

RESEARCH ARTICLE

dBMHCC: A comprehensive hepatocellular carcinoma (HCC) biomarker database provides a reliable prediction system for novel HCC phosphorylated biomarkers

Yen-Wei Chu^{1,2,3,4,5}✉, Ching-Hsuan Chien^{1,5}✉, Mei-I Sung¹, Chi-Wei Chen^{1,6}, Yu-Ting Chen^{1,2,3,5}*

1 Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung, Taiwan, **2** Biotechnology Center, National Chung Hsing University, Taichung, Taiwan, **3** Agricultural Biotechnology Center, National Chung Hsing University, Taichung, Taiwan, **4** Institute of Molecular Biology, National Chung Hsing University, Taichung, Taiwan, **5** Ph.D. Program in Medical Biotechnology, National Chung Hsing University, Taichung, Taiwan, **6** Department of Computer Science and Engineering, National Chung-Hsing University, Taichung, Taiwan

✉ These authors contributed equally to this work.

* yuting@dragon.nchu.edu.tw



OPEN ACCESS

Citation: Chu Y-W, Chien C-H, Sung M-I, Chen C-W, Chen Y-T (2020) dBMHCC: A comprehensive hepatocellular carcinoma (HCC) biomarker database provides a reliable prediction system for novel HCC phosphorylated biomarkers. PLoS ONE 15(6): e0234084. <https://doi.org/10.1371/journal.pone.0234084>

Editor: Dilek Colak, King Faisal Specialist Hospital and Research Center, SAUDI ARABIA

Received: November 3, 2019

Accepted: May 18, 2020

Published: June 4, 2020

Copyright: © 2020 Chu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Database URL: <http://predictor.nchu.edu.tw/dBMHCC>.

Funding: This research was supported by (a) Ministry of Science and Technology, Taiwan, ROC under grant number MOST 106-2313-B-005-035-MY2, 108-2321-B-005-008 and 108-2634-F-005-002. This funding is supported for the salaries on research assistants and graduate students; (b) National Chung Hsing University and Chung-Shan Medical University under grant number NCHU-

Abstract

Hepatocellular carcinoma (HCC), which is associated with an absence of obvious symptoms and poor prognosis, is the second leading cause of cancer death worldwide. Genome-wide molecular biology studies should provide biological insights into HCC development. Based on the importance of phosphorylation for signal transduction, several protein kinase inhibitors have been developed that improve the survival of cancer patients. However, a comprehensive database of HCC-related phosphorylated biomarkers (HCCPMs) and novel HCCPMs prediction platform has been lacking. We have thus constructed the dBMHCC databases to provide expression profiles, phosphorylation and drug information, and evidence type; gathered information on HCC-related pathways and their involved genes as candidate HCC biomarkers; and established a system for evaluating protein phosphorylation and HCC-related biomarkers to improve the reliability of biomarker prediction. The resulting dBMHCC contains 611 notable HCC-related genes, 234 HCC-related pathways, 17 phosphorylation-related motifs and their 255 corresponding protein kinases, 5955 HCC biomarkers, and 1077 predicted HCCPMs. Methionine adenosyltransferase 2B (MAT2B) and acireductone dioxygenase 1 (ADI1), which regulate HCC development and hepatitis C virus infection, respectively, were among the top 10 HCCPMs predicted by dBMHCC. Platelet-derived growth factor receptor alpha (PDGFRA), which had the highest evaluation score, was identified as the target of one HCC drug (Regorafenib), five cancer drugs, and four non-cancer drugs. dBMHCC is an open resource for HCC phosphorylated biomarkers, which supports researchers investigating the development of HCC and designing novel diagnosis methods and drug treatments. **Database URL:** <http://predictor.nchu.edu.tw/dBMHCC>.

CSMU-10811. This funding is applied to research equipment, stationery, and printing fee. This funding is supported for English editing and publication fee. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No authors have competing interests.

Introduction

Liver cancer is the sixth most common cancer in the world but is the second leading cause of cancer death worldwide [1]. Most (70% to 90%) primary liver cancers are hepatocellular carcinoma (HCC), which presents with unnoticeable symptoms, resulting in a poor prognosis and a 5-year survival rate of only 12–15% [2]. To date, surgical procedures, radioactive particle implantation, radiofrequency ablation, hepatic artery chemoembolization, and chemotherapeutics are the most popular therapy strategies for patients in early-stage HCC. Recently, a few tyrosine kinase inhibitors have been developed as chemotherapeutic agents, of which sorafenib is the only systemic therapeutic approved by the FDA and has exhibited notable survival benefits for patients with HCC [3]. Genome-wide molecular biology studies will allow us to better understand the genetic mechanisms of liver cancer and provide new options for treatment.

