

# Phospho-Network Analysis Identifies and Quantifies Hepatitis C Virus (HCV)-induced Hepatocellular Carcinoma (HCC) Proteins Regulating Viral-mediated Tumor Growth

NU T. LU<sup>1,2</sup>, NATALIE M. LIU<sup>1</sup>, JAMES Q. VU<sup>1</sup>, DARSHIL PATEL<sup>1</sup>, WHITAKER COHN<sup>3</sup>,  
JOE CAPRI<sup>3,4</sup>, MARY ZIEGLER<sup>1</sup>, NIKITA PATEL<sup>1</sup>, ANGELA TRAMONTANO<sup>5</sup>,  
ROGER WILLIAMS<sup>6</sup>, JULIAN WHITELEGGE<sup>3</sup> and SAMUEL W. FRENCH<sup>1</sup>

<sup>1</sup>Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA, U.S.A;

<sup>2</sup>Department of Hematology and Oncology, University of California Los Angeles, Los Angeles, CA, U.S.A;

<sup>3</sup>Department of Psychiatry and Biobehavioral Sciences,  
University of California Los Angeles, Los Angeles, CA, U.S.A;

<sup>4</sup>Department of Molecular and Medical Pharmacology,  
University of California Los Angeles, Los Angeles, CA, U.S.A;

<sup>5</sup>Institute for Technology Assessment, Massachusetts General Hospital, Boston, MA, U.S.A;

<sup>6</sup>The Foundation for Liver Research-The Institute for Hepatology, London, U.K.

**Abstract.** *Background: Patients with chronic hepatitis C virus (HCV) infection are at risk of serious complications of cirrhosis and hepatocellular carcinoma (HCC). Mass spectrometry (MS) is a versatile methodology that produces a global proteomic landscape for analysis of cancer mechanisms. Materials and Methods: Using multiplex peptide stable isotopic labeling and immobilized metal affinity chromatography (IMAC), we enriched and quantified the phosphoproteome of HCC, with and without HCV. While raw data identified protein targets based on expression alone, we also used abundance groups for comprehensive functional analysis. Results: Analysis of functional differences highlighted deregulated phosphoprotein networks. This uncovered additional candidates that could be directly derived from the MS data. Cellular processes and pathways that may differ with HCV infection include: cytoskeletal dynamics, insulin response, gene expression, and PI3K/AKT oncogenesis. Conclusion: This function-focused workflow provides a simple framework to analyze MS data.*

*Phosphoproteome quantitation with inclusive functional analysis can generate hypotheses for liver cancer research to improve early screening and identification of molecular targets for therapy.*

The phospho-network promises us more than a few scientific breakthroughs. In unveiling how *Escherichia coli* enhances its bacterial virulence (1, 2), how oncogenic B-RAF stabilizes the anti-apoptotic protein Mcl-1 to promote melanoma survival and chemo-resistance (3), or how type II diabetics develop insulin insensitivity (4), the phospho-network implores our imagination to draw other intricate biological maps.

Patients with chronic hepatitis C virus (HCV) infection risk serious complications of cirrhosis and hepatocellular carcinoma (HCC). What if HCV-infected hepatocytes utilize post-translational modifications (PTM) like phosphorylation to orchestrate fundamental biological changes influencing cancer growth, proliferation, and differentiation? Using selective tools (5, 6) to capture network-wide phosphorylation signals to expose the primary switches of how HCV triggers core HCC pathways, we can create a new platform that generates hypotheses for future studies to earlier diagnose and treat one of the most prevalent types of liver cancer (7, 8).

This study presents the first quantitative phosphoproteome analysis of liver cancer with HCV. Our model uses a simple work-flow to map the differential phosphoproteomic expression between two conditions: HCC (hepatocellular carcinoma) and HCC+V (hepatocellular carcinoma with HCV replicon). While raw data were utilized to identify

*Correspondence to:* Samuel Wheeler French, MD, Ph.D., Department of Pathology and Laboratory Medicine, University of California Los Angeles, UCLA Center for Health Sciences, 10833 Le Conte Ave., Rm 13-145, Los Angeles, CA 90095-1732, U.S.A. Tel: +1 3102672795, Fax: +1 3102672058, e-mail: sfrench@mednet.ucla.edu

**Key Words:** Hepatocellular carcinoma, hepatitis C virus, mass spectrometry, phosphoproteomics.

protein candidates based on expression alone, additional comprehensive functional analyses using abundance thresholds uncovered targets that could not be derived from the original MS results. This function-focused workflow provides a simple methodology for revealing hypotheses that incorporate a broader range of protein networks. In comparison to analyzing mRNA messages, mass spectrometry provides a much closer approximation of the final cellular dynamics that influence HCC transformation. The HCC and HCC+V phosphoproteomes depict unique protein abundances that may differentially regulate processes including: metabolic insulin response, cytoskeletal dynamics impacting cell growth, viral-mediated host mRNA transcription, and phosphatidylinositol-3-kinase/protein kinase B (PI3K/AKT)-driven oncogenic survival. These MS derived hypotheses can serve as the starting point for subsequent molecular research to closely investigate the mechanisms of HCC.

## Materials and Methods

**In solution tryptic digestion.** All experimentation and analysis were performed at the University of California, Los Angeles.  $2 \times 10^8$  cells were harvested with 0.25% trypsin (Fisher Scientific, Grand Island, NY, USA) from Huh-7.5 (HCC) and SGR (HCC+V) cell lines. Cell pellets were lysed in 500  $\mu$ L of 12 mM sodium lauryl sarcosine, 0.5% sodium deoxycholate, and 50 mM triethyl ammonium bicarbonate (TEAB). Samples were then sonicated and heated at 90°C for 5 min each. A bicinchoninic acid protein quantitation assay (Thermo Fischer Scientific) was performed using a spectrophotometer. Protein disulfides were reduced with 5 mM tris (2-carboxyethyl) phosphine (TCEP) in 50 mM TEAB (30 min at RT) and alkylated with 10 mM iodoacetamide in 50 mM TEAB (30 min in the dark at RT). Protein solutions were diluted five-fold with 50 mM TEAB. Lyophilized porcine trypsin (Promega, Madison, WI, USA) was solubilized in 50 mM TEAB and added 1:50 (w/w) ratio to proteins, followed by overnight incubation at 37°C. Sodium deoxycholate was removed from peptide solutions with trifluoroacetic acid (0.5% final concentration), phase transferred with ethyl acetate 1:1(v/v), and centrifuged ( $12,000 \times g$  at RT, 5 min). The upper organic phase was decanted prior to lyophilizing peptides.

**On-column stable isotopic dimethyl labeling.** Supelco Visiprep SPE Vacuum Manifold (Sigma-Aldrich, St. Louis, MO, USA) was assembled with Sep-Pak cartridges and waste containers. Each sample was reconstituted in 2 mL of water. Cartridges were conditioned with 100% methanol (MeOH), 80% acetonitrile (ACN), and 0.5% acetic acid. The cartridges were then equilibrated with 0.6% acetic acid. The sample set was passed through the cartridge and washed with 2.4 mL of 0.25 M 2-(N-Morpholino)-ethanesulfonic acid (MES). Two labeling solutions were then prepared by adding 300  $\mu$ L of formaldehyde (4%,  $\text{CH}_2\text{O}$  or  $^{13}\text{CD}_2\text{O}$ ) and 300  $\mu$ L cyanoborohydride (0.6M,  $\text{NaBH}_3\text{CN}$  or  $\text{NaBD}_3\text{CN}$ ) to 2.4 mL of 0.25 M MES Buffer (final concentration: 0.4%  $\text{CH}_2\text{O}$  or  $^{13}\text{CD}_2\text{O}$  or 60 mM  $\text{NaBH}_3\text{CN}$  or  $\text{NaBD}_3\text{CN}$ ). 3 mL of 0.4%  $\text{CH}_2\text{O}$ , 60mM  $\text{NaBH}_3\text{CN}$  (light isotope labeling solution) was passed through the HCC samples over the course of 10 min. Then, 3 mL of 0.4%  $^{13}\text{CD}_2\text{O}$ , 60 mM  $\text{NaBD}_3\text{CN}$  (heavy isotope labeling

solution) was passed through the HCC + V samples over the course of 10 min. Acetic acid (0.6%) was then used to wash the cartridges. Labeled peptides were eluted with 3 mL of 80% acetonitrile (ACN) and 0.5% acetic acid (9).

**HILIC.** Labeled peptides were combined 1:1 (w/w) and loaded onto a column (TosoHaas TSKgel Amide-80 HR 4.6 mm ID  $\times$  25 cm, 5  $\mu$ m) and fractionated using a Series II 1090 Liquid Chromatography (Hewlett-Packard, Palo Alto, CA, USA). Loading Solvent A (100%  $\text{H}_2\text{O}$ , 0.1% trifluoroacetic acid (TFA)) and Eluting Solvent B (100% ACN, 0.1% TFA) were used. Fractionation was performed as per protocol (5). Peptides were initially held in 80% solvent B for 10 min, eluted for 35 min in 80%-65% solvent B, and then eluted for 5 min in 65%-0% solvent B. After elution, there was a 10 min hold for a final run time of 60 min. A total of 40 fractions (1 min each) were collected.

**IMAC.** Forty fractions were combined into 28 fractions for IMAC phosphopeptide enrichment as per protocol (5). 18  $\mu$ L of PHOS-Select Iron Affinity Gel beads (Sigma) were added to each sample and mixed (30 min. at RT). An Acroprep Nylon 96 Filter Plate (Pall) was prepared by passing through 100% ethanol, 400 mM ammonium hydroxide in 85% ACN, and 250 mM acetic acid in 95% ACN. After samples were passed through, they were washed with 250 mM acetic acid in 30% ACN. Acetic acid was diluted by passing through  $\text{H}_2\text{O}$ . Samples were eluted with 150  $\mu$ L of 400 mM ammonium hydroxide in 30% ACN and 15  $\mu$ L of 10% TFA using centrifuge ( $1150 \times g \times 2$  min). Samples were dried prior to desalting.

**Desalting.** Desalting of samples was performed according to StageTips procedure (10). Phosphopeptide-enriched samples were re-suspended in 100  $\mu$ L of a buffer solution (3% ACN and 0.5% acetic acid). C18 filter tips were conditioned by passing through 100% methanol and then 80% ACN with 0.5% acetic acid. Next, C18 tips were equilibrated with 3% ACN, 0.5% acetic acid and 80  $\mu$ L of the samples were subsequently loaded. The filter was washed with 3% ACN, 0.5% acetic acid. Peptides were eluted with 20  $\mu$ L of 80% ACN, 0.5% acetic acid. The eluted samples were lyophilized and 10  $\mu$ L of 3% ACN, 0.1% FA was mixed with the lyophilized sample, which was then placed into the injection vial.

**HPLC-MS/MS.** Fractionated samples were analyzed with an Eksigent 2D nanoLC attached to a Orbitrap LTQ XL (Thermo Fisher Scientific, Waltham, MA, USA) Peptides were injected onto a laser-pulled nanobore 20 cm  $\times$  75  $\mu$ m C18 column (Acutech Scientific, San Diego, CA, USA) in buffer A (0.1% formic acid) and resolved using a 2-h linear gradient from 3-50% buffer B (80% acetonitrile with 0.1% formic acid). The Orbitrap LTQ XL was operated in data-dependent mode with 60,000 resolution and target auto gain control at  $5 \times 10^6$  for parent scan. The top 12 ions above +1 charge were subjected to collision induced dissociation (CID) set to a value of 35 with target auto gain control of 5000. Dynamic exclusion was set to 30 seconds. Raw data was processed and quantified using Proteome Discoverer (Thermo Fisher Scientific, Waltham, MA, USA).

**MS Quantification.** Multiplex peptide stable isotope dimethyl labeling was used to accurately compare and quantify protein levels. This method converts all primary amines in a peptide mixture to

dimethylamines. By incorporating  $^{13}\text{C}$  and D as the heavy isotopes in one condition, all experimental conditions (*i.e.* HCC and HCC+V) are subsequently distinguished during concurrent LC-MS/MS runs. Protein quantification is achieved by comparing their relative ion abundances. Raw data was interpreted with Proteome Discoverer v.2.0 (Thermo Fisher Scientific, Waltham, MA, USA) and SEQUEST HT (Matrix Science London, UK) against human entries from the UniProt database (released December 20, 2015). Search parameters were performed using: trypsin enzyme digest with up to two missed cleavages; a precursor ion mass tolerance of 10 ppm; CID fragment ion mass tolerance of 0.02 Da; carbamidomethylation as a fixed modification; variable modifications for oxidation of methionine; dimethylation ( $\text{C}_2\text{H}_6$ ) of lysine and the N-terminus; dimethylation  $^{13}\text{C}_2\text{D}_6$  of lysine and the N-terminus; and phosphorylation of serine, threonine, and tyrosine; and a peptide minimum length of six amino acids. An individual ion score cutoff was chosen that led to a FDR of 5% at the PSM level. Phosphorylation sites were localized with PhosphoRS 2.0 embedded within Proteome Discoverer (6). Individual site probabilities are calculated using probability-based PhosphoRS 2.0 with a 75% site probability threshold.

