Medical University, Kaohsiung 80708, Taiwan. Tel.: +886 7 3220377; fax: +886 7 3133434.
E-mail addresses: ktlee@kmu.edu.tw (K.-T. Lee), shchiou@kmu.edu.tw (S.-H. Chiou).

[1] Chun-Hao Huang and Chao-Jen Kuo contributed equally to this paper.

http://dx.doi.org/10.1016/j.bbacli.2015.02.004

required on the ultrasonographic evaluation coupled with poor sensitivity as well as specificity of AFP has limited their prognostic use [7]. Recent developments in proteomic technologies have enabled the reliable and high-throughput identification of protein mixtures in biological tissues, building a solid foundation on which to understand the complex proteome profiles from various sample sources such as tissues, cells, plasma and urine [8,9]. Furthermore, the qualitative and quantitative studies of proteins by means of fast-evolving and state-of-the-art proteomic methodologies, such as shotgun proteomics, have provided comprehensive global analysis of protein expression profiles for various types of cancers, including HCC [10–12]. Although onco-proteomics has been extensively applied to biomarker discovery, there is still a limitation to identify tumor-specific peptides due to the complexities of tumor genomes and transcriptomes such as copy-number aberrations, point mutations, unusual splicing variants and gene fusions [13].

Recent exhaustive genomic studies have identified novel genomic amplifications, mutations, and deletions as frequent events in HCC [14,15]. For example, a novel HCC-promoting gene, *FGF19*, was previously found to be amplified in HCC cells harboring the 11q13.3 amplicon [16]. Sawey et al. further showed that 11q13.3 amplification could be an effective biomarker for patients most likely to respond to anti-FGF19 therapy [16]. In addition, advances in gene-expression profiling technologies have enabled the molecular classification of HCCs into defined subclasses, forming a firm basis on which to build more informative clinical trials [17]. To link cancer proteomics to genomics and transcriptomics, onco-proteogenomics as a new research tool and strategy to integrate these different approaches and revolve around detecting various tumor-specific changes in the proteome, leading to the in-depth understanding of tumor initiation, progression and responses to treatment [13,18]. In addition, by incorporating MS-generated data with genomic information can provide a more complete outlook on how cellular networks and canonical signaling pathways are dysregulated in various types of cancers. Several previous studies have applied multiple proteogenomic strategies to characterize cell lines and primary tumors from colorectal cancer, gastric cancer, etc. [19–21].

In the present study, we aim to identify effective and noninvasive diagnostic biomarkers for early detection of HCC. To attain this goal, we have characterized and compared the urinary proteins between diseased and control groups in order to identify potential biomarker candidates by means of gel-free shotgun proteomic analysis coupled with stable isotope dimethyl labeling [8,12,22] and nanoLC–MS/MS [22–24]. Using an integrated proteogenomics, we have investigated

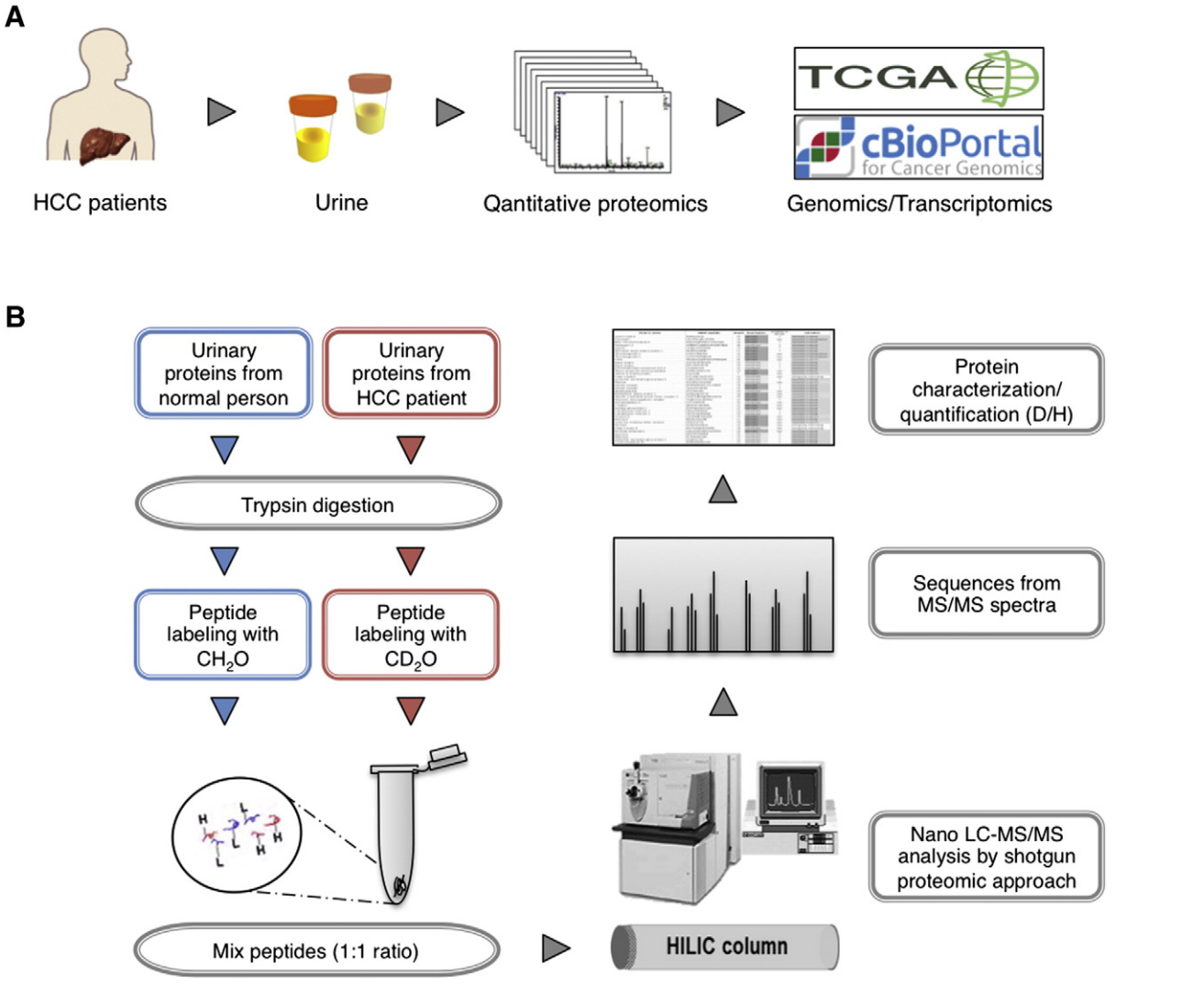

Fig. 1. Using an onco-proteogenomic approach to identify potential urinary biomarkers for HCC. (A) Urine samples from patients who were diagnosed as cases of HCC incidence were collected and analyzed by quantitative proteomics together with urine from normal controls (n = 44). Identified candidates were further investigated and selected using genomic/transcriptomic approaches. (B) Experimental scheme of the procedures used for quantitative proteomics. Upon enzymatic digestion, peptides were differentially stable isotope dimethyl-labeled and combined prior to desalting and fractionation. The quantitative shotgun analysis of proteome changes from clinical urine samples of HCC patients and normal controls was carried out by using HILIC-C18 peptide separation and nanoLC–MS/MS coupled with stable isotope dimethyl labeling.

and selected potential candidates based on their changes on genomic and transcriptomic levels. Gene amplification, mRNA expression and urinary protein levels of both S100A9 and GRN are found to be significantly higher in HCC patients than normal controls. HCC patients with amplification of these two identified genes or their combination also showed poor survival. Our studies establish urinary S100A9 and GRN as potential combinatorial biomarkers for early diagnosis of hepatocellular carcinoma and highlight the utility of onco-proteogenomics for identifying protein biomarkers that can be applied to disease-oriented translational medicine.