Reversible post-translational modifications of proteins have a crucial role in cellular signal sensing and transduction networks, as they influence enzymatic activity and protein interactions, stability, and subcellular location. Recently, more than 200 post-translational modifications have been described, including the addition of functional groups (e.g., phosphorylation, acetylation), chemical modification of amino acids (e.g., citrullination, carbamylation), and structural changes (e.g., disulfide bridges, proteolytic cleavage) [4]. Among them, phosphorylation is the most common and well-studied, as one-third to one-half of all proteins in eukaryotic cells are estimated to be phosphorylated [5]. Protein phosphorylation, which is regulated by protein kinases and protein phosphatases, is involved in almost all cellular processes, including cellular energy balance, antioxidant mechanisms, cellular defense, mitochondrial biogenesis, growth, differentiation, metabolism, and apoptosis. Because of its key role in maintaining the cellular condition, a bifunctional protein kinase mediated signaling pathway may influence the development of many disease conditions, especially cancers, cardiovascular disease, neurodegenerative syndromes, and rare diseases [6]. Several protein kinase inhibitors do improve the survival and quality of life of cancer patients [7]. Hypervascularization is a key component of tumor growth and metastasis. The vascular endothelial growth factor (VEGF), basic fibroblast growth factors (bFGFs), platelet-derived growth factor (PDGF), and transforming growth factor beta (TGFB) are the major angiogenic factors. These growth factors along with their receptor tyrosine kinases (RTKs) and the related RAS/RAF/MEK/ERK and PI3K signaling pathways have become major targets for drug development to treat cancer [8]. That is, building a database of disease-related proteins and their target kinases should support the development of novel biomarkers for improved diagnosis accuracy and of new drug therapies in a more rapid, comprehensive, and efficient manner.

Recently, several HCC-related biomarkers have been predicted or confirmed by experiments. To organize this heterogeneous information, multiple HCC-related databases have been launched. OncoDB.HCC was the first complex cancer genomic database, integrating data from studies of genomic aberration, gene expression, and HCC model organisms [2]. OncoDB.HCC reported 614 HCC genes/proteins with significantly differential expression in HCC tissues [2]. The Encyclopedia of Hepatocellular Carcinoma genes Online (EHCO) systematically collected 2,906 HCC-related genes from PubMed, SAGE, microarray, and proteomics data and also found 40 of 120 evolutionarily conserved and overexpressed genes that can form a highly interactive network based on comparative genomics and interactomics analyses [9]. On a larger scale, 6927 HCC-related genes, including 143 gene signatures, and their supporting evidence—consisting of 98 microarray, proteome, and genomic studies are included in Liverome [10]. Liverome also provides functions for HCC-related gene searches and comparisons [10]. The database of prognostic biomarkers and models for HCC (dbPHCC)

Table 1. Comparison of dbMHCC and other HCC-related databases.

Characteristic	dbMHCC	dbPHCC	Liverome	EHCO	OncoDB.HCC
Year developed	2018	2016	2011	2007	2007
Organism	<i>Homo sapiens</i>	<i>Homo sapiens</i>	<i>Homo sapiens</i>	<i>Homo sapiens</i>	<i>Homo sapiens</i> and mouse
Database size	5955	567	6927	2906	613
HCC biomarkers	5068	567	143	2906	613
Experimental	611	567	143	2906	613
Predicted	4457	0	0	0	0
HCCPMs ^a	1280	0	0	0	0
Experimental	203	0	0	0	0
Predicted	1077	0	0	0	0
Covers	Protein and gene	Protein, gene, and microRNA	Protein and gene	Protein and gene	Protein and gene
Pathway information	✓	✓		✓	
Online prognosis analysis		✓			
Phosphorylation analysis	✓				
Drug information	✓	?			
Web service	✓	✓	✓		✓

^a HCCPMs: Hepatocellular carcinoma phosphorylated markers.

? dbPHCC mentioned that they provide patient sample information, clinical annotation, and prognostic information, drug information and relevant web link, but drug information was not displayed on the web site.

<https://doi.org/10.1371/journal.pone.0234084.t001>

contains 567 biomarkers (323 proteins, 154 genes, and 90 microRNAs) and their reference information, including experimental conditions and prognostic information [11]. These five databases mainly consist of high-throughput data, clinical annotations, and prognostic information, and some of them provide functional enrichment analyses, but they lack the ability to predict biomarkers and analyze phosphorylation (Table 1).

In this study, we constructed a comprehensive database of HCC biomarkers, dbMHCC, which includes a reliable prediction system for finding novel HCC-related phosphorylated biomarkers (HCCPMs). We also provide relevant medical information as a reference for future research and the development of treatment protocols. dbMHCC is freely available at <http://predictor.nchu.edu.tw/dbMHCC>.

Materials and methods

For providing a reliable prediction system to discover novel HCC-related biomarkers with phosphorylation, we constructed a comprehensive HCC biomarker database, dbMHCC (Fig 1). To achieve these approaches, we constructed six databases, HCC-611, GOCU, PhosMotif, SPPKinase, MKA, and Drug database, to gather expression profiles and phosphorylation and drug information. The MySQL has been used as database management tool for dbMHCC. For the prediction of novel HCC biomarker genes, we collected HCC-related pathways and their involved genes. We also established evaluation systems for protein phosphorylation and HCC-related biomarkers to improve the reliability of biomarker prediction. The details are described below.