**Data analysis and software.** Microsoft Office Excel (Microsoft, Redmond, WA, USA) was used for main data organization and presentation. R statistical computing software was used for creating intensity heat maps. WebGestalt (bioinfo.vanderbilt.edu/webgestalt) and Gene Ontology Consortium (geneontology.org) were used for GO enrichment and PathwayCommons analysis to generate protein sets. Protein networks and corresponding pathways were generated from PathwayCommons analysis using the Top10 significance level. All GO enrichment was performed with a significance level of  $p < 0.05$  and for each analysis a hypergeometric statistical method and Benjamini-Hochberg multiple test adjustment were used. Functional proteomic analysis was completed by combining GO enrichment results with UniProt (uniprot.org) search and validation. The heat map depicting proteomic functional results was created by using Python (python.org) software to sort GO enrichment data into different abundance categories and graph the resulting proportion of IDs over the respective total. Protein network diagrams were created using Cytoscape3.3.0 (cytoscape.org) with the GeneMania application (apps.cytoscape.org/apps/genemania) (11). Pathways for the hypothetical protein target figures were created using information from PathwayCommons 2 (v7) (pathwaycommons.org) and UniProt. The hypothetical protein target and mass spectrometry workflow images were made with GeneMania software (<http://www.genemania.org/>) and Servier Medical Art (servier.com/Powerpoint-image-bank).

**Western blot analysis.** Lysates were prepared with TritonX-100 lysis buffer (0.5% Triton X-100, 10 mM Tris, 150 mM NaCl, 5 mM EDTA, 1 mM Sodium Orthovanadate, 40 mM Sodium Fluoride, 50 mM Beta-Glycophosphosphate, 5% Sigma Aldrich Protease Inhibitor Cocktail, and 10% Glycerol). Protein samples were separated and transferred overnight onto PVDF membrane (EMD Millipore, Darmstadt, Germany). Western blots were incubated overnight with primary antibodies, then for 1 h with secondary antibodies. Primary antibodies for BCL-2, p-Op18 Ser25, p-Op18 Ser63 were purchased from Santa Cruz Biotechnology (Dallas, TX, USA); VEGF was purchased from Abcam (Cambridge, MA, USA); Tubulin was purchased from DSHB (Iowa City, IA, USA); and all other remaining antibodies were purchased from Cell Signaling (Danvers, MA, USA). Secondary

Table I. *Phosphoproteome overview.*

Category	No.
Peptides	3712
Phosphopeptides	2382
Abundance Range (HCC+V)/(HCC)	(7.14-0.0285)
Total Phosphosites	3173
P-Serine	2698
P-Threonine	426
P-Tyrosine	49

antibodies were purchased from Jackson ImmunoResearch (West Grove, PA, USA). Western blots were visualized with ECL Prime Western Blot Detection Reagent (GE Healthcare, Buckinghamshire, UK) and Odyssey Fc (Li-Cor, Lincoln, Nebraska, USA).

## Results

**MS Overview.** All results were obtained from the average of the MS biological replicates. Though there was a small subset of exclusive phosphopeptides, the overwhelming majority of phosphorylation events were common to both cell lines. Overall, MS analysis identified 2,382 shared phosphopeptides out of 3,712 peptide groups (Figure 1 and Table I). There were commonly 1-3 phosphorylations on each phosphopeptide, which subsequently totaled to 3,173 phosphosites (p-sites). Phosphorylated serine was over-represented among the identified p-sites at 85% (2,698/3,173). In contrast, threonine and tyrosine constituted a much smaller share of the total p-sites. Only 13% of the p-sites was phosphorylated threonine (426/3,173), with the rarest p-site identified of just 2% (49/3173) being phosphorylated tyrosine. Proteome Discoverer produced a ratio derived from peptides modified with heavy ( $^{13}\text{C}_2\text{D}_6$ ) vs. light ( $\text{C}_2\text{H}_6$ ) dimethylation sites. This gave the relative abundance for the phosphopeptides as ratio of (HCC +V)/(HCC), ranging from 7.14 to 0.0285. Global intensity values were similar between the two conditions (Figure 2a). A small subset of phosphopeptides showing differential intensity was discernable when excluding the 10% highest and lowest values from the range (Figure 2b).

**Most abundant phosphopeptides.** The 30 most abundant phosphopeptides in HCC with and without virus are shown (Table II). The XCorr Sequest HT score ranges from 1.24 to 4.92 for HCC and from 1.39 to 5.33 for HCC+V. Majority of these scores correlated with a high hit confidence level (Table II). Gene ontology enrichment of this top subset revealed that most abundant proteins identified for HCC were pronounced for the biological process category of cytoskeleton organization (ID:GO:0007010, adjusted

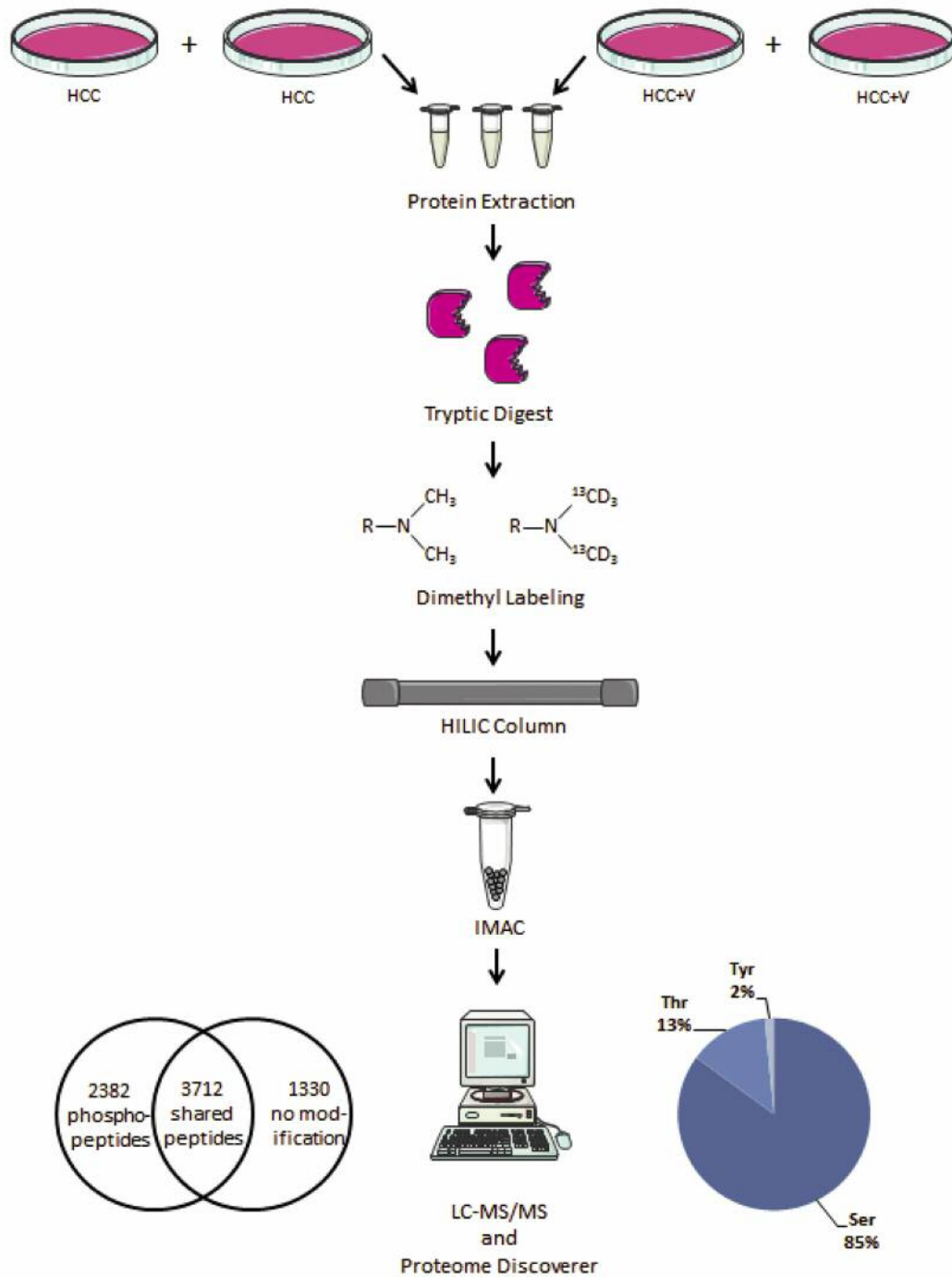
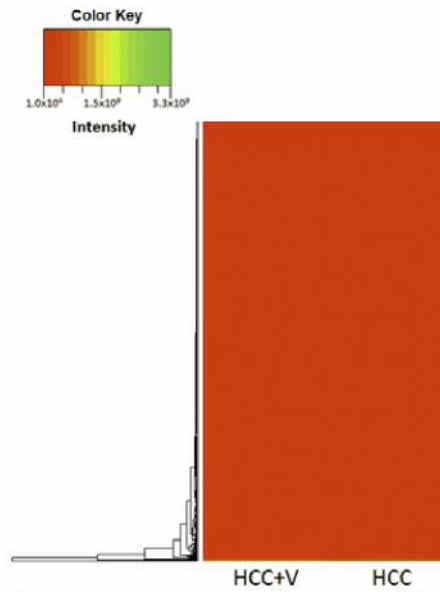


Figure 1. Experimental workflow. The sequence of events for each biological replicate, as well as the resulting quantified phosphopeptides and phosphosite amino acids averaged from the experiments.

$p=0.0186$ ). Conversely, the HCC+V most abundant phosphopeptides predominantly belonged to the execution phase of apoptosis category (ID:GO:0097194, adjusted  $p=0.0307$ ).

**Phosphorylation fold change.** Overall, there were more phosphopeptides with higher abundance in HCC compared to HCC+V. Three hundred and forty (14%) phosphopeptides had  $\geq 1.5$ -fold abundance in HCC+V and 673 (28%)

a) All phosphopeptide intensities.



b) Heat map depicting proteins with the greatest intensity value difference between HCC and HCC+V.

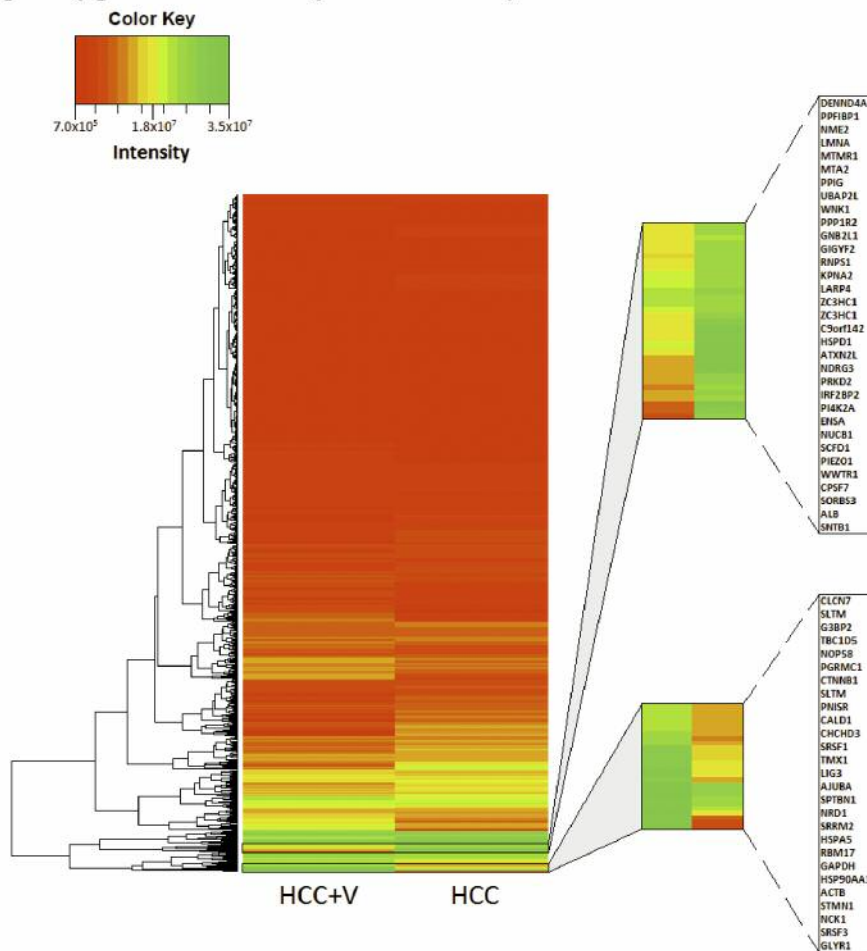


Figure 2. Intensity heat maps. (a) Heat map for all phosphopeptide values. (b) Heat map for phosphopeptides excluding top 10% and bottom 10% of intensity values.