2. Materials and methods

2.1. Sample collection

All the procedures used in this study were approved by the ethical committee of clinical research at the Kaohsiung Medical University Hospital. We collected urine from 44 patients who were diagnosed as cases of HCC incidence and never underwent cholecystectomy as the disease group. We also got signed agreements from 44 patients diagnosed with non-HCC incidence but underwent cholecystectomy for urine

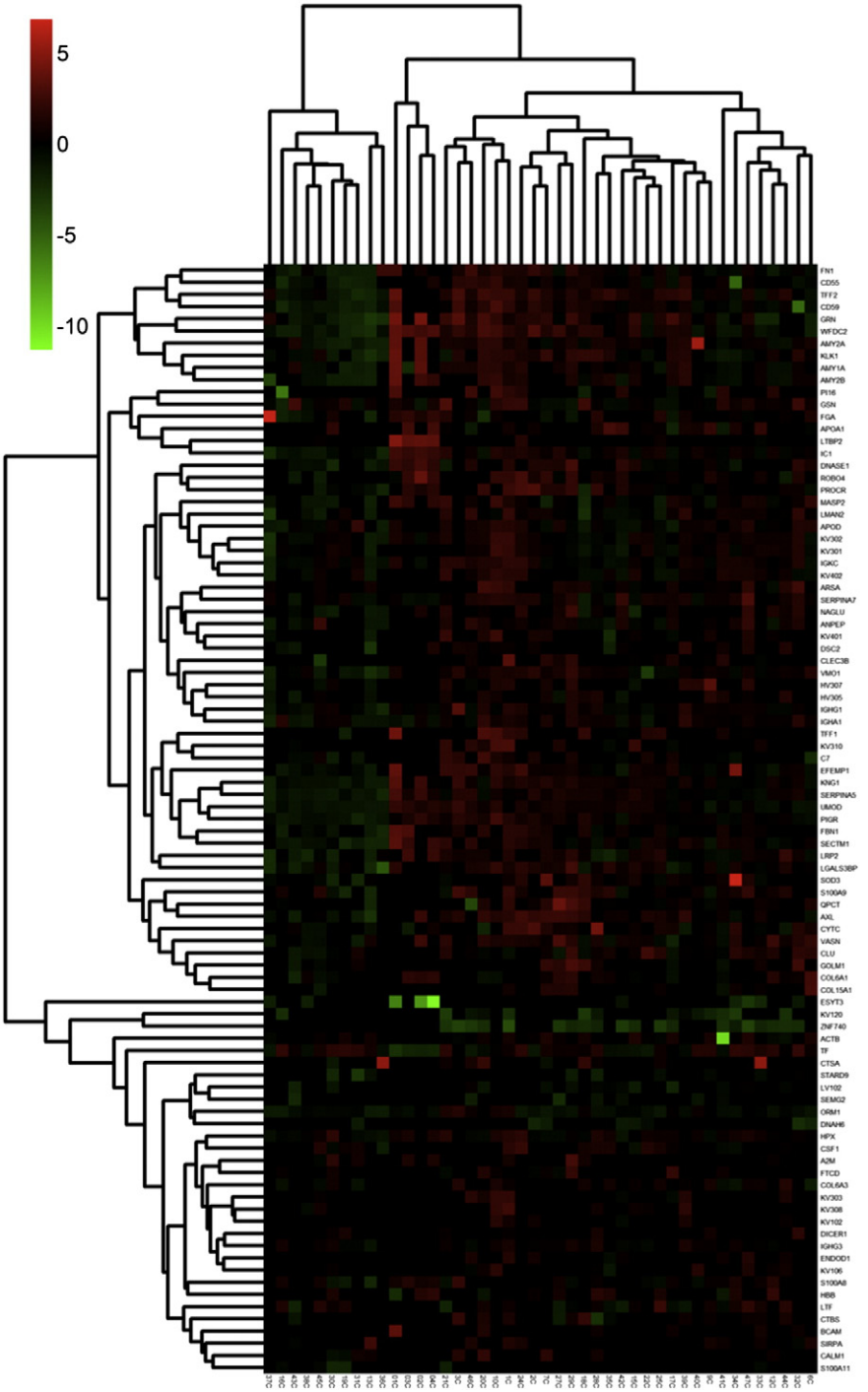

Fig. 2. Hierarchical clustering (HCL) of the proteins differentially expressed (p-value < 0.05) in the urine sample pairs from HCC patients and healthy controls (n = 44). p-Value was calculated by Wilcoxon signed-rank test with paired setting. Heat map was done by R package "pheatmap" with default setting. The value shown in the heatmap is the quantification ratio (D/H) of peptides identified in urine from HCC patients (deuterium labeling) and normal controls (hydrogen labeling).

collection as the normal control. 50 mL urine for each individual was harvested, concentrated by centrifugation and assayed for determination of total protein concentration using Coomassie protein assay reagent, and subsequently was stored at −80 °C until being analyzed.

2.2. Dimethyl labeling and peptide preparation

Volumes of urine containing 100 μg of total proteins were adjusted to 60 μL and treated with 0.7 μL of 1 DTT and 9.3 μL of 7.5% SDS at 95 °C for 5 min before reduction. After the reaction, lysates were further treated with 8 μL of 50 mM IAM at room temperature for alkylation in the dark for 30 min; subsequently proteins were precipitated by adding 52 μL of 50% TCA and incubated on ice for 15 min. After removing the supernatant by centrifugation at 13,000 ×*g* for 5 min, the collected proteins were washed with 150 μL of 10% TCA, vortexed and centrifuged at 13,000 ×*g* for 10 min. The precipitated proteins were washed again with 250 μL distilled H_2O, vortexed and centrifuged thrice under the same condition. The resultant pellets were resuspended with 50 mM NH_4HCO_3 (pH 8.5), then digested with 4 μg of trypsin for 8 h at 37 °C and further dried in a vacuum centrifuge to remove NH_4HCO_3. The lyophilized peptides from HCC and normal urine re-dissolved in 180 μL of 100 mM sodium acetate at pH 5.5 were treated with 20 μL of 4% formaldehyde-H_2 and 20 μL 4% formaldehyde-D_2, respectively [22–24] and mixed thoroughly. The mixtures were vortexed for 5 min, immediately followed by the addition of 10 μL of 0.6 M sodium cyanoborohydride and vortexed for 1 h at room temperature. The resultant liquids were acidified by 10% TFA/H_2O to pH 2.0–3.0 and applied onto the in-house reverse-phase C18 column pre-equilibrated with 200 μL of 0.1% TFA/H_2O (pH 2.0–3.0) for desalting. The column was also washed with 200 μL of 0.1% TFA/H_2O (pH 3.0) and then eluted with a stepwise ACN gradient from 50% to 100% in 0.1% TFA at room temperature. Trichloroacetic acid (TCA), trifluoroacetic acid (TFA), dithiothreitol (DTT), iodoacetamide (IAM), ethylenediaminetetraacetic acid (EDTA), sodium deoxycholate, sodium fluoride (NaF), formaldehyde-H_2, formaldehyde-D_2 and ammonium bicarbonate (NH_4HCO_3), and Triton X-100 were purchased from Sigma Aldrich (St. Louis, MO). Acetonitrile (ACN) and sodium phosphate were obtained from Merck (Darmstadt, Germany). Formic acid (FA), sodium acetate, sodium cyanoborohydride and sodium chloride (NaCl) were purchased from Riedel-de Haven (Seelze, Germany). Protease inhibitors (Complete™ Mini) were purchased from Roche (Mannheim, Germany). Sodium dodecyl sulfate (SDS) and urea were purchased from Amresco (Solon, OH). Modified sequencing-grade trypsin for in-gel digestion was purchased from Promega (Madison, WI). Quantitative reagent for protein contents was purchased from Bio-Rad (Hercules, CA). Water was deionized to 18 MΩ by a Milli-Q system (Millipore, Bedford, MA).