HCC-611 database construction

The information on HCC genes in HCC-611 database was obtained from “Significant gene” entries of OncoDB.HCC [2]. After updating and universalizing each gene accession number (Ensembl gene ID) and protein accession number (RefSeq accession number and UniProtKB/SwissProt accession number), the unique HCC genes were collected. For further search

High resolution figures for review

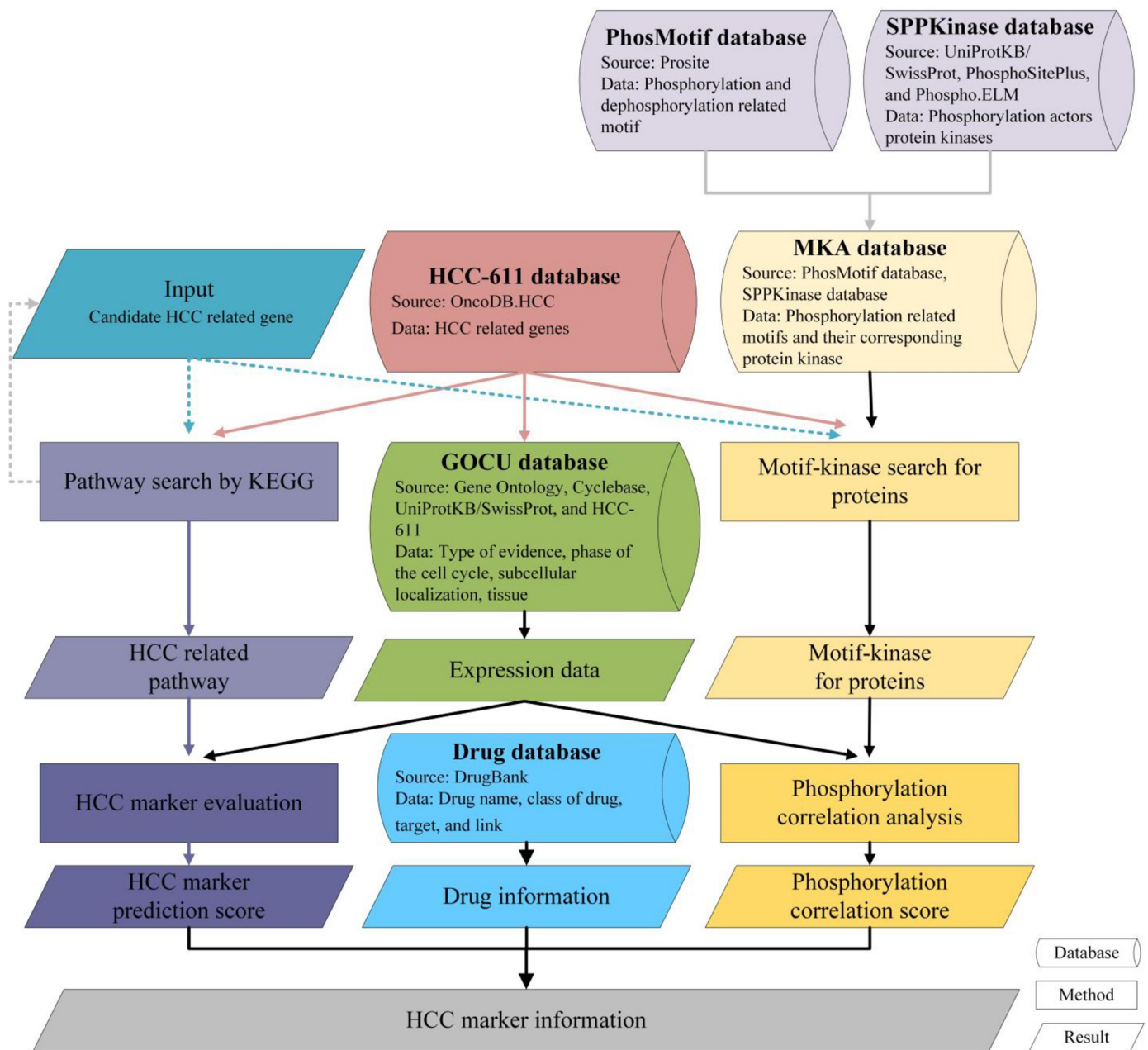


Fig 1. Flow chart of dbMHCC construction. The six developed databases are shown, along with the methods and resulting information used to generate the dbMHCC.

<https://doi.org/10.1371/journal.pone.0234084.g001>

functions, we incorporated the corresponding National Center for Biotechnology Information (NCBI) gene ID, gene name (Ensembl gene ID), and UniProtKB/SwissProt accession number in HCC-611.

MKA database construction

The MKA database, which integrates the information from PhosMotif and SPPKinase databases, provides phosphorylation-related motifs and their corresponding protein kinases. Based on sliding window, a comparison of each coding motif with consensus sequences is used to predict the phosphorylation function of the corresponding protein. Such motifs are good indicators for the prediction of corresponding kinases of novel HCC-related proteins. Therefore, we used “kinase OR phosph” and “phosphatase” as keywords for collecting the phosphorylation- and dephosphorylation-related motifs from Prosite [12], while constraining the search to human studies. SPPKinase database information was obtained from UniProtKB/SwissProt, PhosphoSitePlus [13, 14], and Phospho.ELM [15] phosphorylation databases. We first carried out a five-step screening process, consisting of (a) selecting human genes with experimental validation, (b) confirming the accession numbers of the related proteins, (c) finding the corresponding UniProtKB/SwissProt accession numbers for proteins collected from RefSeq, (d) judging the correctness of the protein sequences according to UniProtKB for proteins collected from RefSeq, and (e) confirming the accuracy of phosphorylation sites by Perl. The phosphorylation site should be screened for S, T, Y or H, after which the information about the corresponding kinases was presented in a uniform format.