Table II. Most abundant phosphopeptides. List of the top 1% (top 30) most abundant phosphopeptides identified in (2a) HCC and (2b) HCC+V. Peptide stringency was set at a high ion cutoff score of FDR=5%.

(IIa) HCC List

Protein ID	Protein Name	Peptide Sequence	XCorr Sequest HT Score	Confidence	Abundance (HCC+V)/ (HCC)
Q68CZ2-1	Tensin-3	[R].HPPFSPPEPPLSSPASQHK.[G]	3.64	High	0.03
O95817	BAG family molecular chaperone regulator 3	[R].SSTPLHSPSPIR.[V]	2.46	High	0.04
Q9HCM4	Band 4.1-like protein 5	[K].QLEMENSPLLSPR.[S]	3.48	High	0.05
Q14999-2	Cullin-7	[R].SHDWSSLATR.[G]	1.24	Medium	0.05
Q13522	Protein phosphatase 1 regulatory subunit 1A	[K].STLAMSPR.[Q]	2.41	High	0.06
O15061-1	Synemin	[R].SLRQQLDELSWATALAEGER.[D]	1.48	High	0.06
Q13017-1	Rho GTPase-activating protein 5	[R].THSDASDDAEFTTSK.[T]	4.01	High	0.06
Q86X27-1	Ras-specific guanine nucleotide-releasing factor RalGPS2	[R].SAASREDLVGPEVGASPGSGR.[K]	4.92	High	0.06
P35658-5	Nuclear pore complex protein Nup214	[R].ITPPAAKPGSPQAK.[S]	2.07	High	0.07
P46821	Microtubule-associated protein 1B	[K].SDISPLTPR.[E]	2.82	High	0.08
Q6ZSZ5	Rho guanine nucleotide exchange factor 18	[R].SLSPILPGR.[H]	2.56	High	0.08
Q6UUV7-1	CREB-regulated transcription coactivator 3	[R].SSSGLQSSRSNPSIQATLTK.[T]	4.14	High	0.08
Q15139	Serine/threonine-protein kinase D1	[K].SPSEFSGR.[E]	2.19	High	0.08
P24752	Acetyl-CoA acetyltransferase, mitochondrial	[R].SRSPLLR.[R]	2.49	Medium	0.08
O75815-1	Breast cancer anti-estrogen resistance protein 3	[R].TGSEPALSPAVVR.[R]	2.90	High	0.08
A6NGB9	WAS/WASL-interacting protein family member 3	[K].TISGLIPPASPR.[L]	2.59	High	0.08
P29966	Myristoylated alanine-rich C-kinase substrate	[K].AEDGATPSPNETPKK.[K]	3.32	High	0.09
P02768-1	Serum albumin	[K].QEPERNECFQHKDDNPPLPR.[L]	2.49	Medium	0.09
O75385	Serine/threonine-protein kinase ULK1	[R].NLQSPQFQTPR.[S]	1.95	High	0.09
Q9H1Z4	WD repeat-containing protein 13	[R].SVSRGSYQLQAQMNR.[A]	2.33	Medium	0.10
Q9NYB9	Abl interactor 2	[R].HTPPTIGGSLPYR.[R]	1.83	High	0.10
Q9NVD7	Alpha-parvin	[K].SPSVPKSPTPKSPPSR.[K]	2.05	High	0.10
Q9GZR1	Sentrin-specific protease 6	[R].HCSTYQTPPLSPASKK.[C]	2.59	High	0.10
P33991	DNA replication licensing factor MCM4	[R].IAEPSVCGR.[C]	1.86	High	0.10
O43741	5'-AMP-activated protein kinase subunit beta-2	[R].DLSSPPGPYGQEMYAFR.[S]	2.40	High	0.10
Q53SF7-1	Cordon-bleu protein-like 1	[R].DQTASAPATPLVVK.[H]	1.97	High	0.10
Q9BTU6	Phosphatidylinositol 4-kinase type 2-alpha	[R].VAAAAGSGSPPGSPG HDRERQPLLDL.[A]	2.23	High	0.11
P08670	Vimentin	[R].SLYASSPGGVYATRSSAVR.[L]	2.62	High	0.11
O60716-1	Catenin delta-1	[R].SQSSHSYDDSTLPLIDR.[N]	2.57	High	0.12
O95819-3	Mitogen-activated protein kinase kinase kinase 4	[R].AASSLNLSNGETESVK.[T]	2.05	High	0.12

(IIb) HCC+V List

Protein ID	Protein Name	Peptide Sequence	XCorr Sequest HT Score	Confidence	Abundance (HCC+V)/ (HCC)
P16402	Histone H1.3	[R].KASGPPVSELITK.[A]	2.61	High	7.14
Q96A00	Protein phosphatase 1 regulatory subunit 14A	[R].QPSPSHDGLSPLQDR.[A]	3.42	High	5.88
Q49A26-1	Putative oxidoreductase GLYR1	[R].KLSLSEK.[V]	1.49	High	5.56
Q5VZL5	Zinc finger MYM-type protein 4	[R].KKSIVAVEPR.[S]	2.57	High	5.56
Q6P3W7	SCY1-like protein 2	[K].RASLTLEEK.[Q]	2.33	High	5.26
O60825-1	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2	[R].RNSFTPLSSNTIR.[R]	1.94	High	5.00
Q5SYE7-2	NHS-like protein 1	[R].KNSGAEAAQLSER.[T]	2.24	High	5.00
Q92615	La-related protein 4B	[R].KNSFGYR.[K]	1.69	Medium	5.00
P15924-1	Desmoplakin	[R].SMSFQGIR.[Q]	2.73	High	4.76
Q02363	DNA-binding protein inhibitor ID-2	[R].KNSLSDHSLGISR.[S]	1.86	High	4.76
Q15149-4	Plectin	[R].SSSVGSSSYPIPAVSR.[T]	2.88	High	4.76
P21860-1	Receptor tyrosine-protein kinase erbB-3	[K].RESGPGIAPGPEPHGLTNK.[K]	2.87	High	4.55

Table II. Continued

Table II. Continued

Protein ID	Protein Name	Peptide Sequence	XCorr Sequest HT Score	Confidence	Abundance (HCC+V)/ (HCC)
P84103	Serine/arginine-rich splicing factor 3	[R].GPPPSWGR.[R]	1.78	High	4.17
Q9H2H9	Sodium-coupled neutral amino acid transporter 1	[R].RSLTNSHLEK.[K]	1.67	Medium	4.17
Q9UHB6-4	LIM domain and actin-binding protein 1	[K].SEVQQPVHPKPLSPDSR.[A]	3.71	High	4.17
Q9H501	ESF1 homolog	[K].FKIDSNISPK.[K]	2.85	High	4.00
Q9UKV3-1	Apoptotic chromatin condensation inducer in the nucleus	[R].RLSQPESAEK.[H]	1.39	High	4.00
P02458-2	Collagen alpha-1(II) chain	[K].QGAPGASGDRGPPGPVGP PGLTGPAEPGR.[E]	3.38	High	3.85
Q16875-4	6-phosphofructo-2-kinase/ fructose-2,6-bisphosphatase 3	[R].RNSVTPLASPEPTK.[K]	1.92	High	3.85
P20700	Lamin-B1	[R].AGGPTTPLSPTR.[L]	1.53	High	3.70
Q9NYB0	Telomeric repeat-binding factor 2-interacting protein 1	[K].ENARSPSSVTGNALWK.[A]	2.00	High	3.57
Q09666-1	Neuroblast differentiation-associated protein AHNAK	[K].ASLGSLEGEAEAEASSPK.[G]	2.85	High	3.45
Q16649	Nuclear factor interleukin-3-regulated protein	[K].GPIHSPVELK.[H]	2.59	High	3.45
P16422	Epithelial cell adhesion molecule	[R].QCQCTSVGAQNTVICSK.[L]	2.58	High	3.33
P31323	cAMP-dependent protein kinase type II-beta regulatory subunit	[R].RASVCAEAYNPDEEEDDAESR.[I]	5.33	High	3.33
Q15642	Cdc42-interacting protein 4	[R].APSDSSLGTSPDGRPELR.[G]	2.03	High	3.33
Q8N3X1-2	Formin-binding protein 4	[R].RPILQLSPPGPR.[G]	2.49	High	3.33
Q8TEX9-2	Importin-4	[K].ACQSCPSENTAALQAALAR.[V]	2.92	High	3.33
P51858-3	Hepatoma-derived growth factor	[K].GNAEGSSDEEGKLVIDEPAKEK.[N]	3.27	High	3.23
Q09161	Nuclear cap-binding protein subunit 1	[R].KTS DANETEDHLESLICK.[V]	2.07	High	3.23

phosphopeptides had  $\geq 1.5$ -fold abundance in HCC. To visually represent the phosphorylation fold change difference between the two conditions, the abundance ratio is depicted as  $\text{Log}_{10} [(HCC+V)/HCC]$  (Figure 3). Additionally, there were 1,369 other phosphopeptides that did not reach this 1.5-fold threshold.

**Functional analysis.** Shared phosphopeptides were used for gene ontology (GO) enrichment to assess differences in cellular component, biological process, and molecular function. Outcomes of GO analysis are shown and classified into the three abundance groups:  $\geq 1.5$ -fold in HCC,  $\geq 1.5$ -fold in HCC+V, and  $< 1.5$ -fold in HCC and HCC+V. Functional analysis revealed that HCC and HCC+V shared the most resemblance among cellular components (Figure 4a), with one exception being a subset of cytoskeletal categories (*i.e.* actin cytoskeleton) that were most abundant in HCC. Additionally, the two conditions had several similarities in molecular function (Figure 4b), especially among binding categories. However, the phosphoproteins among signaling categories like Ras/Rho GTPase activity were highest in the HCC abundance group. HCC and HCC+V differed most significantly when comparing biological processes (Figure 4c). The phosphoproteins within many biological process

categories were prominent for very specific abundance groups. For instance, cytoskeletal dynamics and signaling transduction (*i.e.* response to insulin stimulus, Ras/Rho signal transduction) were predominantly enriched in HCC. Conversely, other biological processes such as regulation of gene expression and transcription were notable in HCC+V.

**Protein network crosstalk and pathway analysis.** The abundance groups were used to identify a defining network of proteins, whose crosstalk activity among signal transduction pathways may depict the cellular dynamics of liver cancer. The protein networks for HCC and HCC+V are composed of a set of 76 and 41 phosphoproteins respectively (Figure 5a and 5b). The network components and their abundances are also provided (Table III). The corresponding pathways that each network could be differentially regulating through phosphorylation are listed (Table IV). The HCC pathways included several Ras/GTPase mediated signaling transduction such as insulin-like growth factor 1 (IGF1) and ADP-ribosylation factor 6 (Arf6) pathways. In comparison, HCC+V pathways included vascular endothelial growth factor (VEGF) and PI3K/AKT signaling events. Additionally, there were certain pathways (TRAIL and PAR1 signaling) that were common to both networks.