2.3. Hydrophilic interaction chromatography (HILIC) for peptide separation

HILIC was performed on an L-7100 pump system with quaternary gradient capability (Hitachi, Tokyo, Japan) using a TSK gel Amide-80 HILIC column (2.0 × 150 mm, 3 μm; Tosoh Biosciences, Tokyo, Japan) [25–27] with a flow rate of 200 μL/min. Two buffers were used for gradient elution: solvent (A), 0.1% TFA in water, and solvent (B), 0.1% TFA in 100% ACN. The eluted fractions after being desalted from the in-house reverse-phase C18 column were each dissolved in 25 μL of solution containing 85% ACN and 0.1% TFA and then injected into the 20 μL sample loop. The gradient was processed as follows: 98% (B) for 5 min, 98–85% (B) for 5 min, 85–0% (B) for 40 min, 0% (B) for 5 min, 0–98% (B) for 2 min and 98% (B) for 3 min. A total of 10 fractions were collected (1.2 mL for each fraction) and dried in a vacuum centrifuge.

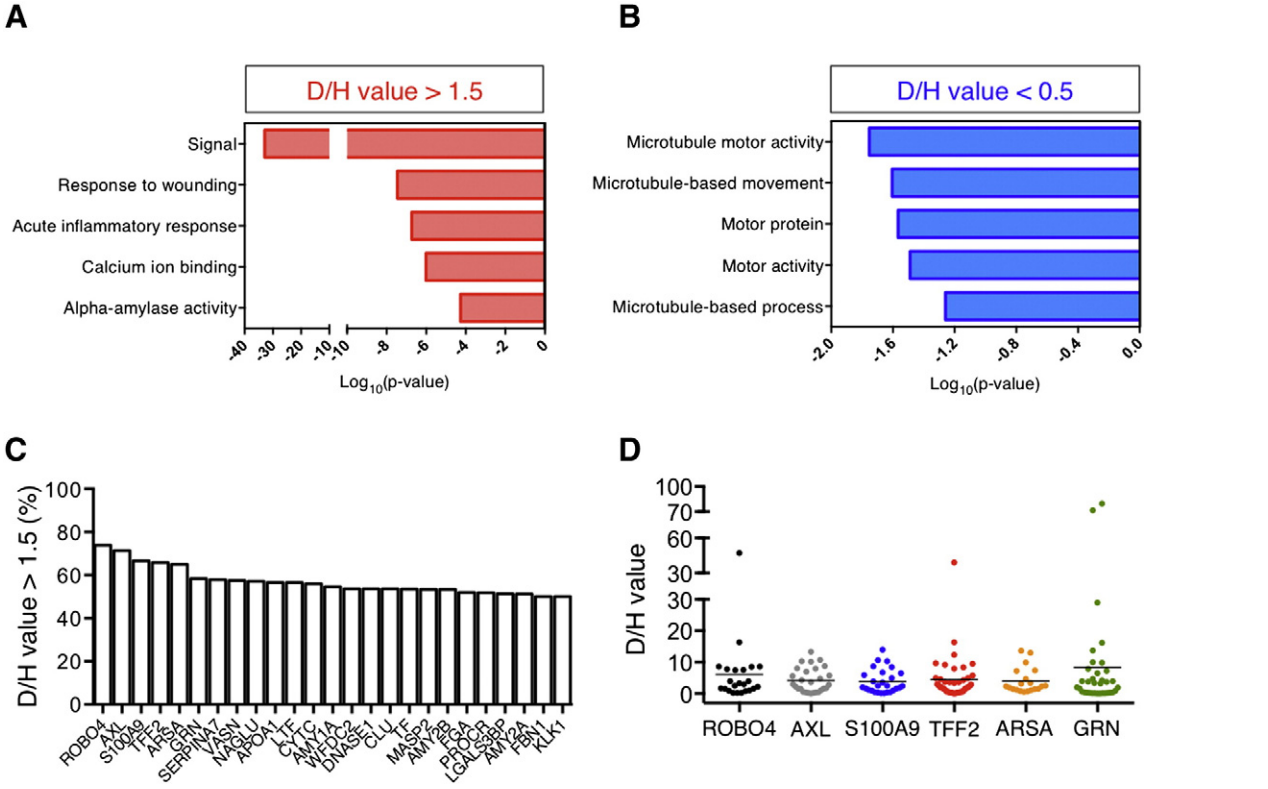

Fig. 3. Identification of top up-regulated proteins in urine samples from HCC patients (n = 44). (A, B) Gene ontology (GO) analysis was employed to classify major functional processes among (A) 82 up-regulated (D/H > 1.5, $p < 0.05$) and (B) 7 down-regulated (D/H < 0.5, $p < 0.05$) proteins. All identified GO term, p-value are shown in Table S2. (C) Identification of up-regulated proteins showing at least 50% penetrance. The y-axis is the frequency of up-regulated proteins (D/H > 1.5) identified in urine samples from HCC patients. (D) D/H values of top 6 candidates from individual sample pairs.

Table 1
The top 6 up-regulated proteins in urine samples from HCC patients. The value shown in the table is the quantification ratio (D/H) of peptides identified in urine from HCC patients (deuterium labeling) and normal controls (hydrogen labeling). The candidates were identified in at least 20 HCC urine sample pairs and the ratio of D/H > 1.5 were over 58%. p-Value was calculated by Wilcoxon signed-rank test with paired setting.

Protein names	Gene names	Uniprot	Mass (kDa)	D/H >1.5	Ratio (%)	p-Value
Roundabout homolog 4	ROBO4	Q8WZ75	107.457	17/23	73.9	0.0004
Tyrosine-protein kinase receptor UFO	AXL	P30530	98.336	20/28	71.4	0.0002
Protein S100-A9 (calgranulin-B)	S100A9	P06702	13.242	18/27	66.7	0.0004
Trefoil factor 2	TFF2	Q03403	14.284	27/41	65.9	0.00005
Arylsulfatase A	ARSA	P15289	53.588	13/20	65	0.0001
Granulins	GRN	P28799	63.544	21/36	58.3	0.001

2.4. NanoLC–MS/MS analysis

The lyophilized powders were reconstituted in 10 μL of 0.1% FA in H_2O and analyzed by LTQ Orbitrap XL (Thermo Fisher Scientific, San Jose, CA). Reverse-phase nanoLC separation was performed on an Agilent 1200 series nanoflow system (Agilent Technologies, Santa Clara, CA). A total of 10 μL sample from collected fractions was loaded onto an Agilent Zorbax XDB C18 precolumn (0.35 mm, 5 μm), followed by separation using in-house C18 column (i.d. 75 μm × 15 cm, 3 μm). The mobile phases used were (A) 0.1% FA in water and (B) 0.1% FA in 100% ACN. A linear gradient from 5% to 95% of (B) over a 70-min period at a flow rate of 300 nL/min was applied. The peptides were analyzed in the positive ion mode by applying a voltage of 1.8 kV to the injection needle. The MS was operated in a data-dependent mode, in which one full scan with m/z 400–1600 in the Orbitrap using a scan rate of 30 ms/scan. The fragmentation was performed using the CID mode with collision energy of 35 V. A repeat duration of 30 s was applied to exclude the same m/z ions from the reselection for fragmentation. The Xcalibur software (version 2.0.7, Thermo Fisher Scientific, San Jose, CA) was used for the management of instrument control, data acquisition, and data processing.