GOCU database construction

The GOCU database provides gene expression information, including tissue-specific information, subcellular localization of the resulting protein, expression profile during the cell cycle, and the evidence type for this information. The evidence for HCC-related genes that was obtained from OncoDB.HCC and gathered from small- or large-scale experimental analyses was classified into type E1 or E2, respectively.

To investigate novel HCC-related genes, we isolated all genes that are expressed in liver, blood, or the lymphatic system from UniProtKB/SwissProt. They were classified into three types: type T1 genes were validated by protein-level experiments, type T2 genes were validated at the mRNA level, and type T3 genes were obtained from the “Sequence origin” section of UniProtKB/SwissProt, without additional specification. Their subcellular localization information was collected from Gene Ontology [16] and UniProtKB/SwissProt. Phase information was obtained from Cyclebase [17]. After evaluating their consistency and regularity, the experimental data of first cell cycle set was retained. To allow for a consistent time period across different experiments, we divided the cell cycle into 100 time points and calculated the expression level of the gene at each time point by interpolation.

Motif-kinase search for proteins

To predict the corresponding protein kinase, the amino acid sequences of target proteins obtained from UniProtKB were screened for phosphorylation-related motifs from the MKA database. The corresponding protein kinase of the characterized motif was defined as the corresponding protein kinase for the target protein.

Phosphorylation correlation analysis

Considering the need for co-expression and protein interaction between a catalytic kinase and its targeted protein, we imported data about tissue and subcellular localization and the expression phase from GOCU database to carry out a phosphorylation correlation analysis. The scoring method for the tissue and subcellular localization and phase factor data was as follows:

1. Scoring for tissue (S_t): Since this study was focused on Hepatocellular carcinoma, the genes which express in liver, blood, or the lymphatic system were collected for novel HCC-related genes prediction. In the reaction of protein phosphorylation, the catalytic kinase and its targeted protein process the phosphor group transferring in the form of protein. In order to assess the association between liver cancer-specific proteins and protein kinases, we classify proteins based on the credibility of the data source. If both genes with validation at the protein level show that they have highest chance to process the protein phosphorylation, thus the score is given as 10 (type T1). As one of them was validated at RNA level (type T2), that is, its expression in protein level was not confirmed, that reduced the credibility. Thus the score of type T2 was reduced as 6. The lowest credibility is "Sequence origin", thus if one of them was type T3 gene obtained from the "Sequence origin" section of UniProtKB/SwissProt, without additional specification (type T3), the score was as low as 2 left. In contrast, if one of the pair was not expressed in the liver, blood, or lymphatic system, there was no chance to interact and process the reaction, the score was 0.
2. Scoring for subcellular localization (S_l): The information of subcellular localization was obtained from Gene Ontology and UniProtKB/SwissProt. Compared to UniProtKB/SwissProt, GO provided more abundance data with clear validation evidence. dbMHCC obtained 6173 and 2709 data form GO and UniprotKB/SwissProt, respectively. Due to the imbalance of the data of these two databases, a two-layer process was built for this scoring system. The first layer consisted of an analysis of the GO term from Gene Ontology and was scored based on its validation evidence. As shown in S1 Table, a gene belongs to type L1, if it has experimental evidence; type L2, if it has computational analysis evidence; type L3, if its evidence consists of a curatorial statement; or type L4, if it was automatically assigned evidence. The credibility of the validation evidence decreases in accordance with the experimental, computational analysis, automatically assigned and automatically assigned evidence. If the query protein and protein kinase are both type L1 genes, the pair is scored a 10. If one of the pair belongs to type L2 and the other is type L1 or type L2, the score will be an 8. If one of the pair is type L3 and the other is not type L4, the score will be a 4. If either of the pair is type L4, the score will be a 2. If the query protein and protein kinase belong to type L3 or L4 or are without common GO terms, with a score lower than 8, the scoring system will start the second-layer processing. Based on the subcellular localization information from UniProtKB/SwissProt, if the query protein and protein kinase have the same subcellular localization, the score of the second layer will be a 6; if they do not, the final score is simply the score from the first layer.
3. Scoring for co-expression during a phase of the cell cycle (S_p): Co-expression is curial for protein-protein interaction, and also for the phosphor group transferring between kinase and its target protein. Thus, if the query protein and the corresponding protein kinase are both expressed at the same time point during the cell cycle, the score will be a 10; if they are not, the score will be a 0.