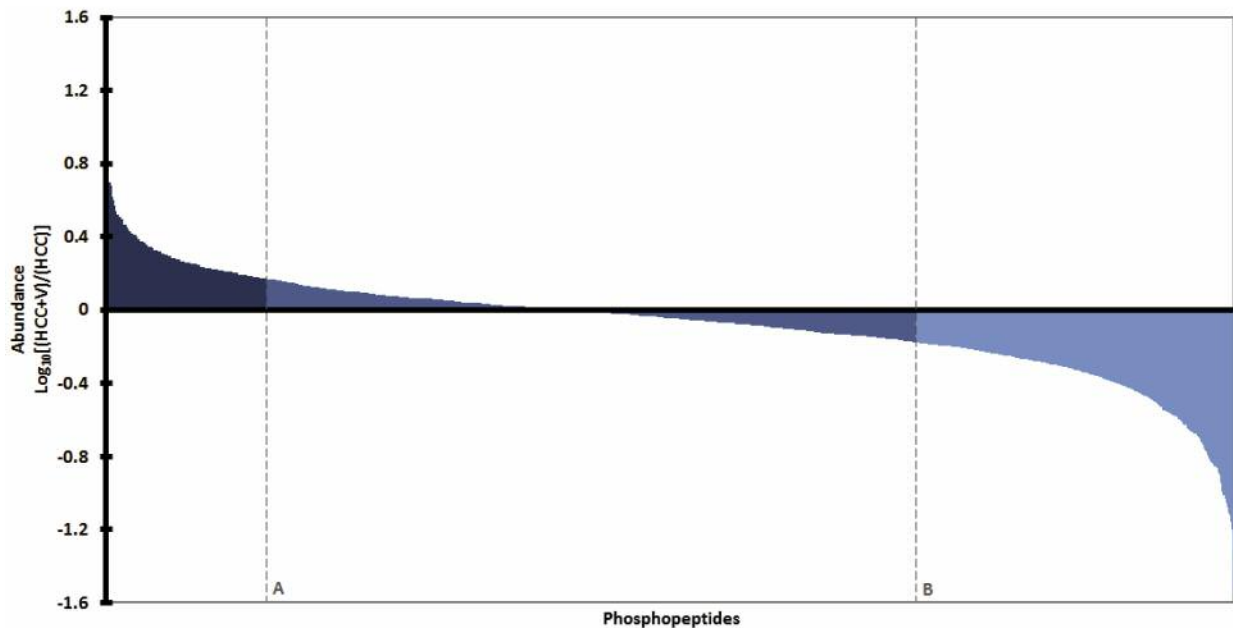


Figure 3. Phosphorylation landscape. The protein abundance of phosphopeptides in the HCC and HCC+V abundance groups. The abundance groups are gated on a threshold of  $\geq 1.5$ -fold higher relative abundance (dotted lines A and B), which is calculated as the ratio of HCC+V over HCC.  $\log_{10}$  of the relative abundance ratio is depicted, resulting in positive y-values (left of line A) for the HCC+V group and negative y-values (right of line B) for the HCC group.

An individual protein was selected from the protein networks for further pathway analysis of upstream kinases and downstream targets to characterize the interactions that could be developed as candidates for further study. From the HCC network, insulin receptor substrate 1 (IRS1) participates in multiple pathways controlling cell growth, proliferation, and inflammatory response (Figure 6a). From the HCC+V network, stathmin (STMN1) is targeted by several kinases in the CDK, MAPK, and PAK families that affect cell proliferation and survival to subsequently induce cytoskeletal deregulation (Figure 6b).

**Initial validation.** Western blots were performed for preliminary validation of mass spectrometry-quantified proteins and overall functional analysis. All selected proteins on western blot analysis, with the exception of PAK1, were parallel our MS quantitative data (Figure 7). These proteins were chosen for western blot validation based on network and pathway analyses suggesting their pivotal roles in viral or non-viral HCC.

**Stathmin.** Of the proteins selected in our initial validation, STMN1 appears as a promising candidate for future study (Figure 8). MS data characterizes that HCV potentially controls downstream signals *via* phosphorylation of STMN1 at multiple serines (16, 25, 38, 63), but most predominantly

*via* S38 and S63. Western blot analysis of phosphorylated STMN1 correlated with MS results for S16. However, western blot findings differed from MS by showing a higher protein expression for P-STMN1 S25 and no notable difference for P-STMN1 S38.

## Discussion

While MS is a powerful tool to study the phosphoproteome and other PTMs, the overwhelming amount of information that can be generated requires a systematic approach to analyze the typically large data sets produced. A simple framework of narrowing-down a global phosphoproteomic screen by focusing on protein function, pathways, and networks, enables us to visualize how HCV may influence cancer growth and identify the protein candidates for future molecular testing. Herein we presented the first study analyzing the phosphorylation landscape of HCC cell lines with or without virus. We demonstrate that an IMAC enrichment technique is successful and suitable for relative phosphoproteomic quantitation. This method strongly detected phosphorylated serine, that could reflect that the natural distribution of phosphorylation sites *in vitro* is predominantly serine (12). A more selective antibody-based MS technique could overcome this limitation and might be necessary if tyrosine and threonine phosphorylation is of



further interest (13, 14). Historically, MS has been used for the identification of unmodified proteins. The capabilities of MS are enhanced by adding relative quantitation and focusing on PTMs like phosphorylation to explore the particular proteins that may be involved in differential regulation. This enables MS to more closely capture *in vivo* conditions that could more easily address clinical challenges such as screening for liver cancer in high risk patient populations. A current hurdle in the field of proteomics lies in the bioinformatics analysis of MS results, as there are numerous and diverse softwares available. In order to capture the dynamic cellular responses, discern relevant signals from noise, and define novel protein interactions, two core concepts facilitated our framework in interpreting the HCC and HCC+V phosphoproteomes. The first important step begins with asking the right questions of interest to guide data analysis. In our experience, having a very specific question can expedite how one strategically processes a large data set. For instance, our main question was to interrogate how HCV can influence liver cancer transformation and subsequently this question determined the various techniques we selected to further examine our data.

To address this question, we initially surveyed the MS results with the most global view of intensity and abundance. An initial assessment of intensity values appears similar for our experimental conditions, given the two cell lines are hepatocyte derived. Furthermore, when intensity values have a large range, differences may be misleadingly depicted as similar. By excluding the extreme intensity values (*i.e.* lowest and highest 10% values) previously hidden differences between the tested conditions can be unmasked. Looking at intensity could be one option to initially visualize different phosphopeptide molar responses and provide a comparison for quantitation values, as an extended view of intensity would logically progress towards a closer look at different protein abundances.

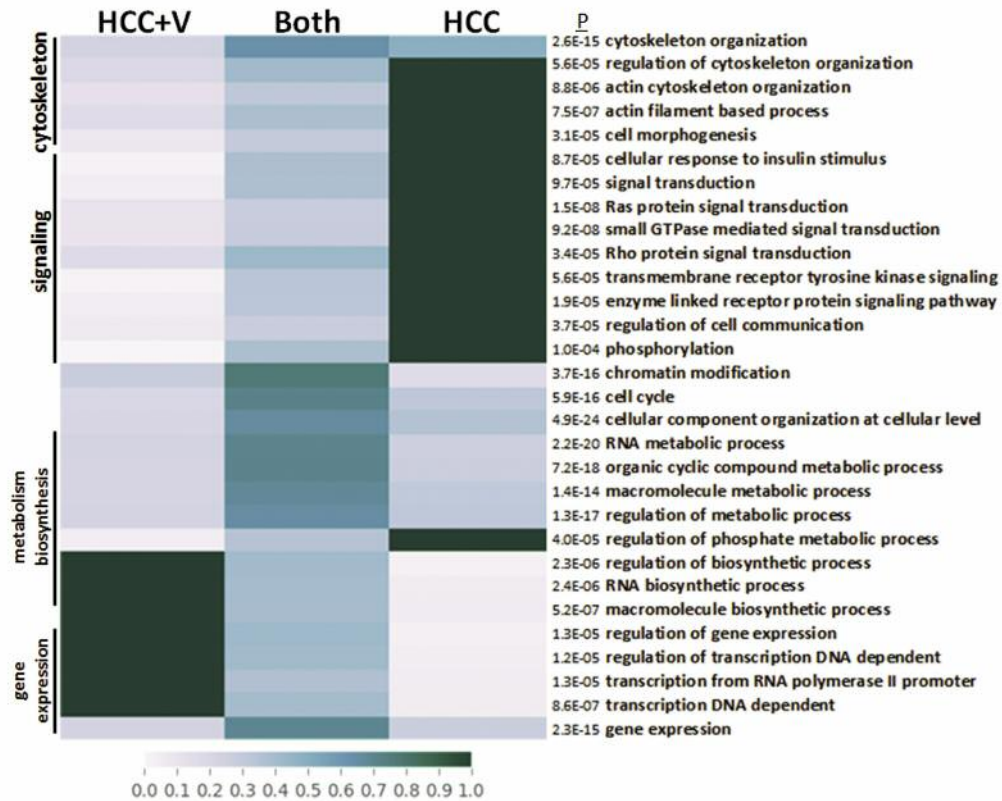
We began interrogating our results by looking at the 1% (top 30) most abundant hits for each experimental condition. Phosphoproteins pertaining to cytoskeletal dynamics were noted for HCC and phosphoproteins related to apoptosis were identified for HCC+V. Although this cursory examination is a simple step that can be used to start analysis, MS experiments should supplement these results with a more comprehensive look at protein dynamics to find deregulated pathways and functions that cannot be easily derived from the top hits. Given that changes in cellular dynamics may stem from large or small differences in protein levels, analysis should encompass the smaller subset of phosphopeptides with the highest abundance as well as those with lower abundance which may still influence relevant biological functions. Next, to be more inclusive of different protein levels than just the top 1%, we separated our phosphoproteome data into different thresholds of

abundance. We focused our subsequent functional analysis on the population of  $\geq 1.5$ -fold abundant phosphopeptides to narrow down proteins that may have a greater role in cellular dynamics based on this degree of quantitative changes within the cell. Although a threshold of 1.5 fold abundance was chosen for this data set, other studies have implemented a more conservative threshold (2.0-fold) for analysis (15, 16). Selection of a more liberal or conservative threshold could be dictated by the total number of peptides obtained from MS. For instance, our abundance threshold (1.5-fold) may be considered more conservative within the context of a smaller data set. This approach is motivated by previous studies noting that significant targets may emerge from the combination of lower abundant proteins that may still be impacting important functional variation, even though their expression may not be the highest (17). By setting a  $\geq 1.5$ -fold threshold we obtained approximately double the amount of HCC phosphopeptides for analysis compared to HCC+V. The disparity between these two groups could reflect the more limited amount of available MS database information for viral HCC. Furthermore, the smaller amount of HCC+V abundant phosphopeptides may be due to HCV infected hepatocytes predominantly utilizing a different PTM.

To further characterize this global difference between phosphopeptide abundance, our subsequent analysis seeks to investigate the functional features of each  $\geq 1.5$ -fold abundant population. GO annotation of biological processes, molecular functions, and cellular components enabled us to visualize the functional landscape of each cell line. Given the two cell lines are hepatocyte derived, it is expected that the two cell lines share many similar cellular components of their phosphopeptides. Yet, the presence or absence of the virus has a more distinct impact on the molecular functions and biological processes of the identified phosphopeptides (Figure 4). HCC abundant phosphopeptides were primarily enriched among Ras/Rho signal transduction activity and cytoskeletal organization categories. This suggests that without the virus, HCC is influenced by deregulation of cell growth signaling pathways and cytoskeletal dynamics. In contrast, HCC+V abundant phosphopeptides were highest among certain gene expression categories such as regulation of transcription. The amount of HCC+V-abundant phosphopeptides in these categories could reflect how the virus activates differential pathways within the host cell to promote its own survival. Therefore, hepatocyte transformation and growth could occur through alternate means when cells are infected with hepatitis C.