2.5. Protein database search, characterization and quantification

Peptides were identified by peak lists converted from the nanoLCMS/MS spectra by bioinformatics searching against *Homo sapiens* taxonomy in the Swiss-Prot databases for exact matches using the Mascot search program (http://www.matrixscience.com) [28,29]. Parameters were set as follows: a mass tolerance of 10 ppm for precursor ions and 0.8 Da for fragment ions; no missed cleavage site allowed for trypsin; carbamidomethyl cysteine as fixed modification; dimethylation specified as standard of the quantification; oxidized methionine and deamidated asparagine/glutamine as optional modification. Peptides were considered positively identified if their Mascot individual ion score was higher than 20 ($p < 0.05$). Subsequently, the analysis of peptide quantification ratio (D/H) for normal (hydrogen labeling) and HCC (deuterium labeling) from urine was carried out by Mascot Distiller program (version 2.3, Matrix Science Ltd., London, U.K.) using the average area of the first 3 isotopic peaks across the elution profile. The Mascot search data as well as quantification resulting from each fraction were also merged by this program that combined the peptide ratios matching the same sequence obtained from different fractions or at different retention times and charge states [22]. The identified proteins with up- and down-regulation were further

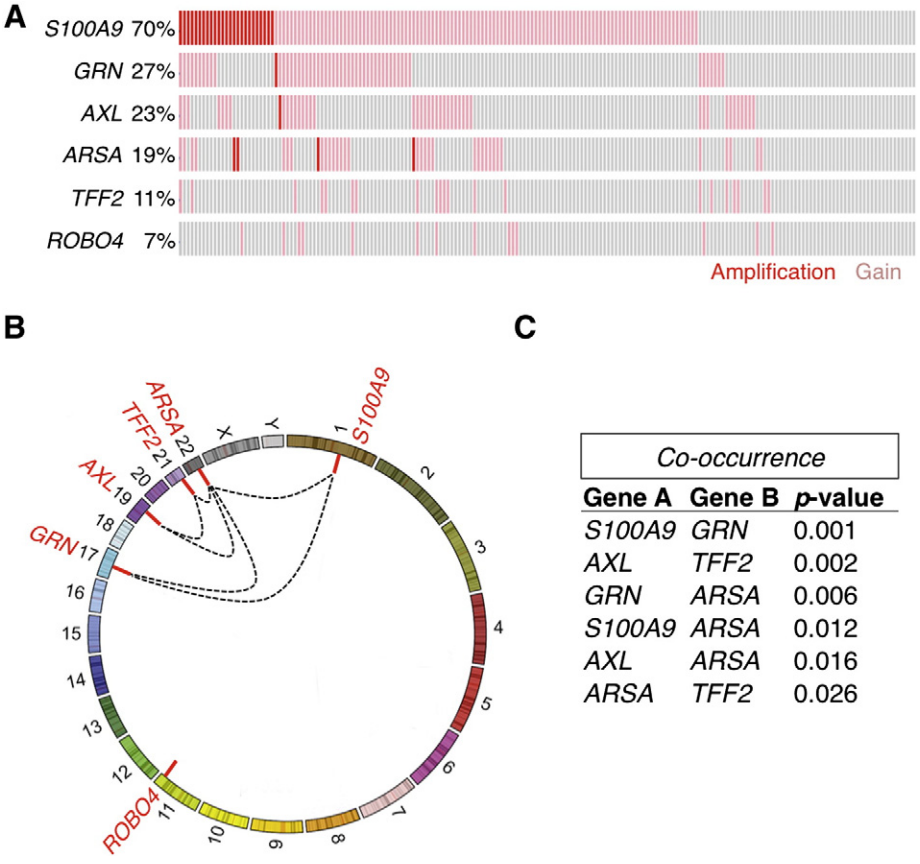

Fig. 4. Genomic analysis of identified biomarkers in HCC tumor samples (n = 193). (A) Putative copy-number alterations from copy number (GISTIC) algorithm for *S100A9*, *GRN*, *AXL*, *ARSA*, *TFF2* and *ROBO4* in each individual sample, with dark red indicating amplification and light red indicating gain. (B, C) Analysis of co-occurrence of candidate genes in HCC tumor samples. Fisher's exact test was used for statistical calculations. Data analysis is based on available TCGA data processed by the MSKCC cBio Core (www.cbioportal.org).

categorized based on their biological process and molecular function using the PANTHER classification system (http://www.pantherdb.org) as described in the previous studies [30–32].

2.6. Gene ontology (GO) and pathway analysis

By employing DAVID Bioinformatics Resources [33], GO analysis was performed to determine which functional processes were differentially represented in the protein list from quantitative proteomic analysis of urine samples from HCC patients and normal controls. 83 up-regulated and 8 down-regulated proteins from proteomic analysis were used for GO analysis.

2.7. Genomic and transcriptomic data analysis

All cancer genome datasets and bioinformatics tools for visualizing different parameters for analysis of genomic and transcriptomic data are accessible through the MSKCC cBio Core homepage (www.cbioportal.org) [34,35]. Co-occurrence of gene amplifications in HCC was performed as described previously [36] by analyzing The Cancer Genome Atlas (TCGA) (193 tumor samples). In brief, statistically significant CNAs in HCCs were analyzed for frequency and co-occurrence in individual samples, and Fisher's exact test was used to calculate p-values for co-occurrence of identified candidate genes. Comparison of copy number aberrations and gene expression was also based on available TCGA datasets for HCC.

2.8. Survival analysis of human HCC patients

Survival data analysis of patients with gene amplifications is based on available TCGA data processed by the MSKCC cBio Core (www.cbioportal.org) [34,35]. Statistical analysis was performed as described previously [37].

3. Results and discussion

One-year survival rates for newly diagnosed hepatocellular carcinoma (HCC) are lower than 50% owing to its unresectability that results from late detection. Although numerous investigations on the abnormalities of HCC based on genetic, biochemical and pathological characterization have been conducted, non-invasive and sensitive diagnostic or prognostic markers are very limited. In this regard, quantitative proteomics is an ideal tool to identify potential biomarkers in urine as it can systematically monitor protein variation on a large scale. In the current study, we employ an optimized shotgun-based quantitative proteomic strategy for analyzing the urinary proteins from 44 pairs of HCC patients and normal controls (Fig. 1A). Further cross-analyzing these proteomic results together with genomic and transcriptomic