The kinase catalyzed phosphorylation correlation score (S_k) between the query protein and corresponding protein kinase is calculated as in Eq (1), and the highest S_k is selected as the kinase correlation score.

$$S_k = \frac{W_t S_t + W_l S_l + W_p S_p}{W_t + W_l + W_p} \quad (1)$$

W_t , W_l , and W_p are the weight of the tissue, subcellular localization, and phase, respectively. To evaluate whether the specific protein and protein kinase can really interact in the cell, the

environmental factors, including tissue and subcellular localization and the period of expression during the cell cycle. For HCC biomarker, expressing in specific tissues (liver, blood, or the lymphatic system) is the most important criteria, thus we set 5 (half of total weight) as weight of the tissue (W_t). The subcellular localization and the period of expression during the cell cycle were needed to be considered equally. While the data size of phase was less than the data size of subcellular localization, thus we set the weight of phase (W_p) as 2, and the weight of subcellular localization (W_l) as 3.

Curation of candidate HCC-related genes

We do not currently know all pathway(s) involved in the generation and development of HCC. Several studies have, however, shown that the mutation of HCC-related genes leads to a chain reaction that affects the normal physiological mechanisms and leads to the generation and development of HCC [18]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) [19] database constructed its network based on chemical genomics information, which can provide HCC-related pathways and all the genes involved in these pathways. To look for novel candidates for HCC marker genes, we found all the related pathways in the KEGG database using a search based on HCC-611 via DBGET [20], which we referred to collectively as the HCC-related pathways. All the genes involved in these HCC-related pathways then served as candidates for HCC-related genes.

HCC marker evaluation

HCC markers were evaluated based on the importance of their involved pathways, i.e., if more than one gene in a given pathway was experimentally confirmed to be associated with HCC. Based on the reliability of the evidence types of query gene, we developed a scoring system to assess the possibility of HCC markers as follows:

1. If the query gene belongs to HCC-611 with type E1 evidence, confirmed by wet-lab experimental data from previous reports, it has the highest credibility and its S_c will be set as 10. If the query gene with type E2 evidence, which is identified by microarray or proteomic reports, its S_c will be set as 8.
2. If the query gene was not experimentally confirmed as a HCC-associated gene, the other genes involved in the same pathway (OGSPs) are assessed. Each OGSP that is associated with type E1 evidence will be scored based on Eq (2).

$$S_c = W_1 \left(\frac{N_e + 0.8 N_m}{N_p} \right) + W_2 \quad (2)$$

Whereas if the OGSP was supported by only type E2 evidence, it will be scored according to Eq (3).

$$S_c = W_2 \left(\frac{N_m}{N_p} \right) \quad (3)$$

Where S_c is the score for a particular HCC marker in a single pathway; W_1 and W_2 are the weights of a given OGSP proven by type E1 and E2 evidence, respectively; N_m and N_e are the numbers of genes proven by type E2 and E1 evidence in that pathway, respectively; and N_p is the number of all genes involved in the pathway. According the reliability of evidence of OGSP and limited the query gene without experimental evidence to the highest score to 8, W_1 was given as 5, and W_2 as 3.

Based on the influence of all the pathways that the query gene is involved in, the HCC marker prediction score is calculated as the average of weights, which will be scored according to Eq (4).

$$S_{cm} = \sum_{c=1}^n S_c W(S_c) / \sum_{w=1}^n W(S_c) \quad (4)$$

Where S_{cm} is the score for HCC (cancer) marker prediction. $W(S_c)$ is the weight for S_c , and n is the number of pathways which the query gene involved in. If $S_c = 10$, then $W(S_c) = 10$; if $S_c = 8$, then $W(S_c) = 8$; if $5 \leq S_c < 8$, then $W(S_c) = 5$; and if $S_c < 5$, then $W(S_c) = 3$.

Drug information collection

To provide the related drug information, we used DrugBank 5.0.9 [21], which includes 10,505 drugs and 16,804 targets. A gene name was used with the Targets search in DrugBank. Because the name of a gene from different resources can differ, this will influence the accuracy of the search. Thus, we downloaded the Drug Bank 5.0.9 database and used the UniProtID as the basis for our search. In addition to HCC marker prediction and phosphorylation correlation evaluation, dBMHCC also provides the drug relation by Biomarker Prediction. User can use Gene Id, Accession Number or Gene Symbol as search term by Single Search function. While user use Gene Id as criteria, that may correspond to more than one protein. dBMHCC uses all the corresponding proteins as search criteria and provide their related drug information obtained from DrugBank database, which includes identification, pharmacological information, interactions, references, clinical trial information, pharmacoeconomic data, properties, spectra, taxonomy, and other details.

Results and discussion

HCC-611 database

HCC-611 database collected 609 protein-encoding genes and two non-protein-encoding genes from OncoDB.HCC, and HCC-611 provided the information, including chromosome location, gene name, NCBI gene ID, and UniProtKB/SwissProt accession number for these 611 significant HCC genes (Among them, 19 entries were extracted as examples shown in [S2 Table](#)).

GOCU database

The expression pattern of a gene is an important aspect of its ability to function as a reliable HCC-related biomarker. Thus, we constructed the GOCU database for further evaluation. Considering that different experimental methods may provide different credibility, the 611 significant HCC genes were classified based on different scales of experimental analyses. In HCC-611, 421 genes had supporting evidence from small-scale analyses and were classified as Type E1, and the other 190 genes were identified based on large-scale RNA analyses and were classified as Type E2.