To further elaborate on biological changes perturbed by the virus, we aimed to look at the signaling events that could be enhanced to examine the possible crosstalk among abundant phosphopeptides in HCC or HCC+V. Analysis identified distinctive protein networks that corresponded with different sets of signaling transduction pathways. While

(4a) Cellular Component



(4b) Molecular Function

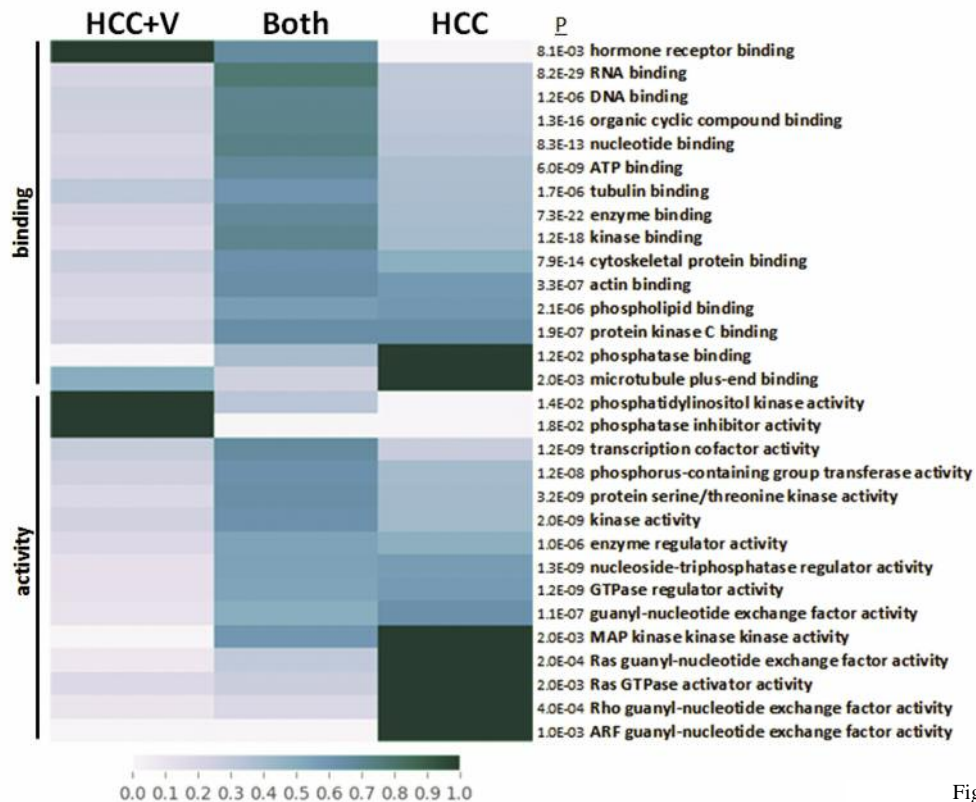


Figure 4. Continued

## (4c) Biological Process

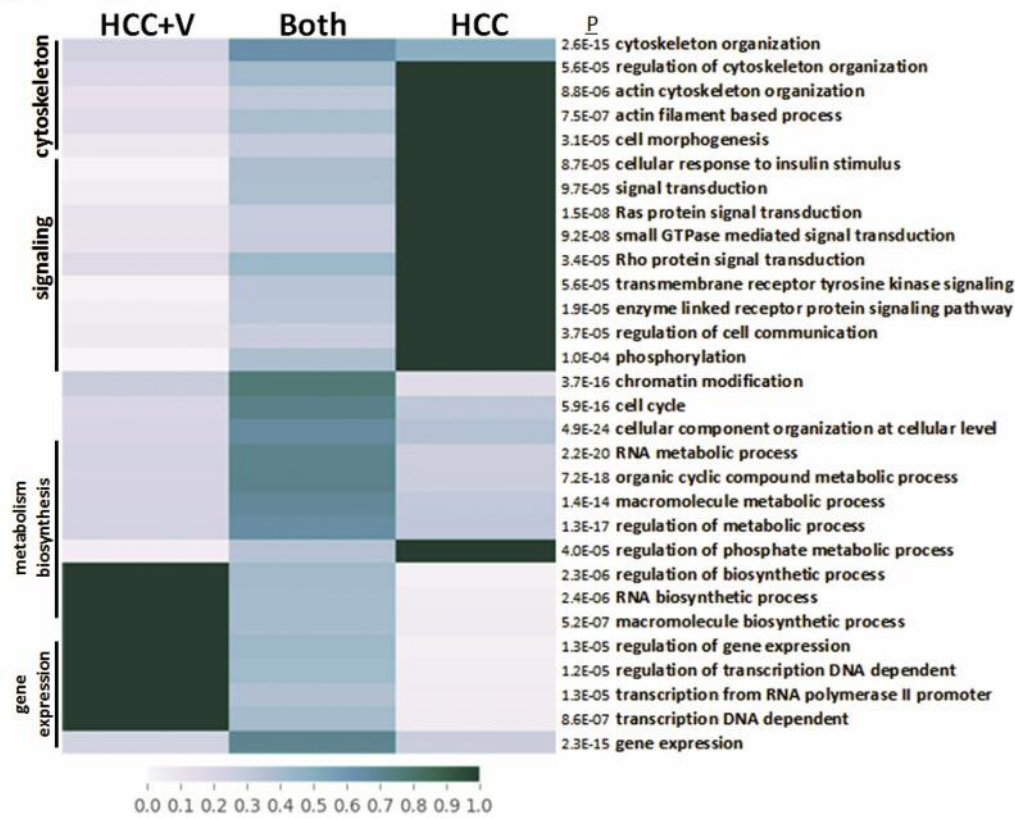


Figure 4. Functional analysis heat map. Comparison of gene ontology analysis organized by abundance groups: (HCC:  $\geq 1.5$ -fold abundance in HCC, HCC+V:  $\geq 1.5$ -fold abundance in HCC+V, and Both:  $< 1.5$ -fold abundance in HCC and HCC+V). The number of phosphoproteins per abundance group is normalized to the total number of category proteins identified from the entire set of shared phosphopeptides. The heat maps with p-values (adjusted for multiple testing) are divided into (a) cellular component (b) molecular function and (4c) biological process.

some of these pathways were shared, most were unique, presumably based on the presence or absence of the virus. Investigating the characteristic signaling of an experimental condition could be exploited for diagnostic or therapeutic development in multiple ways. One option is to use this method to identify any previously studied mechanisms that have known available agents, facilitating research of pathway inhibition or excitation. For example, our analysis suggests that the effect of the VEGF pathway (Table IV) in HCC+V could be easily studied by using inhibitors like bevacizumab (Avastin; Genentech Inc., South San Francisco, CA, USA) (18). In addition, this method of analysis can generate individual protein candidates for further research that have not been previously associated with HCC±V. By inspecting the crosstalk protein sets, one can choose a rational target that has both higher abundance and possible functional consequences through the manipulation of multiple pathways. For instance, we selected IRS1 from the HCC phosphoprotein network as a possible new candidate

for research (Figure 6a). Similarly, the protein STMN1 was selected from the HCC+V set of crosstalk phosphoproteins as a target of interest for further study of viral infected HCC (Figure 6b).

Our analysis takes into account an inclusive spectrum of: the top 1% most abundant phosphoproteins from MS data, the set of  $> 1.5$ -fold abundant proteins that could be influencing changes in biological dynamics and signaling pathways of interest, and even those proteins that may not be on the original MS data but are identified after bioinformatics analyses. This comprehensive strategy provides a basic framework for the logical classification of a large proteomic data set. The emphasis of such an approach is on functional changes within each experimental condition that may be driven by any degree of protein abundance. Specifically, if cellular function can be linked in a non-biased fashion to different changes in protein abundances, MS global data analyses would generate hypotheses that can more closely mirror *in vivo* conditions.



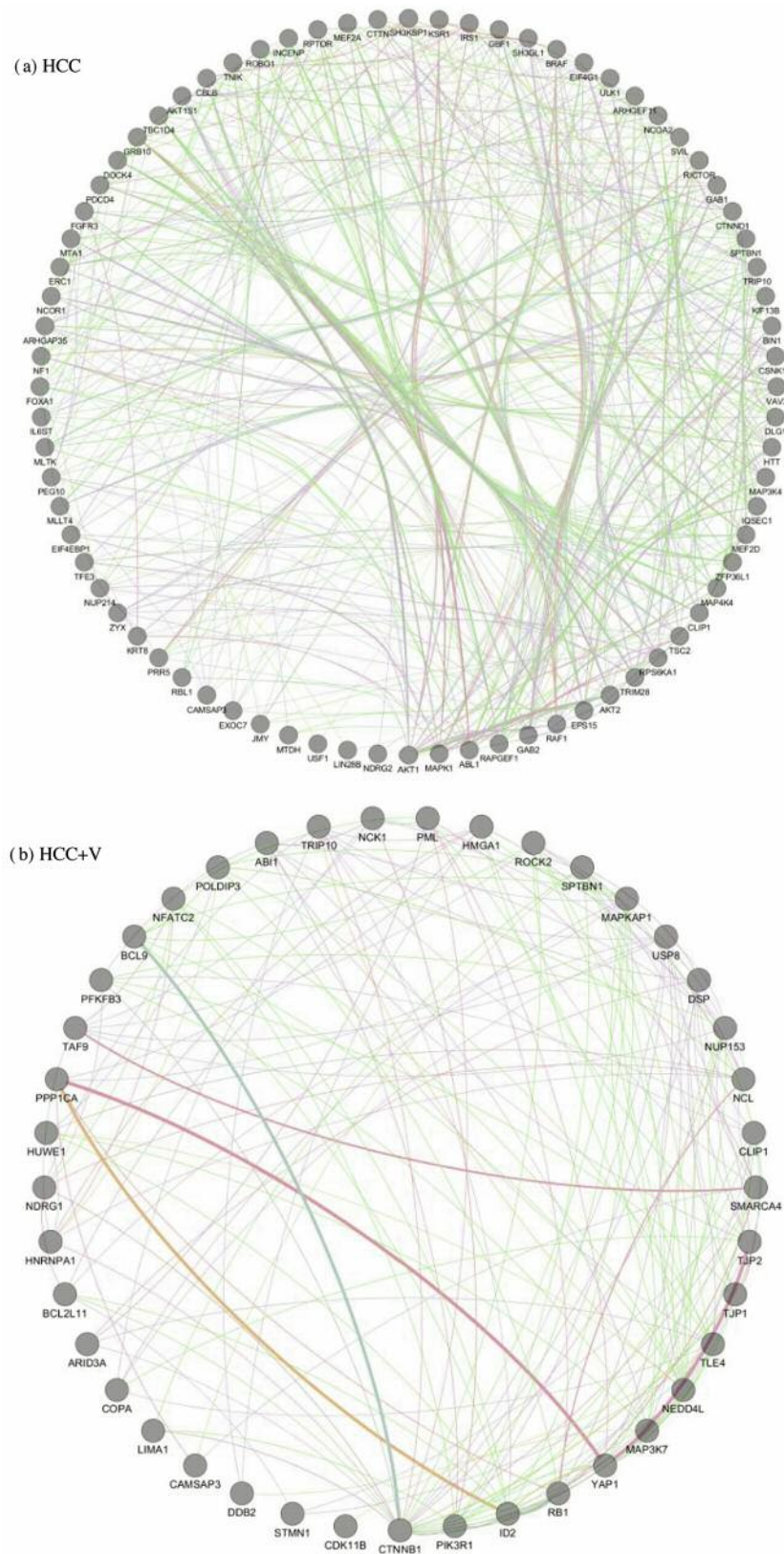


Figure 5. Protein networks. Protein networks for (A) HCC and (B) HCC+V. Edges represent the following relationships: genetic interactions (green), physical interactions (red), pathway (light blue), predicted (orange), co-expression (purple), co-localization (dark blue), and shared protein domain (olive).

Table III. *Phosphoprotein Crosstalk Networks. List of the phosphoprotein crosstalk networks generated from the (IIIa) HCC and (IIIb) HCC+V abundance groups. The abundance groups are gated on a threshold of  $\geq 1.5$ -fold higher relative abundance.*