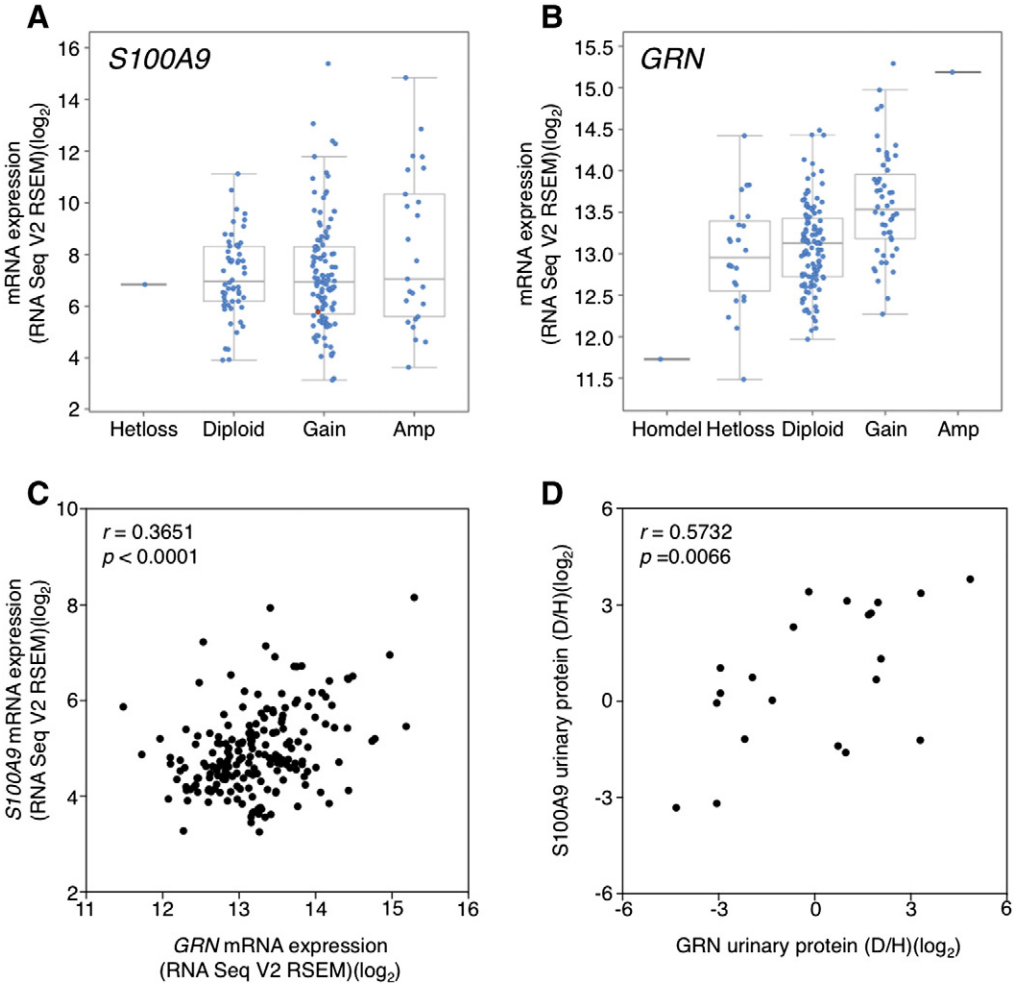

Fig. 5. Co-expression of S100A9 and GRN in tumor and urine samples from HCC patients. (A–C) Transcriptomics analysis of identified biomarkers in HCC tumor samples (n = 193). Gene expression and copy number (GISTIC) algorithm for (A) *S100A9* and (B) *GRN* indicates that amplification and gain on genomic levels correspond to their gene expression. Data analysis is based on available TCGA data processed by the MSKCC cBio Core (www.cbioportal.org). (C) Scatter plot illustrating the correlation between *S100A9* and *GRN* expression levels in HCC tumor samples (n = 193). (D) Scatter plot illustrating the correlation between S100A9 and GRN protein levels in urine samples from HCC patients (n = 44).

data was used to identify potential urinary HCC biomarkers, alone or their combinations (Fig. 1A).

3.1. Quantitative proteomic analysis of differentially expressed proteins in urine samples from HCC patients by a shotgun approach

By means of gel-free shotgun proteomic analysis coupled with stable isotope dimethyl labeling [8,12,22] and nanoLC–MS/MS [22–24], we have characterized and compared the urinary proteins between diseased and control groups (n = 44) (Fig. 1B). The process for identifying differentially released proteins including dimethyl labeling, enzyme digestion and peptide mass fingerprinting (PMF), followed by the analysis of urinary-peptide quantification ratio (D/H) for HCC patients (deuterium labeling) and normal controls (hydrogen labeling) (Fig. 1B) [22,23,38]. In total, we have systematically mapped around 1000 proteins among the 44 pairs of urine samples (Table S1). As shown in Fig. 2 and Table S1, expression levels of 91 identified proteins were significantly different ($p < 0.05$), which included 83 up-regulated (red color) and 8 down-regulated (green color) proteins. In addition, by clustering the D/H ratio, we found that there are two subgroups among all the 44 patients, i.e. 34 patients with many up-regulated proteins while 10 patients with more distinct expression profiles, suggesting that early diagnosis using multiple biomarkers may give better prediction than single one.

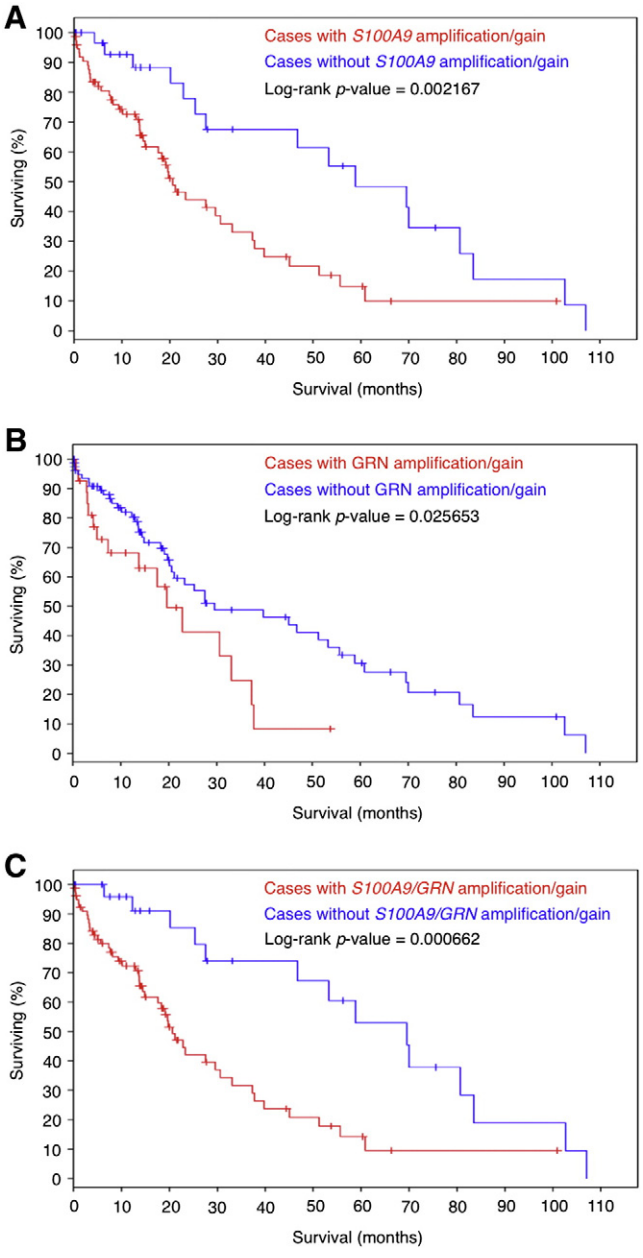

Fig. 6. Survival association of gene amplification and gain of *S100A9*, *GRN* and their combination in HCC patients (n = 193). (A) Survival curves of HCC patients comparing cases with and without *S100A9* amplification/gain. (B) Survival curves of HCC patients comparing cases with and without *GRN* amplification/gain. (C) Survival curves of HCC patients comparing cases with and without both *S100A9* and *GRN* amplification/gain. Statistical tests were performed as described previously [31]. Data analysis is based on available TCGA data processed by the MSKCC cBio Core (www.cbioportal.org).