To narrow down the screening base of tissue information, we focused on gene expression in particular tissues, i.e., liver, blood, or lymph system, from which we isolated 18,468 proteins from UniProtKB/SwissProt (Among them, 12 entries were extracted as examples shown in [S3 Table](#)). In addition, the subcellular localization information can indicate the possibility for protein-protein interactions in the microenvironment. We obtained 17,429 proteins with 1166 GO terms from Gene Ontology (Among them, 13 entries were extracted as examples shown in [S4 Table](#)) and 6233 proteins belonging to six categories of subcellular localization, including Secreted, Cell Membrane, ER/Golgi, Mitochondrion, Cytoplasm, and Nucleus, from

UniProtKB/SwissProt (Among them, 11 entries were extracted as examples shown in [S5 Table](#)). In addition, co-expression is also a key element for protein-protein interactions. We obtained 537 gene expression profiles from Cyclebase as a reference for co-expression. To allow consideration of uniform time periods from different experiments, we separated the cell cycle into 100 time points and calculated the expression level by interpolation ([S1 Fig](#)). The HCC evidence type and the co-expression profile relative to tissue and subcellular localization, and the phase of the cell cycle serve as the criteria for HCC marker and phosphorylation prediction to improve the reliability of dBMHCC.

MKA database

The MKA database, which integrates the PhosMotif and SPPKinase databases, provides information about phosphorylation-related motifs and their corresponding protein kinases. First, we collected 252 phosphorylation-related motifs from Prosite for PhosMotif database construction. Among them, 243 are related to phosphorylation, 32 are related to dephosphorylation, and 23 have dual functions. In the dBMHCC, the PhosMotif database provides the following information: Prosite Documentation Entry, Prosite Discriminator Entry, Entry Name, Description, Pattern, Max Repeat, Site, and Skip Flag. After a five-step screening process, data about a phosphorylation site and its corresponding protein kinases isolated from UniProtKB/SwissProt, Phospho.ELM, and PhosphoSitePlus were integrated into the SPPKinase database. Here information about 4161 experimentally identified phosphorylation sites from 1415 phosphorylated proteins and 348 corresponding protein kinases were collected and unified by using the UniProtKB/SwissProt accession number (Among them, 20 entries were extracted as examples shown in [S6 Table](#)). After integrating the PhosMotif and SPPKinase databases, 17 phosphorylation-related motifs were identified along with their corresponding 255 protein kinases in the MKA database ([S7 Table](#)). Among the 17 motifs, the bifunctional PS00004 and PS00018 motifs had 102 and 5 corresponding protein kinases, respectively. For the shorter patterns of the PS00004, PS00005, PS00006, and PS00007 motifs, a variety of corresponding protein kinases were identified. For example, there are 167 protein kinases that can target the PS00006 motif.

Based on the importance of protein phosphorylation in signal transduction and cellular responses, we proposed that protein phosphorylation should have a key function in HCC development. To identify the corresponding protein kinases for HCC-specific proteins, we used the genes from HCC-611 to search within MKA to find their corresponding protein kinases via motif screening. Among the HCC-611 entries, 13 phosphorylation-related motifs were found in 608 proteins, and 253 protein kinases were predicted to correspond to these 13 motifs. In the future, we may establish a BLAST tool in MKA to provide a corresponding protein kinase analysis of unknown proteins.

Correlation analysis of phosphorylation

With MKA, we can find phosphorylation-related motifs within HCC-specific protein sequences and can search for their corresponding protein kinases, but this association alone cannot prove that the specific protein and protein kinase can really interact in the cell. Therefore, some environmental factors, including tissue and subcellular localization and the period of expression during the cell cycle, were considered for phosphorylation correlation analysis. According to the validation evidences, we built a scoring method, Eq (1). Although the query protein may be associated with several protein kinases, the highest of the kinase correlation scores is treated as S_k . For the strongly suggestive ($S_k \geq 6.8$), it was given as $S_t \geq 6$, $S_l \geq 6$, and $S_p = 10$. That is the data of tissue should be validated at the protein or RNA level, the data of

subcellular localization should have experimental evidence from GO or annotated in UniProtKB/SwissProt, and the query protein and the corresponding protein kinase are both expressed at the same time point during the cell cycle. With such kinds of evidences, they are very likely to undergo a phosphorylation reaction. Considering the less data of expression profile in the phase of cell cycle, we ignore the phase criteria but keep the data with $S_i \geq 6$, and $S_l \geq 6$ as suggestive ($6.8 > S_k \geq 4.8$). When the S_k is less than 4.8, the evidences were not sufficient to support it. We obtained 71 genes that were strongly suggestive, 1373 genes that were suggestive, and 4511 genes that were not supported. There are 24.2% of HCC-related proteins with strongly suggestive or suggestive score, based on reliable experimental evidences.

HCC-related pathways

Based on HCC-611, we identified 391 genes involved in 234 pathways, which were then referred to as HCC-related pathways. Among these, the bladder cancer (hsa05219) pathway contains the most genes with a high level of supporting evidence: 24 genes with type E1 evidence among 38 genes in the pathway. The correlation between bladder and liver cancer development needs further investigation.