## (IIIa) HCC Network

Protein ID	Protein Name	Peptide Sequence	XCorr Sequest HT Score	Confidence	Abundance (HCC+V)/ (HCC)
P35658	Nuclear pore complex protein Nup214	[R].ITPPAAKPGSPQAK.[S]	2.07	High	0.07
O75385	Serine/threonine-protein kinase ULK1	[R].NLQSPTQFQTPR.[S]	1.95	High	0.09
O60716	Catenin delta-1	[R].SQSSHSYDDSTLPLIDR.[N]	2.57	High	0.12
O95819	Mitogen-activated protein kinase kinase kinase kinase 4	[R].AASSLNLSNGETESVK.[T]	2.05	High	0.12
P28482	Mitogen-activated protein kinase 1	[R].VADPDHDTGFLTEYVATR.[W]	1.86	High	0.14
Q9NQ78	Kinesin-like protein KIF13B	[R].ERPDLAPAPGSPFR.[V]	2.82	High	0.14
Q8N9B5	Junction-mediating and -regulatory protein	[R].SSAWAEGGSPR.[S]	2.87	High	0.15
Q6DN90	IQ motif and SEC7 domain-containing protein 1	[K].LQHSTSLR.[K]	1.41	High	0.16
Q53EL6	Programmed cell death protein 4	[R].DSGRGDSVSDSGDALR.[S]	2.44	High	0.18
P42566	Epidermal growth factor receptor substrate 15	[K].IGTPTRPCPLPPGKR.[S]	1.63	High	0.21
Q96B36	Proline-rich AKT1 substrate 1	[R].TEARSSDEENGPPSS PDLDRIAASMR.[A]	2.86	High	0.21
Q9UKE5	TRAF2 and NCK-interacting protein kinase	[K].SEGSPVLPHEPAK.[V]	2.79	High	0.22
P55196	Afadin	[R].VTRSQEELREDK.[A]	2.67	Medium	0.23
Q13322	Growth factor receptor-bound protein 10	[R].SIQPQVSPR.[Q]	2.23	High	0.25
P22607	Fibroblast growth factor receptor 3	[K].GLGSPTVHK.[I]	2.68	High	0.26
Q9UN36	Protein NDRG2	[R].SRTASLTSAASVDGNR.[S]	3.16	High	0.27
O95425	Supervillin	[R].SPSFGDPQLSPEARPR.[CV]	2.63	High	0.27
P40189	Interleukin-6 receptor subunit beta	[K].SHIAQWSPHTPPR.[H]	2.42	High	0.27
P42858	Huntingtin	[K].EKEPGEQAQVPLSPK.[K]	3.86	High	0.27
O15085	Rho guanine nucleotide exchange factor 11	[R].NSVLSDPGLDSPRTSPVIMAR.[V]	2.60	High	0.27
Q9Y6N7	Roundabout homolog 1	[K].TFNSPNLKDGR.[F]	1.95	High	0.29
Q12959	Disks large homolog 1	[R].EQMMNSSISSGSGSLR.[T]	4.60	High	0.30
Q9UPT5	Exocyst complex component 7	[K].SSSSSGVPYSPAIPNKR.[K]	1.48	High	0.31
O00499	Myc box-dependent-interacting protein 1	[K].GNKSPSPPDGSPAATPEIR.[V]	3.96	High	0.31
Q13541	Eukaryotic translation initiation factor 4E-binding protein 1	[K].FLMECRNSPVTK.[T]	1.42	High	0.32
O75376	Nuclear receptor corepressor 1	[K].AQLSPGIYDDTSAR.[R]	2.76	High	0.32
Q9UQC2	GRB2-associated binding protein 2	[R].RNTLPAMDNSR.[L]	2.28	High	0.34
P30622	CAP-Gly domain-containing linker protein 1	[K].TASESISNLSEAGSIK.[K]	2.84	High	0.34
Q86TG7	Retrotransposon-derived protein PEG10	[K].ASKSSPAGNSPAPL.[-]	3.43	High	0.35
Q02078	Myocyte-specific enhancer factor 2A	[R].ASPNLIGATGANSLGK.[V]	2.43	High	0.36
Q14814	Myocyte-specific enhancer factor 2D	[K].VIPAKSPPPPTHSTQLGAPSR.[K]	3.31	High	0.36
Q01082	Spectrin beta chain, non-erythrocytic 1	[K].TALPAQSAATLPR.[T]	3.94	High	0.37
P85299	Proline-rich protein 5	[R].SKSYNTPLLNPVQEHEA EGAAAGGTSIR.[R]	3.28	High	0.37
Q15942	Zyxin	[K].LGHPREALSAGTGSPQP PSFTYAQQR.[E]	3.27	High	0.37
P55317	Hepatocyte nuclear factor 3-alpha	[R].KDPGASNPADSPLHR.[G]	3.78	High	0.39
Q13263	Transcription intermediary factor 1-beta	[K].RSRSGEGEVSGLMR.[K]	2.32	High	0.40
P52735	Guanine nucleotide exchange factor VAV2	[R].ASSRSPVFTPR.[V]	1.65	Medium	0.42
P19532	Transcription factor E3	[R].AASDPLLSSVPAVSK.[A]	4.16	High	0.44
Q9NQS7	Inner centromere protein	[R].HSPIAPSSPSPQVLAQK.[Y]	2.19	High	0.46
O60343	TBC1 domain family member 4	[R].AQGVRSPLLR.[Q]	1.67	High	0.47
P05787	Keratin, type II cytoskeletal 8	[R].AFSSRSYTSGPSR.[I]	1.35	Medium	0.47
Q15642	Cdc42-interacting protein 4	[R].VLSNRGDSLSR.[H]	1.75	High	0.48
Q8N122	Regulatory-associated protein of mTOR	[R].VLDTSSLTQSAPASPTNK.[G]	3.01	High	0.49
Q9Y6R4	Mitogen-activated protein kinase kinase kinase 4	[R].HSSPTEERDEPAYPR.[G]	1.97	High	0.49
Q92538	Golgi brefeldin A resistant guanine nucleotide exchange factor 1	[R].ADAPDAGAQSDE LPSYHQNDVSLDR.[G]	5.80	High	0.49
Q8IVT5	Kinase suppressor of Ras 1	[R].RTESVPSDINNPDVDR.[A]	2.93	High	0.49
Q86UE4	Metadherin	[R].TVEVAEGEAVRTPQSVTAK.[Q]	1.84	High	0.50

Table III. *Continued*

Table III. *Continued*

Protein ID	Protein Name	Peptide Sequence	XCorr Sequest HT Score	Confidence	Abundance (HCC+V)/ (HCC)
Q04637	Eukaryotic translation initiation factor 4 gamma 1	[K].RSFSKEVEER.[S]	2.09	Medium	0.50
P04049	RAF proto-oncogene serine/ threonine-protein kinase	[R].AAHTEDINACTLTTSPPR.[L]	2.89	High	0.50
Q13191	E3 ubiquitin-protein ligase CBL-B	[R].KCTDRQNSPVTSPGSSPLAQR.[R]	2.71	High	0.51
P15056	Serine/threonine-protein kinase B-raf	[R].SNPKSPQKPIVR.[V]	1.96	Medium	0.52
P00519	Tyrosine-protein kinase ABL1	[R].SSSFREMDGQPERR.[G]	2.06	Medium	0.53
Q6R327	Rapamycin-insensitive companion of mTOR	[K].SNSVSLVPPGSSHTLPR.[R]	2.46	High	0.55
Q13480	GRB2-associated-binding protein 1	[R].SSPKTPPRRPVPVADCEPPPVDR.[N]	2.35	High	0.55
Q96B97	SH3-domain kinase binding protein 1	[R].ANSPSLFGTEGKPK.[M]	2.58	High	0.56
P31749	RAC-alpha serine/threonine-protein kinase	[R].SGSPSDNSGAEMEVSALAKPK.[H]	3.33	High	0.57
Q8N110	Dedicator of cytokinesis protein 4	[R].ASPLLSDKHK.[H]	2.21	High	0.57
Q14247	Cortactin	[K].TQTPPVSPAPQPTTEERL PSSPVYEDAASF.[A]	5.11	High	0.59
Q9P1Y5	Calmodulin-regulated spectrin-associated protein 3	[R].HPLLSSGGPQSPLR.[G]	2.41	High	0.59
Q6ZNI7	Protein lin-28 homolog B	[R].EVGGGHGCTSPFPQEAR.[A]	3.81	High	0.61
P49815	Tuberlin	[R].GQPEGPLPSSSPRSPGLRPR.[G]	2.20	High	0.61
Q13330	Metastasis-associated protein MTA1	[R].SNMSPHGLPAR.[S]	2.79	High	0.61
P21359	Neurofibromin	[K].GSEGYLAATYPTVGQTSPPR.[A]	3.84	High	0.62
Q13905	Rap guanine nucleotide exchange factor 1	[K].LSKSDEQLSSLDSDSGQCSR.[N]	1.78	High	0.62
Q07352	Zinc finger protein 36, C3H1 type-like 1	[R].RHSVTLPSKK.[F]	2.24	High	0.62
P48730	Casein kinase I isoform delta	[R].GTQEVAPPTPLTPTSHTA NTSPRPVSGMER.[E]	2.41	High	0.63
P31751	RAC-beta serine/threonine-protein kinase	[K].CGSPSDSSTTEEMAVASK.[A]	2.10	High	0.64
P28749	Retinoblastoma-like protein 1	[R].SFAPSTPLTGR.[R]	1.77	High	0.64
Q9NYL2	Mitogen-activated protein kinase kinase kinase MLT	[R].VQSALNPHQSPDFKR.[S]	2.64	High	0.64
Q8IUD2	ELKS/RAB6-interacting/CAST family member 1	[KR].LPRSPR.[EL]	1.69	High	0.65
Q9NRY4	Rho GTPase activating protein 35	[R].TSFVSGSDDELGPRI.[K]	4.24	High	0.65
Q15596	Nuclear receptor coactivator 2	[R].MSPGVAGSPR.[I]	1.94	High	0.65
Q99961	Endophilin-A2	[K].IAASSSFR.[S]	1.69	High	0.65
P35568	Insulin receptor substrate 1	[R].HSSETFSSTPSATR.[V]	2.95	High	0.65
Q15418	Ribosomal protein S6 kinase alpha-1	[R].TPKDSPIPPSAGAHQLFR.[G]	2.10	Medium	0.66
P22415	Upstream stimulatory factor 1	[R].THPYSPK.[SI]	1.44	High	0.66

(IIIb) HCC+V Network

Protein ID	Protein Name	Peptide Sequence	XCorr Sequest HT Score	Confidence	Abundance (HCC+V)/ (HCC)
Q02363	DNA-binding protein inhibitor ID-2	[R].KNSLSDHSLGISR.[S]	1.86	High	4.76
P15924	Desmoplakin	[R].SMSFQGR.[Q]	2.73	High	4.76
Q9UHB6	LIM domain and actin-binding protein 1	[K].SEVQQPVHPKPLSPDSR.[A]	3.71	High	4.17
Q16875	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3	[R].RNSVTPLASPEPTK.[K]	1.92	High	3.85
Q15642	Cdc42-interacting protein 4	[R].APSDSSLGTPSDGRPELR.[G]	2.03	High	3.33
Q96PU5	E3 ubiquitin-protein ligase NEDD4-like	[R].SLSSPTVTLSAPLEGAK.[D]	3.50	High	3.13
P16333	Cytoplasmic protein NCK1	[K].RKPSVPDSASPADDSEFVDPGER.[L]	4.21	High	2.94
P17096	High mobility group protein HMG-I/HMG-Y	[R].KQPPVSPGTALVGS QKEPSEVPTPK.[R]	2.49	High	2.94
P06400	Retinoblastoma-associated protein	[K].ISEGLPTPTKMTPR.[S]	1.97	High	2.78
Q04727	Transducin-like enhancer protein 4	[K].SSSVSPSASFR.[G]	2.61	High	2.56
P30622	CAP-Gly domain-containing linker protein 1	[K].TASESISNLSEAGSIKKGER.[E]	2.31	High	2.56
P29590	Protein PML	[K].AVSPPHLDGPPSPR.[S]	1.69	High	2.5
Q9UDY2	Tight junction protein 2	[R].SRSWEDSPER.[G]	2.73	High	2.38
P46937	Transcriptional coactivator YAP1	[R].QASTDAGTAGALTPQHVR.[A]	4.92	High	2.38
Q9BPZ7	Mitogen-activated protein kinase associated protein 1	[R].TSFSFQK.[E]	1.31	Medium	2.33

Table III. *Continued*

Table III. Continued

Protein ID	Protein Name	Peptide Sequence	XCorr Sequest HT Score	Confidence	Abundance (HCC+V)/ (HCC)
P35222	Catenin beta-1	[R].RTSMGGTQQQFVEGVR.[M]	5.08	High	2.33
Q92466	DNA damage-binding protein 2	[R].SRSPLELEPEAKK.[L]	2.21	High	2.22
P27986	Phosphatidylinositol 3-kinase regulatory subunit alpha	[K].TRDQYLMWLTQKGVR.[Q]	1.71	High	2.13
P49790	Nuclear pore complex protein Nup153	[K].CIACQAAKLSPR.[D]	1.65	Medium	2.08
P40818	Ubiquitin carboxyl-terminal hydrolase 8	[R].SYSSPDITQAIQEEER.[K]	2.14	High	2.00
P51532	Transcription activator BRG1	[R].GRPPAEKLSNPNNLT.[K]	2.92	High	2.00
Q13469	Nuclear factor of activated T-cells, cytoplasmic 2	[R].TSPIMSPR.[T]	1.56	High	1.92
O75116	Rho-associated protein kinase 2	[R].RPSRQLAPNKPS.[-]	2.41	Medium	1.90
P16949	Stathmin	[R].RKSHEAEVLK.[Q]	2.14	High	1.85
Q92597	Protein NDRG1	[R].SRTASGSSVTSLDGTR.[S]	2.95	High	1.82
Q9P1Y5	Calmodulin-regulated spectrin-associated protein 3	[R].KFSPSQVPVQTR.[S]	1.97	High	1.82
P09651	Heterogeneous nuclear ribonucleoprotein A1	[K].SESPKEPEQLRK.[L]	2.61	High	1.79
O43521	BCL2-like 11 (apoptosis facilitator)	[R].RSSLLSR.[S]	2.17	High	1.75
P62136	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	[K].YGQFSLNPGGRPITPPR.[N]	1.62	Medium	1.72
Q9BY77	Polymerase delta-interacting protein 3	[K].RSSPAAFINPPIGTVTPALK.[L]	2.71	High	1.69
Q01082	Spectrin beta chain, non-erythrocytic 1	[R].TSSKESSPIPSPTSDR.[K]	2.96	High	1.69
O43318	Mitogen-activated protein kinase kinase 7	[K].RMSADMSEIAR.[I]	2.44	High	1.69
Q8IZP0	Abl interactor 1	[R].TNPPTQKPPSPMSGGR.[G]	2.72	High	1.69
Q99856	AT-rich interactive domain- containing protein 3A	[R].AAAAGLGHPPASPGGSED GPPGSEEDAAAR.[E]	7.67	High	1.67
Q07157	Tight junction protein 1	[R].KSREDLSAQPVQTK.[F]	3.82	High	1.61
O00512	B-cell CLL/lymphoma 9 protein	[R].IPVEGPLSPSRGDFPK.[G]	2.01	High	1.61
P53621	Coatomer subunit alpha	[K].KNLSPGAVESDVR.[G]	2.80	High	1.59
Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1	[K].GSKSPAKVSDGGSSSTDFK.[M]	3.04	High	1.59
P21127	Cyclin-dependent kinase 11B	[R].GTSRPPPEGGLGYSQLGDDDLK.[E]	4.40	High	1.52
P19338	Nucleolin	[K].GATPGKALVATPGKK.[G]	5.15	High	1.52
Q16594	Transcription initiation factor TFIID subunit 9	[R].LSVGSVTSRSPSTPTLGT PTPQTMVSSTK.[V]	3.32	High	1.52