The quantitative proteomic result was further subjected to gene ontology (GO) analysis to pinpoint proteins and processes that might underlie the change of urinary proteome in HCC patients (Fig. 3A, B and Table S2). Interestingly, proteins overexpressed ($D/H > 1.5$, $p < 0.05$) in HCC patients compared with healthy controls were linked to "signal," "response to wounding," "acute inflammatory response," "calcium ion binding," and "alpha-amylase activity" (Fig. 3A and Table S2). This analysis also revealed a significant correlation between down-regulated proteins ($D/H < 0.5$, $p < 0.05$) and microtubule motor activity (Fig. 3B and Table S2). The roles of signal peptides, cytokines and secreted proteins have been extensively studied due to their importance on the tumor microenvironment of HCC [39]. Therefore, the peptides in urine may reflect the signals involved in cancer cell-host interaction and inflammatory process [40,41]. The wound repair process was also shown to be essential for the growth of cancer stem cells in solid tumors [42]. Interestingly, calcium ion binding proteins were previously identified in another proteomic study that compared HCV-induced HCC with cirrhosis [43].

For a "hit" to undergo further analysis we used the scoring criterion of at least 50% penetrance and D/H value is more than 1.5; these included 25 proteins (Fig. 3C) in total and the top six proteins are ROBO4, AXL, S100A9, TFF2, ARSA, and GRN (Fig. 3D and Table 1). ROBO4 expression has been found to be confined to vasculature and it was overexpressed in tumor endothelial cells in comparison to normal adult endothelial cells [44]. Also, several candidates were related to HCC proliferation and invasion. AXL promotes tumor invasion through the transcriptional activation of Slug in HCC, and its inhibition was sufficient to suppress Slug expression and decreased the invasiveness of HCC cell lines [45]. S100A9 can activate the MAPK signaling pathway to also promote the proliferation and invasion of HCC cells [46]. Granulin (GRN)–epithelin precursor (GEP) is shown to be a potential therapeutic target owing to its overexpression promotes growth and invasion of HCC [47,48]. Taken together, the comparative proteome data from urine samples not only systematically mapped up- and down-regulated proteins that reflect the dysfunction of liver cells but also identified potential and valuable candidates for further investigations.

3.2. Genomic amplification of S100A9 and GRN co-occurs in tumors from HCC patients

Recent advances in proteogenomics which integrates MS-generated data with genomic and transcriptomic information are able to link DNA, RNA and protein data together to improve our understanding of biology in various fields [13]. This integrated approach has been also applied to study cancer biology as it can systematically identify tumor specific-peptides and monitor variation on a large scale of cancer cells' genomic, transcriptional and translational landscape. Here, by employing both onco-genomic and transcriptomic analyses, we aim to identify reliable biomarkers from our candidates as revealed in the proteomics result for HCC patients (Fig. 1A).

To better define the potential of identified hits, we determined the extent of their gene amplifications from cancer genome datasets at cBioPortal at the Memorial Sloan-Kettering Cancer Center derived from The Cancer Genome Atlas (TCGA) project, totaling 193 primary tumor samples of HCC (Fig. 4A). According to these data, these top six

hits harbor varied characteristics at genomic levels, for example, *S100A9* (70%), *GRN* (27%), and *AXL* (23%) are amplified in over 20% HCC tumor samples while amplification of ROBO4 is only 7% (Fig. 4A). The high complexity and heterogeneity of HCCs of both etiological and genetic aspects result in some difficulties for finding appropriate diagnosis and therapy [49]. Chromosome gains were very common in HCC, preferentially affecting chromosomes 1, 8, 17 and others [50,51]. The amplifications of *S100A9* and *GRN* with their detected locations correspond to chromosomes 1 and 17, respectively.

To further understand the association between these genomic events, Fisher's exact test was used to calculate the relevance of their co-occurrence (Fig. 4B and C). Interestingly, 5 out of 6 candidate genes were significantly co-amplified with each other, such as S100A9/GRN ($p = 0.001$), and AXL/TFF2 ($p = 0.002$). These results, which establish the amplification of S100A9 and GRN genes, may serve as co-indicators for detecting HCC tumors.

3.3. Co-expression of S100A9 and GRN in tumor and urine samples from HCC patients

Previous clinical and histopathological evidences suggest that HCC is a heterogeneous disease. Although genomic approaches to the analysis of HCC hold promise for identifying disease-related markers, the genomic aberrations, such as amplification and deletion, cannot usually reflect the change of mRNA or protein expression. Therefore, an alternate approach is to study tumors at the level of their gene expression profiles [17]. Also, combing data of the alterations on the genomic, transcriptional and translational levels may help us identify the most valuable markers for disease diagnosis.

By analyzing gene expression together with copy number (GISTIC) algorithm, copy number gain and amplification were correlated with *S100A9* and *GRN* mRNA overexpression in primary HCC tumors, as shown in Fig. 5A and B, respectively. Since these 2 genes were identified to be co-amplified on genomic levels, we next asked whether their gene expression also co-occurs on mRNA levels. We found that there is a significant correlation between the mRNA expression levels of *S100A9* and *GRN* ($r = 0.3651$, $p < 0.0001$) (Fig. 5C), suggesting that they may cooperate with each other during the hepatocarcinogenesis. To further validate this finding on protein levels, we analyzed the D/H ratio of *S100A9* and *GRN* among 44 pairs of HCC patients and normal controls, the urinary protein level of S100A9 was found highly correlated with GRN ($r = 0.5732$, $p = 0.0066$). Thus, these results establish that the S100A9 and GRN may serve as combinatorial biomarkers on genomic, transcriptional, and protein levels and can be co-detected in both tumor and urine samples.

3.4. Amplifications of S100A9/GRN and their combination are associated with poor survival of HCC patients

To substantiate our findings in survival benefit of patients, we next asked whether amplifications of *S100A9* and *GRN* are associated with survival outcome in human HCC. We analyzed genomic data from a cohort of 193 HCC patients with available survival data from TCGA [37,52]. We found that amplifications of either *S100A9* or *GRN* were significantly correlated with poor survival (Fig. 6A and B) and thus might predict more aggressive disease progression. More importantly, co-amplification of *S100A9* or *GRN* is associated with worse overall survival than that of individual ones, suggesting they can potentially serve as combinatorial biomarkers for early diagnosis of HCC.

Overexpression of S100A9 and GRN can promote both growth and invasion of HCC cells through different signaling pathways [46–48]. Previous studies have shown that HCC tumor invasion and metastasis predict poor survival in patients [53,54]. Therefore, the amplification and expression of *S100A9* and *GRN* may result in poor survival in patients with HCC thorough tumor metastasis in different organs. Together, this result indicates that amplification of these two individual genes or their combination can predict not only a high risk of extrahepatic metastasis but also a worse survival of HCC patients.

4. Conclusions

In this study, the utility of shotgun-based quantitative proteomics has been adapted and extended to compare the urinary proteins between diseased and control groups in order to identify potential biomarker candidates. This proteomic approach can be coupled with genomic and transcriptomic analysis to be applied as a powerful onco-proteogenomic tool for the detection of important protein markers involved in the tumorigenesis of HCC. Among these, S100A9 and GRN were shown to be co-amplified and co-expressed in tumor and urine samples from HCC patients with poor survival. Our results pinpoint that urinary S100A9 and GRN as potential combinatorial biomarkers for early detection of HCC and highlight the utility of proteogenomics for identifying protein markers that may be valuable to diagnosis and prognosis of diverse types of cancers and other diseases.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbacli.2015.02.004.