Given the complex nature of signaling networks, one gene could be involved in more than one pathway. In an extreme case, a single gene could be involved in several pathways and may function as a central hub in the development of cancer. The mitogen-activated protein kinases MAPK1 and MAPK3, which are each involved in 69 pathways (S8 Table), may be examples of this type of gene. MAPK1 and MAPK3 are the key components of MAPK signal transduction, which mediates diverse biological functions such as cell growth, adhesion, survival, and differentiation through the regulation of transcription, translation, and cytoskeletal rearrangements [22]. In particular, the MAPK1- and MAPK3-mediated Ras/MAPK pathway is significantly activated in human HCCs and is correlated with a poor prognosis [23]. Thus this kind of HCC-related gene may serve as a diagnosis indicator. As many HCC-related genes (337 of 611) are involved in <10 pathways (S9 Table), we believe that most HCC-related genes participate in specific pathways. Among these 337 HCC-related genes, 125 are involved in only one pathway. For example, the well-known oncogenic genes *NRG1*, *CD82*, *STMN1*, and *MCL1* are involved in the ErbB, p53, MAPK, and PI3K-Akt signaling pathways, respectively. In addition, *GPC3*, a member of heparan sulfate proteoglycans, has been identified as a biomarker for the diagnosis and prognosis of liver cancer [24].

HCC marker evaluation

All the genes involved in the HCC-related pathways identified above (5955 genes) were further evaluated as HCC markers based on their level of experimental confirmation. The HCC markers with type E1 or type E2 evidence are known biomarkers. To investigate the novel HCC marker, we evaluated the possibility according to the importance of their involved pathways. The more genes in a pathway are verified with more credible evidence to be related to liver cancer, the more important this pathway is. Thus, for suggestive, if the query gene was not experimentally confirmed as a HCC-associated gene, all the other genes involved in the same pathway should be associated with large scale experimental evidences E2 at least ($S_{cm} \geq 3$, calculated by Eq (4)). The query gene with $S_{cm} < 3$ will be noted as not supported. While 1/4 of total genes in the same pathway were associated with type E1 evidence, the query gene will be classified as strongly suggestive ($S_{cm} \geq 4.25$). We obtained 677 genes that were strongly suggestive, 4391 genes that were suggestive, and 887 genes that were not supported with respect to being a HCC marker. Based on GO enrichment and KEGG analysis, the most significant gene functions of HCC markers were identified as follows: 4834 HCC markers (11.4%) were