After extensive bioinformatics-driven hypotheses are generated, a second key concept in analyzing MS results must include well thought-out ways to validate the quantitation of the phosphopeptides. We used western blot to demonstrate that the MS results were measurable on the protein level (Figure 7). Western blot analysis was also utilized to validate the functional analysis of HCC and HCC+V. For instance, western blot results supported our signaling pathway analysis (Table IVb) that suggested higher p-AKT expression in HCC+V. The different p-AKT signal for our two cell lines is likely dependent on specific upstream pathways mediated by this kinase. Literature search also provided another component of validation, as other scientific studies supported our hypotheses that deregulation of Rho/GTPase (19), insulin metabolic (20), and PI3K/AKT pathways (21-23) may be responsible for differential growth between HCC and HCC+V.

One particular protein of significant interest after extensive analysis is STMN1. STMN1 is a cytoskeletal protein that regulates microtubule dynamics for cell-cycle,

proliferation, and motility (24). Its tightly regulated phosphorylation/dephosphorylation controls cell cycle division and differentiation through inducing depolymerization from the microtubule plus end in a process referred to as “catastrophe” (25). Overexpressed in various cancers including prostate (26, 27), breast (28, 29), liver (30, 31), and oral squamous cell (32, 33), STMN1 is associated with increased proliferation, higher invasive potential, and chemo-resistance (34). Patients with endometrial tumors that highly express p-STMN S38 were also reported to have inferior prognosis (35). STMN1’s identification and quantitation on our MS data confirms literature and our hypothesis that this protein plays an important role in viral-mediated liver cancer transformation. MS data show STMN1 regulates HCC+V via all phosphosites (S16, S25, S38, S63) but most predominantly at S38 and S63 (Figure 8). However, this observation was not noted on western blot analysis. This discrepancy between antibody-mediated and mass spectrometry quantitation of protein levels may represent how MS technique has evolved over time to become more

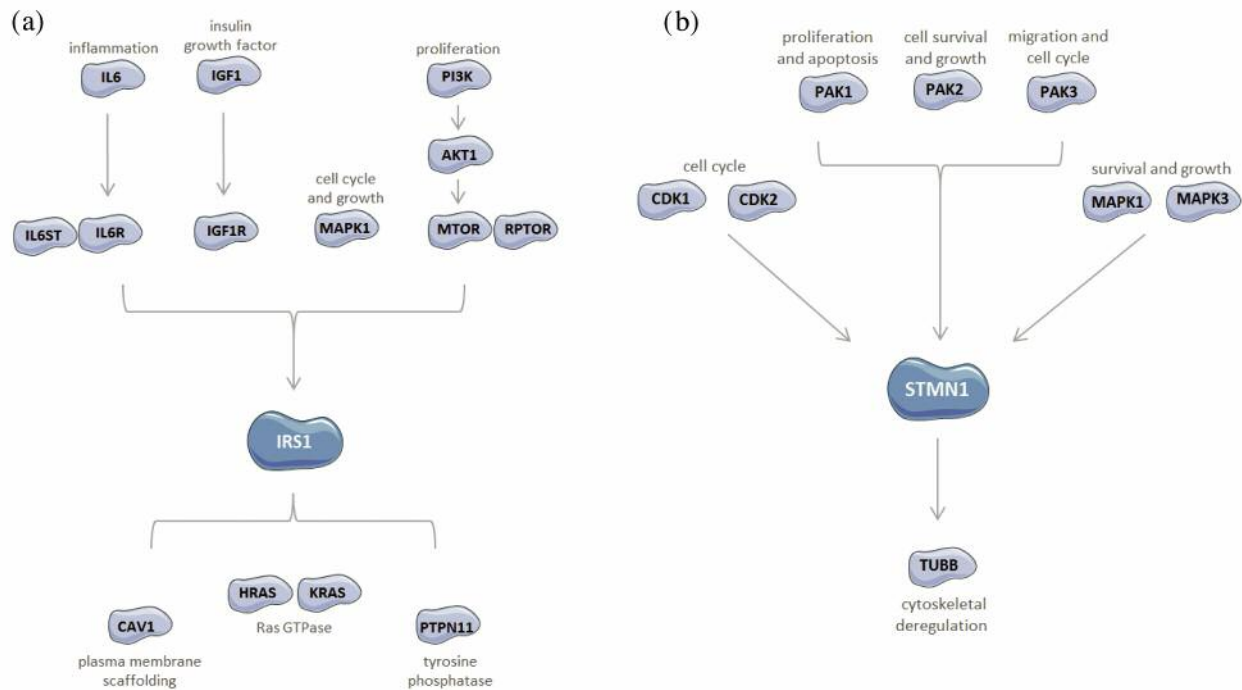


Figure 6. Hypothetical protein targets. Diagrams with associated upstream and downstream effectors for targets of interest selected from the phosphoprotein crosstalk networks. Pathways shown are for (a) the HCC abundant protein IRS1 and (b) the HCC+V abundant protein STMN1.

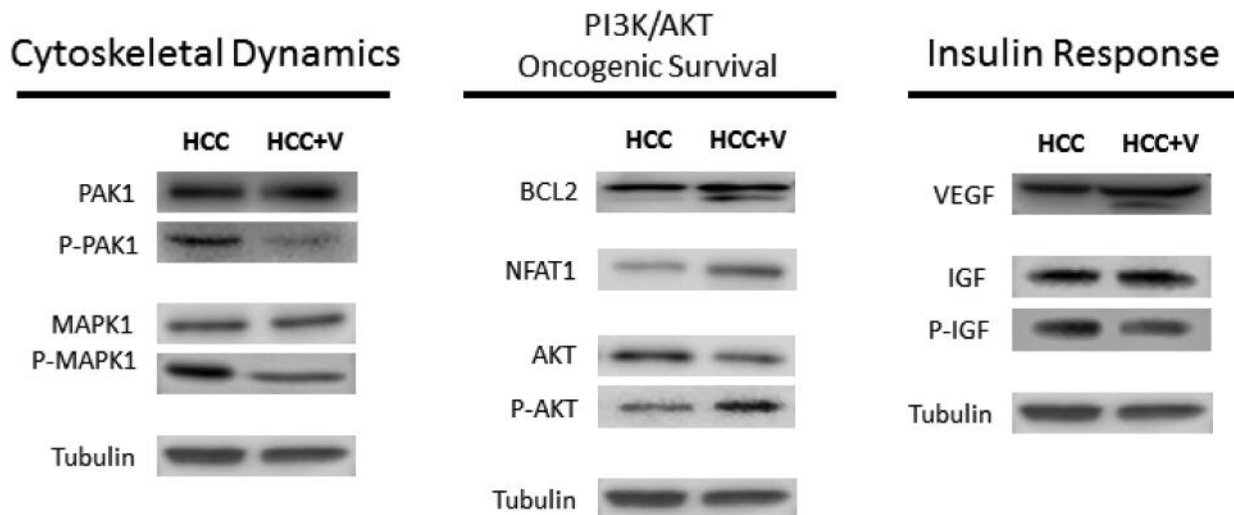


Figure 7. Western blot validation. Pathway and functional analysis validation using western blot with total protein and selective phospho-antibodies. Proteins were sorted into pathway categories based on literature search.

sensitive at detecting protein changes (36) . As a result, western blots should serve as a starting point for verification of MS quantitation while other techniques are used concurrently to explore biological/functional dynamics in depth.

Our strategy for the analysis of MS quantitative data is a simple function-focused workflow to investigate differences between two conditions, resulting in hypotheses for future study. These hypotheses can serve as a starting point on which additional experiments can build a more in-depth













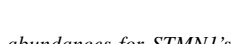

STMN1 Phosphosite	Phosphopeptide Sequence	Abundance (HCC+V)/(HCC)	HCC	HCC+V	
Ser16	[K].RASGQAFELILSPR.[S]	0.87			STMN
Ser25	[R].ASGQAFELILSPR.[S]	0.85			STMN S16
Ser38	[R].SKESVPEFPLSPPK.[K]	0.94			STMN S25
Ser38	[K].ESVPEFPLSPPK.[K]	1.28			STMN S38
Ser38	[K].ESVPEFPLSPPKKK.[D]	1.67			STMN S38
Ser63	[R].RKSHEAEVLK.[Q]	1.85			Tubulin

Figure 8. Stathmin phosphorylation comparison. The table shows phosphopeptide sequences and abundances for STMN1's four phosphorylation sites from MS results, as well as western blot analysis for total STMN1 and its phosphorylated forms.

understanding of the underlying mechanisms that promote HCC with or without virus. However, several factors should be taken into consideration in order to make the most out of future MS experiments. Quantitation of a cell line's phosphoproteome may be most useful in comparison to that cell line's global proteome, taken from either pre-published literature or a parallel MS experiment (37). In a similar manner, if comparing the phosphoproteomes of different conditions (*i.e.* drug treatment) it is indispensable to have a control condition as a baseline for both the quantitation results and subsequent functional analysis (38). Likewise, multiple biological replicates for phosphoproteomic analysis ensure greater confidence in MS results (39). One concern to address when studying PTM like phosphorylation is the transient nature of these modifications, that may require samples from multiple time points to further validate proteomic data. Another limitation to our study is that the HCC+V cell line's HCV replicon was engineered without certain viral elements that could affect infectivity (40). This may partially explain any discrepancies between the MS results utilizing a replicon in comparison to an unmodified HCV. Another challenge to contemplate is the differences between absolute and relative MS quantitation. In our experiment that compares only two cell lines, relative quantitation was appropriate. However, absolute quantitation may have more versatility in a broader range of experimental situations (41). Furthermore, the bioinformatics analysis of MS data relies on previously established databases, which can limit the scope and accuracy of our analysis. Lastly, proteomic analysis of patient samples could yield more rigorous data, especially when looking at the phosphoproteomic maps across different tissues.

While MS has many basic experimental uses, it also has promising applications in the clinical setting. Recent

Table IV. Phosphoprotein Crosstalk Signaling Pathways. List of the top signal transduction pathways effected by the phosphoprotein crosstalk networks with corresponding *P*-values (adjusted for multiple testing) for (IVa) HCC and (IVb) HCC+V.