Transparency Document

The Transparency document associated with this article can be found, in the online version.

Acknowledgements

This work was supported in part by the Kaohsiung Medical University for the program of “Aim for the Top Universities Grant” (Grant No. KMU-TP103E15) and the National Science Council (NSC Grants 99-2745-B-037-005 and 102-2314-B-037-037-MY2 to S.H. Chiou and K.T. Lee), Taipei, Taiwan. We thank the core facility grant from the Quantitative Proteomics Center (NSC Grant 99-2745-B-037-005) at the Center for Research Resources and Development (CRRD), Kaohsiung Medical University under the auspices of National Science Council. The authors have declared that no conflict of interest exists.

References

[1] A. Villanueva, V. Hernandez-Gea, J.M. Llovet, Medical therapies for hepatocellular carcinoma: a critical view of the evidence, Nat. Rev. Gastroenterol. Hepatol. 10 (2013) 34–42.
[2] P.A. Farazi, R.A. DePinho, Hepatocellular carcinoma pathogenesis: from genes to environment, Nat. Rev. Cancer 6 (2006) 674–687.
[3] A.J. Sanyal, S.K. Yoon, R. Lencioni, The etiology of hepatocellular carcinoma and consequences for treatment, Oncologist 15 (Suppl. 4) (2010) 14–22.
[4] J.M. Llovet, S. Ricci, V. Mazzaferro, P. Hilgard, E. Gane, J.F. Blanc, A.C. de Oliveira, A. Santoro, J.L. Raoul, A. Forner, M. Schwartz, C. Porta, S. Zeuzem, L. Bolondi, T.F. Greten, P.R. Galle, J.F. Seitz, I. Borbath, D. Haussinger, T. Giannaris, M. Shan, M. Moscovici, D. Voliotis, J. Bruix, Sorafenib in advanced hepatocellular carcinoma, N. Engl. J. Med. 359 (2008) 378–390.
[5] C.H. Huang, A. Lujambio, J. Zuber, D.F. Tschaharganeh, M.G. Doran, M.J. Evans, T. Kitzing, N. Zhu, E. de Stanchina, C.L. Sawyers, S.A. Armstrong, J.S. Lewis, C.J. Sherr, S.W. Lowe, CDK9-mediated transcription elongation is required for MYC addiction in hepatocellular carcinoma, Genes Dev. 28 (2014) 1800–1814.
[6] L. Goyal, M.D. Muzumdar, A.X. Zhu, Targeting the HGF/c-MET pathway in hepatocellular carcinoma, Clin. Cancer Res. 19 (2013) 2310–2318.
[7] B. Zhang, B. Yang, Combined alpha fetoprotein testing and ultrasonography as a screening test for primary liver cancer, J. Med. Screen. 6 (1999) 108–110.
[8] S.H. Chiou, C.Y. Wu, Clinical proteomics: current status, challenges, and future perspectives, Kaohsiung J. Med. Sci. 27 (2011) 1–14.
[9] T. Ideker, V. Thorsson, J.A. Ranish, R. Christmas, J. Buhler, J.K. Eng, R. Bumgarner, D.R. Goodlett, R. Aebersold, L. Hood, Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, Science 292 (2001) 929–934.
[10] M.P. Washburn, D. Wolters, J.R. Yates 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, Nat. Biotechnol. 19 (2001) 242–247.
[11] A.L. McCormack, D.M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, J.R. Yates 3rd, Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level, Anal. Chem. 69 (1997) 767–776.
[12] M.J. MacCoss, W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark, J.R. Yates 3rd, Shotgun identification of