#### (IVa) HCC Pathways

HCC Signaling pathways	<i>p</i> -Value
TRAIL signaling pathway	3.11E-33
Thrombin/protease-activated receptor (PAR) pathway	1.11E-32
PAR1-mediated thrombin signaling events	1.11E-32
IGF1 pathway	3.53E-32
IFN-gamma pathway	3.66E-32
ErbB1 downstream signaling	3.67E-32
Arf6 signaling events	3.67E-32
Arf6 downstream pathway	3.67E-32
Arf6 trafficking events	3.67E-32

#### (IVb) HCC+V Pathways

HCC+V Signaling pathways	<i>p</i> -Value
Proteoglycan syndecan-mediated signaling events	6.70E-17
TRAIL signaling pathway	6.70E-17
Glypican pathway	6.70E-17
ErbB receptor signaling network	6.70E-17
PAR1-mediated thrombin signaling events	6.79E-17
Signaling events mediated by VEGFR1 and VEGFR2	6.79E-17
VEGF and VEGFR signaling network	6.79E-17
LKB1 signaling events	6.79E-17
Thrombin/protease-activated receptor (PAR) pathway	6.79E-17
Class I PI3K signaling events mediated by Akt	8.30E-17

applications of this technique include utilizing MS for analysis of clinical patient samples and amassing this data into a resource for proteomic profiling to facilitate the

prediction of patient survival, prognosis, and effective therapeutics (42). Pioneered at larger cancer centers such as the MD Anderson Hospital, this method of clinical proteomic profiling is growing in popularity and is starting to be implemented at cancer treatment centers worldwide (43). Although more standardized models are necessary, MS proteomics could possibly be combined with metabolomics to create a sophisticated screen to assess for biomarkers pertaining to disease relapse or cancer incidence in high risk patients. In the future, MS could be utilized to evaluate the risk of HCC development in HCV patients by screening for HCC+V characteristic targets or differential protein aggregates. MS would have an advantage over current microarrays and genetic panels used for clinical screening as proteins and their post-translational modifications have closer biological proximity compared to DNA (44). Similarly, the protein abundance detected by MS might not correlate with mRNA/cDNA amounts used in microarray analysis (45). While challenges remain with processing MS extensive datasets and establishing standardized models, MS quantitation is a versatile technique for studying the proteome and has great potential to become an important player in the diagnosis and treatment of HCC. By harnessing the quantitative power of MS to use differential protein abundances for the analysis of function, pathway, and protein networks, one transforms a complex dataset into a meaningful platform to generate future hypotheses.

## Acknowledgements

This research utilized core support from the UCSD/UCLA NIDDK Diabetes Research Center P30DK063491. The Authors would like to thank Dr. Asim Dasgupta and Dr. Santanu Raychaudhuri (Department of Microbiology, Immunology, and Molecular Genetics; University of California Los Angeles) for providing us with the HCV replicon cell line. They thank Johnny Lin from the UCLA Statistical Consulting Center (Institute of Digital Research and Education; University of California Los Angeles) for his assistance with R software. They also thank Roxanna Khosravi and Dr. Jane Cox for editing assistance. Finally, they thank Sarah Bassilian for her help.

## References

- Whitmore SE and Lamont RJ: Tyrosine phosphorylation and bacterial virulence. *Int J Oral Sci* 4: 1-6, 2012.
- Scholz R, Imami K, Scott NE, Trimble WS, Foster LJ and Finlay BB: Novel Host Proteins and Signaling Pathways in Enteropathogenic *E. coli* Pathogenesis Identified by Global Phosphoproteome Analysis. *Mol Cell Proteomics* 14: 1927-1945, 2015.
- Becker TM, Boyd SC, Mijatov B, Gowrishankar K, Snoyman S, Pupo GM, Scolyer RA, Mann GJ, Kefford RF, Zhang XD and Rizos H: Mutant B-RAF-Mcl-1 survival signaling depends on the STAT3 transcription factor. *Oncogene* 33: 1158-1166, 2014.
- Bouzakri K, Karlsson HK, Vestergaard H, Madsbad S, Christiansen E and Zierath JR: IRS-1 serine phosphorylation and insulin resistance in skeletal muscle from pancreas transplant recipients. *Diabetes* 55: 785-791, 2006.
- Zappacosta F, Scott GF, Huddleston MJ and Annan RS: An optimized platform for hydrophilic interaction chromatography-immobilized metal affinity chromatography enables deep coverage of the rat liver phosphoproteome. *J Proteome Res* 14: 997-1009, 2015.
- Taus T, Kocher T, Pichler P, Paschke C, Schmidt A, Henrich C and Mechtler K: Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* 10: 5354-5362, 2011.
- Bosch FX, Ribes J, Diaz M and Cleries R: Primary liver cancer: worldwide incidence and trends. *Gastroenterology* 127: S5-S16, 2004.
- Wallace MC, Preen D, Jeffrey GP and Adams LA: The evolving epidemiology of hepatocellular carcinoma: a global perspective. *Expert Rev Gastroenterol Hepatol* 9: 765-779, 2015.
- Wilson-Grady JT, Haas W and Gygi SP: Quantitative comparison of the fasted and re-fed mouse liver phosphoproteomes using lower pH reductive dimethylation. *Methods* 61: 277-286, 2013.
- Rappsilber J, Mann M and Ishihama Y: Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2: 1896-1906, 2007.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
- Nita-Lazar A, Saito-Benz H and White FM: Quantitative phosphoproteomics by mass spectrometry: past, present, and future. *Proteomics* 8: 4433-4443, 2008.
- Macek B, Mann M and Olsen JV: Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* 49: 199-221, 2009.
- van der Mijn JC, Labots M, Piersma SR, Pham TV, Knol JC, Broxterman HJ, Verheul HM and Jimenez CR: Evaluation of different phospho-tyrosine antibodies for label-free phosphoproteomics. *J Proteomics* 127: 259-263, 2015.
- Humphrey SJ, Yang G, Yang P, Fazakerley DJ, Stockli J, Yang JY and James DE: Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. *Cell Metab* 17: 1009-1020, 2013.
- Nuhse TS, Bottrill AR, Jones AM and Peck SC: Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. *Plant J* 51: 931-940, 2007.
- Huang HL, Cendan CM, Roza C, Okuse K, Cramer R, Timms JF and Wood JN: Proteomic profiling of neuromas reveals alterations in protein composition and local protein synthesis in hyper-excitable nerves. *Mol Pain* 4: 33, 2008.
- Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, Berlin J, Baron A, Griffing S, Holmgren E, Ferrara N, Fyfe G, Rogers B, Ross R and Kabbinavar F: Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* 350: 2335-2342, 2004.
- Wong CM, Yam JW, Ching YP, Yau TO, Leung TH, Jin DY and Ng IO: Rho GTPase-activating protein deleted in liver cancer suppresses cell proliferation and invasion in hepatocellular carcinoma. *Cancer Res* 65: 8861-8868, 2005.

- 20 Tovar V, Alsinet C, Villanueva A, Hoshida Y, Chiang DY, Sole M, Thung S, Moyano S, Toffanin S, Minguez B, Cabellos L, Peix J, Schwartz M, Mazzaferro V, Bruix J and Llovet JM: IGF activation in a molecular subclass of hepatocellular carcinoma and pre-clinical efficacy of IGF-1R blockade. *J Hepatol* 52: 550-559, 2010.
- 21 Mannova P and Beretta L: Activation of the N-Ras-PI3K-Akt-mTOR pathway by hepatitis C virus: control of cell survival and viral replication. *J Virol* 79: 8742-8749, 2005.
- 22 Liu Z, Tian Y, Machida K, Lai MM, Luo G, Fong SK and Ou JH: Transient activation of the PI3K-AKT pathway by hepatitis C virus to enhance viral entry. *J Biol Chem* 287: 41922-41930, 2012.
- 23 Cheng D, Zhang L, Yang G, Zhao L, Peng F, Tian Y, Xiao X, Chung RT and Gong G: Hepatitis C virus NS5A drives a PTEN-PI3K/Akt feedback loop to support cell survival. *Liver Int* 35: 1682-1691, 2015.
- 24 Belletti B, Nicoloso MS, Schiappacassi M, Berton S, Lovat F, Wolf K, Canzonieri V, D'Andrea S, Zucchetto A, Friedl P, Colombatti A and Baldassarre G: Stathmin activity influences sarcoma cell shape, motility, and metastatic potential. *Mol Biol Cell* 19: 2003-2013, 2008.
- 25 Rubin CI and Atweh GF: The role of stathmin in the regulation of the cell cycle. *J Cell Biochem* 93: 242-250, 2004.
- 26 Friedrich B, Gronberg H, Landstrom M, Gullberg M and Bergh A: Differentiation-stage specific expression of oncoprotein 18 in human and rat prostatic adenocarcinoma. *Prostate* 27: 102-109, 1995.
- 27 Ghosh R, Gu G, Tillman E, Yuan J, Wang Y, Fazli L, Rennie PS and Kasper S: Increased expression and differential phosphorylation of stathmin may promote prostate cancer progression. *Prostate* 67: 1038-1052, 2007.
- 28 Curmi PA, Nogues C, Lachkar S, Carelle N, Gonthier MP, Sobel A, Lidereau R and Bieche I: Overexpression of stathmin in breast carcinomas points out to highly proliferative tumours. *Br J Cancer* 82: 142-150, 2000.
- 29 Golouh R, Cufer T, Sadikov A, Nussdorfer P, Usher PA, Brunner N, Schmitt M, Lesche R, Maier S, Timmermans M, Foekens JA and Martens JW: The prognostic value of Stathmin-1, S100A2, and SYK proteins in ER-positive primary breast cancer patients treated with adjuvant tamoxifen monotherapy: an immunohistochemical study. *Breast Cancer Res Treat* 110: 317-326, 2008.
- 30 Yuan RH, Jeng YM, Chen HL, Lai PL, Pan HW, Hsieh FJ, Lin CY, Lee PH and Hsu HC: Stathmin overexpression cooperates with p53 mutation and osteopontin overexpression, and is associated with tumour progression, early recurrence, and poor prognosis in hepatocellular carcinoma. *J Pathol* 209: 549-558, 2006.
- 31 Hsieh SY, Huang SF, Yu MC, Yeh TS, Chen TC, Lin YJ, Chang CJ, Sung CM, Lee YL and Hsu CY: Stathmin1 overexpression associated with polyploidy, tumor-cell invasion, early recurrence, and poor prognosis in human hepatoma. *Mol Carcinog* 49: 476-487, 2010.
- 32 Kouzu Y, Uzawa K, Koike H, Saito K, Nakashima D, Higo M, Endo Y, Kasamatsu A, Shiiba M, Bukawa H, Yokoe H and Tanzawa H: Overexpression of stathmin in oral squamous-cell carcinoma: correlation with tumour progression and poor prognosis. *Br J Cancer* 94: 717-723, 2006.
- 33 Harada K, Ferdous T, Harada T and Ueyama Y: High expression of stathmin 1 is a strong prognostic marker in oral squamous cell carcinoma patients treated by docetaxel-containing regimens. *Clin Exp Med* 2015.
- 34 Kavallaris M: Microtubules and resistance to tubulin-binding agents. *Nat Rev Cancer* 10: 194-204, 2010.
- 35 Wik E, Birkeland E, Trovik J, Werner HM, Hoivik EA, Mjos S, Krakstad C, Kusonmano K, Mauland K, Stefansson IM, Holst F, Petersen K, Oyan AM, Simon R, Kalland KH, Ricketts W, Akslen LA and Salvesen HB: High phospho-Stathmin(Serine38) expression identifies aggressive endometrial cancer and suggests an association with PI3K inhibition. *Clin Cancer Res* 19: 2331-2341, 2013.
- 36 Mann M: Can proteomics retire the western blot? *J Proteome Res* 7: 3065, 2008.
- 37 Wu R, Dephoure N, Haas W, Huttlin EL, Zhai B, Sowa ME and Gygi SP: Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. *Mol Cell Proteomics* 10: M111 009654, 2011.
- 38 Bennetzen MV, Marino G, Pultz D, Morselli E, Faergeman NJ, Kroemer G and Andersen JS: Phosphoproteomic analysis of cells treated with longevity-related autophagy inducers. *Cell Cycle* 11: 1827-1840, 2012.
- 39 Karp NA and Lilley KS: Design and analysis issues in quantitative proteomics studies. *Proteomics* 7(Suppl 1): 42-50, 2007.
- 40 Bartenschlager R: Hepatitis C virus replicons: potential role for drug development. *Nat Rev Drug Discov* 1: 911-916, 2002.
- 41 Kirkpatrick DS, Gerber SA and Gygi SP: The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* 35: 265-273, 2005.
- 42 Yuan Y, Van Allen EM, Omberg L, Wagle N, Amin-Mansour A, Sokolov A, Byers LA, Xu Y, Hess KR, Diao L, Han L, Huang X, Lawrence MS, Weinstein JN, Stuart JM, Mills GB, Garraway LA, Margolin AA, Getz G and Liang H: Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat Biotechnol* 32: 644-652, 2014.
- 43 MD Anderson named as one of two Genome Characterization Centers. MD Anderson News Release, 2015.
- 44 Witze ES, Old WM, Resing KA and Ahn NG: Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* 4: 798-806, 2007.
- 45 Margolin AA and Califano A: Theory and limitations of genetic network inference from microarray data. *Ann NY Acad Sci* 1115: 51-72, 2007.

Received April 2, 2016

Revised June 16, 2016

Accepted July 1, 2016