protein modifications from protein complexes and lens tissue, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 7900–7905.
[13] J.A. Alfaro, A. Sinha, T. Kislinger, P.C. Boutros, Onco-proteogenomics: cancer proteomics joins forces with genomics, Nat. Methods 11 (2014) 1107–1113.
[14] P. Laurent-Puig, P. Legoix, O. Bluteau, J. Belghiti, D. Franco, F. Binot, G. Monges, G. Thomas, P. Bioulac-Sage, J. Zucman-Rossi, Genetic alterations associated with hepatocellular carcinomas define distinct pathways of hepatocarcinogenesis, Gastroenterology 120 (2001) 1763–1773.
[15] H. Okamoto, K. Yasui, C. Zhao, S. Arii, J. Inazawa, PTK2 and EIF3S3 genes may be amplification targets at 8q23–q24 and are associated with large hepatocellular carcinomas, Hepatology 38 (2003) 1242–1249.
[16] E.T. Sawey, M. Chanrion, C. Cai, G. Wu, J. Zhang, L. Zender, A. Zhao, R.W. Busuttil, H. Yee, L. Stein, D.M. French, R.S. Finn, S.W. Lowe, S. Powers, Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by oncogenomic screening, Cancer Cell 19 (2011) 347–358.
[17] Y. Hoshida, S.M. Nijman, M. Kobayashi, J.A. Chan, J.P. Brunet, D.Y. Chiang, A. Villanueva, P. Newell, K. Ikeda, M. Hashimoto, G. Watanabe, S. Gabriel, S.L. Friedman, H. Kumada, J.M. Llovet, T.R. Golub, Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma, Cancer Res. 69 (2009) 7385–7392.
[18] R.C. Rivers, C. Kinsinger, E.S. Boja, T. Hiltke, M. Mesri, H. Rodriguez, Linking cancer genome to proteome: NCI's investment into proteogenomics, Proteomics 14 (2014) 2633–2636.
[19] X. Wang, R.J. Slebos, D. Wang, P.J. Halvey, D.L. Tabb, D.C. Liebler, B. Zhang, Protein identification using customized protein sequence databases derived from RNA-Seq data, J. Proteome Res. 11 (2012) 1009–1017.
[20] P.J. Halvey, X. Wang, J. Wang, A.A. Bhat, P. Dhawan, M. Li, B. Zhang, D.C. Liebler, R.J. Slebos, Proteogenomic analysis reveals unanticipated adaptations of colorectal tumor cells to deficiencies in DNA mismatch repair, Cancer Res. 74 (2014) 387–397.
[21] P.F. Aquino, D.B. Lima, J. de Saldanha da Gama Fischer, R.D. Melani, F.C. Nogueira, S.R. Chalub, E.R. Soares, V.C. Barbosa, G.B. Domont, P.C. Carvalho, Exploring the proteomic landscape of a gastric cancer biopsy with the shotgun imaging analyzer, J. Proteome Res. 13 (2014) 314–320.
[22] C.J. Wu, Y.W. Chen, J.H. Tai, S.H. Chen, Quantitative phosphoproteomics studies using stable isotope dimethyl labeling coupled with IMAC-HILIC-nanoLC–MS/MS for estrogen-induced transcriptional regulation, J. Proteome Res. 10 (2011) 1088–1097.
[23] H.M. Santos, P. Kouvonen, J.L. Capelo, G.L. Corthals, Isotopic labelling of peptides in tissues enhances mass spectrometric profiling, Rapid Commun. Mass Spectrom. 26 (2012) 254–262.
[24] P.J. Boersema, T.T. Aye, T.A. van Veen, A.J. Heck, S. Mohammed, Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates, Proteomics 8 (2008) 4624–4632.
[25] T. Yoshida, Peptide separation in normal phase liquid chromatography, Anal. Chem. 69 (1997) 3038–3043.
[26] T. Yoshida, Calculation of peptide retention coefficients in normal-phase liquid chromatography, J. Chromatogr. A 808 (1998) 105–112.
[27] A.J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, J. Chromatogr. 499 (1990) 177–196.
[28] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis 20 (1999) 3551–3567.
[29] M. Hirosawa, M. Hoshida, M. Ishikawa, T. Toya, MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming, Comput. Appl. Biosci. 9 (1993) 161–167.
[30] H. Mi, A. Muruganujan, P.D. Thomas, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees, Nucleic Acids Res. 41 (2013) D377–D386.
[31] Y. Nikolsky, J. Bryant, Protein networks and pathway analysis. Preface, Methods Mol. Biol. 563 (2009) v–vii.
[32] P.D. Thomas, A. Kejariwal, N. Guo, H. Mi, M.J. Campbell, A. Muruganujan, B. Lazareva-Ulitsky, Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools, Nucleic Acids Res. 34 (2006) W645–W650.
[33] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[34] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (2012) 401–404.
[35] J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (2013) l1.
[36] B.S. Taylor, N. Schultz, H. Hieronymus, A. Gopalan, Y. Xiao, B.S. Carver, V.K. Arora, P. Kaushik, E. Cerami, B. Reva, Y. Antipin, N. Mitsiades, T. Landers, I. Dolgalev, J.E. Major, M. Wilson, N.D. Socci, A.E. Lash, A. Heguy, J.A. Eastham, H.I. Scher, V.E. Reuter, P.T. Scardino, C. Sander, C.L. Sawyers, W.L. Gerald, Integrative genomic profiling of human prostate cancer, Cancer Cell 18 (2010) 11–22.
[37] S. Roessler, H.L. Jia, A. Budhu, M. Forgues, Q.H. Ye, J.S. Lee, S.S. Thorgeirsson, Z. Sun, Z.Y. Tang, L.X. Qin, X.W. Wang, A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients, Cancer Res. 70 (2010) 10202–10212.
[38] J. Webster, D. Oxley, Protein identification by MALDI-TOF mass spectrometry, Methods Mol. Biol. 800 (2012) 227–240.
[39] G.C. Leonardi, S. Candido, M. Cervello, D. Nicolosi, F. Raiti, S. Travali, D.A. Spandidos, M. Libra, The tumor microenvironment in hepatocellular carcinoma (review), Int. J. Oncol. 40 (2012) 1733–1747.
[40] A. Budhu, X.W. Wang, The role of cytokines in hepatocellular carcinoma, J. Leukoc. Biol. 80 (2006) 1197–1213.
[41] V. Hernandez-Gea, S. Toffanin, S.L. Friedman, J.M. Llovet, Role of the microenvironment in the pathogenesis and treatment of hepatocellular carcinoma, Gastroenterology 144 (2013) 512–527.
[42] A.V. Kurtova, J. Xiao, Q. Mo, S. Pazhanisamy, R. Krasnow, S.P. Lerner, F. Chen, T.T. Roh, E. Lay, P.L. Ho, K.S. Chan, Blocking PGE-induced tumour repopulation abrogates bladder cancer chemoresistance, Nature 517 (2015) 209–213.
[43] V.R. Mas, D.G. Maluf, K.J. Archer, K. Yanek, K. Bornstein, R.A. Fisher, Proteomic analysis of HCV cirrhosis and HCV-induced HCC: identifying biomarkers for monitoring HCV-cirrhotic patients awaiting liver transplantation, Transplantation 87 (2009) 143–152.
[44] M.E. Avci, O. Konu, T. Yagci, Quantification of SLIT-ROBO transcripts in hepatocellular carcinoma reveals two groups of genes with coordinate expression, BMC Cancer 8 (2008) 392.
[45] H.J. Lee, Y.M. Jeng, Y.L. Chen, L. Chung, R.H. Yuan, Gas6/Axl pathway promotes tumor invasion through the transcriptional activation of Slug in hepatocellular carcinoma, Carcinogenesis 35 (2014) 769–775.
[46] R. Wu, L. Duan, L. Ye, H. Wang, X. Yang, Y. Zhang, X. Chen, Y. Weng, J. Luo, M. Tang, Q. Shi, T. He, L. Zhou, S100A9 promotes the proliferation and invasion of HepG2 hepatocellular carcinoma cells via the activation of the MAPK signaling pathway, Int. J. Oncol. 42 (2013) 1001–1010.
[47] J.C. Ho, Y.C. Ip, S.T. Cheung, Y.T. Lee, K.F. Chan, S.Y. Wong, S.T. Fan, Granulin–epithelin precursor as a therapeutic target for hepatocellular carcinoma, Hepatology 47 (2008) 1524–1532.
[48] S.T. Cheung, S.Y. Wong, K.L. Leung, X. Chen, S. So, I.O. Ng, S.T. Fan, Granulin–epithelin precursor overexpression promotes growth and invasion of hepatocellular carcinoma, Clin. Cancer Res. 10 (2004) 7629–7636.
[49] A. Fujimoto, Y. Totoki, T. Abe, K.A. Boroevich, F. Hosoda, H.H. Nguyen, M. Aoki, N. Hosono, M. Kubo, F. Miya, Y. Arai, H. Takahashi, T. Shirakihara, M. Nagasaki, T. Shibuya, K. Nakano, K. Watanabe-Makino, H. Tanaka, H. Nakamura, J. Kusuda, H. Ojima, K. Shimada, T. Okusaka, M. Ueno, Y. Shigekawa, Y. Kawakami, K. Arihiro, H. Ohdan, K. Gotoh, O. Ishikawa, S. Ariizumi, M. Yamamoto, T. Yamada, K. Chayama, T. Kosuge, H. Yamaue, N. Kamatani, S. Miyano, H. Nakagama, Y. Nakamura, T. Tsunoda, T. Shibata, H. Nakagawa, Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators, Nat. Genet. 44 (2012) 760–764.
[50] U. Zimmermann, D. Feneux, G. Mathey, F. Gayral, D. Franco, P. Bedossa, Chromosomal aberrations in hepatocellular carcinomas: relationship with pathological features, Hepatology 26 (1997) 1492–1498.
[51] H. Yano, A. Nanashima, S. Hidaka, M. Haseba, K. Tanaka, H. Yamaguchi, T. Nakagoe, Y. Tagawa, T. Nagayasu, Numerical aberrations of chromosome 17 and the p53 locus in small hepatocellular carcinomas, Anticancer Res. 24 (2004) 111–115.
[52] H.L. Jia, Q.H. Ye, L.X. Qin, A. Budhu, M. Forgues, Y. Chen, Y.K. Liu, H.C. Sun, L. Wang, H.Z. Lu, F. Shen, Z.Y. Tang, X.W. Wang, Gene expression profiling reveals potential biomarkers of human hepatocellular carcinoma, Clin. Cancer Res. 13 (2007) 1133–1139.
[53] K. Uchino, R. Tateishi, S. Shiina, M. Kanda, R. Masuzaki, Y. Kondo, T. Goto, M. Omata, H. Yoshida, K. Koike, Hepatocellular carcinoma with extrahepatic metastasis: clinical features and prognostic factors, Cancer 117 (2011) 4475–4483.
[54] B.S. Ko, T.C. Chang, C. Hsu, Y.C. Chen, T.L. Shen, S.C. Chen, J. Wang, K.K. Wu, Y.J. Jan, J.Y. Liou, Overexpression of 14-3-3 epsilon predicts tumour metastasis and poor survival in hepatocellular carcinoma, Histopathology 58 (2011) 705–711.