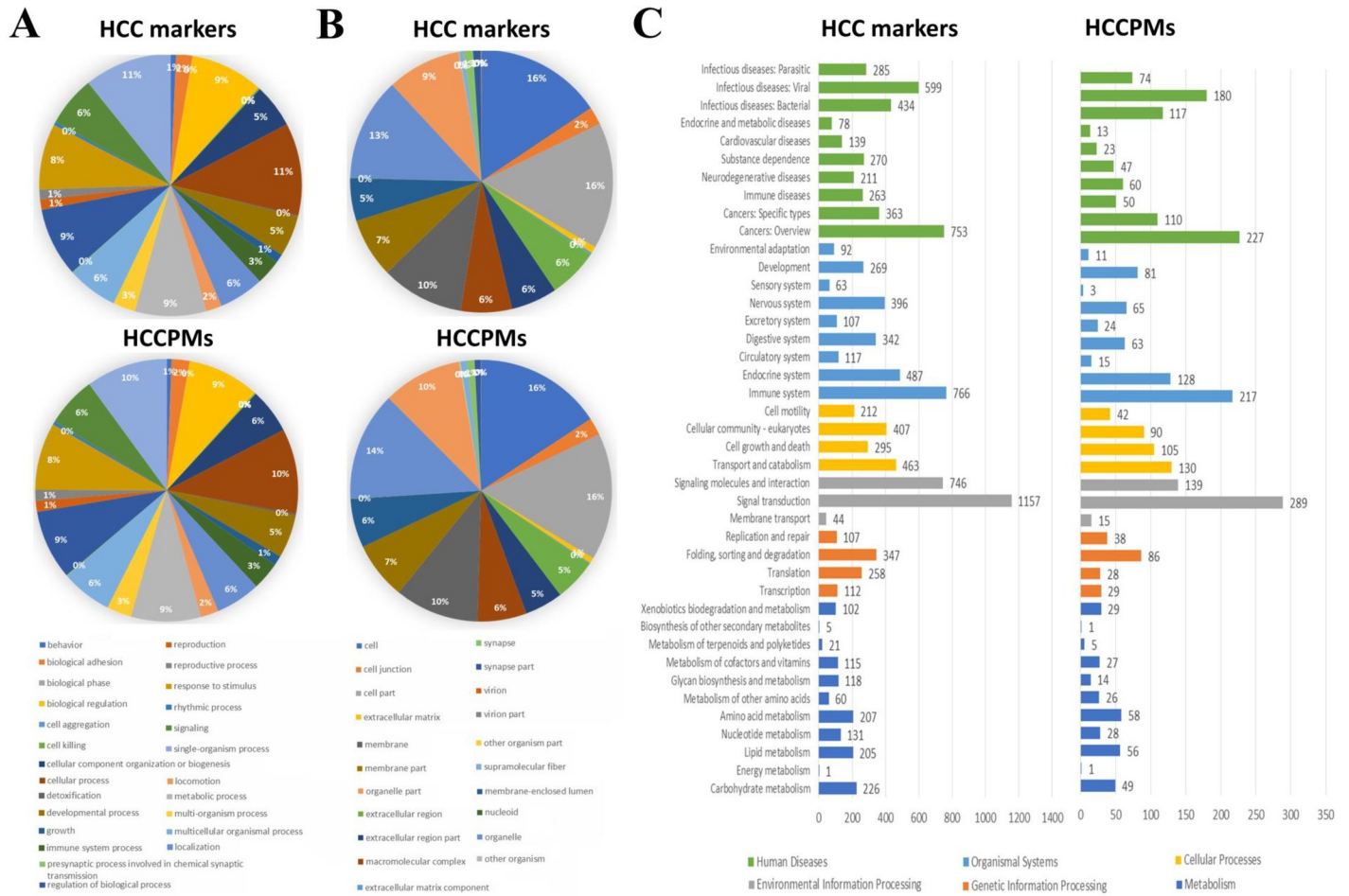


Fig 2. Functional analyses of HCC markers and HCCPMs. The distribution of HCC markers and HCCPMs functionally involved in Biological process (A) and Cellular component (B) was identified by GO enrichment analysis. (C) Based on KEGG analysis, the gene functions of HCC markers and HCCPMs were classified into Human Diseases, Organismal Systems, Cellular Processes, Environmental Information Processing, Genetic Information Processing, and Metabolism.

<https://doi.org/10.1371/journal.pone.0234084.g002>

involved in Biological process, 4799 (15.7%) were involved in Cellular component; 4441 (45.1%) were involved in Binding of Molecular function, and 1157 (14.3%) were involved in Signal transduction pathways (Fig 2). These results are consistent with the known mechanisms of cancer development.

In dbMHCC, the average score from the HCC marker evaluation process was 3.745 (out of 10), and the top nine scores are shown in Table 2. The top three genes in Table 2 are the death-associated protein kinases *DAPK3*, *DAPK1*, and *DAPK2*. DAPK, a tumor suppressor, is a positive mediator of the p38 mitogen-activated protein kinase inhibitor SB203580 and thyroid hormone induction of autophagy of HCC cells [25, 26]. Tropomyosin-receptor kinase (TRK)-fused gene, *TFG*, is a novel HCC-related gene, which has a role in controlling cell size and regulating apoptosis and cell proliferation by fusing with different partners [27]. For example, *TFG* can fuse with neurotrophic tyrosine kinase receptor 1 (NTRK1), neuron-derived orphan receptor 1 (NOR1), and anaplastic lymphoma receptor tyrosine kinase (ALK) involving in extraskeletal myxoid chondrosarcoma, anaplastic large cell lymphoma, inflammatory myofibroblastic tumors, lung cancer, and neuroblastoma [28, 29]. These findings may indicate that *TFG* has high potential as an HCC marker. In addition, matrix metalloproteinase 1, *MMP1*, is

Table 2. Top nine candidates of HCC marker genes selected based on S_{cm} score.

Gene ID	Gene Name	S_{cm}
1613	<i>DAPK3</i>	5.37
1612	<i>DAPK1</i>	5.37
23604	<i>DAPK2</i>	5.37
10342	<i>TFG</i>	4.69
8030	<i>CCDC6</i>	4.69
8031	<i>NCOA4</i>	4.69
4312	<i>MMP1</i>	4.64
1870	<i>E2F2</i>	4.58
1871	<i>E2F3</i>	4.58

<https://doi.org/10.1371/journal.pone.0234084.t002>

another novel HCC-related gene. MMP1 is involved in the breakdown of extracellular matrix during normal and disease processes, and has been identified as a crucial factor for invasion and metastasis in bladder cancer, stomach cancer, breast cancer, and esophageal squamous cell carcinoma [30–33]. According to these results, the top three predicted HCC markers are involved in HCC, and the following six predicted markers are involved in other cancers. That is, dbMHCC did provide a reliable HCC marker prediction tool.

HCC phosphorylated marker (HCCPM) evaluation

Phosphorylation and HCC relationship are two key factors of HCC phosphorylated marker (HCCPM) evaluation. Thus the scoring system was incorporated the evaluation scores of S_k and S_{cm} , which were calculated by Eqs (1) and (4), respectively. Based on their classification criteria, if the query proteins were considered to be strongly suggestive in both correlation analysis of phosphorylation ($S_k \geq 6.8$) and HCC marker evaluation ($S_{cm} \geq 4.25$), it will be considered strongly suggestive in HCCPM evaluation. If the query proteins were considered to be not supported in both correlation analysis of phosphorylation ($S_k < 4.8$) and HCC marker evaluation ($S_{cm} < 3$), it will be considered to be not supported in HCCPM evaluation. While the scores were between the levels of strongly suggestive and not supported in HCCPM evaluation, the query protein will be considered to be suggestive. We obtained 19 genes that were strongly suggestive, 1261 genes that were suggestive, and 4675 genes that were not supported with respect to the evaluation scores S_k and S_{cm} . Based on GO enrichment and KEGG analysis, the most significant gene functions of HCCPMs were identified as follows: 1241 HCCPMs (10.5%) were involved in Biological process, 1269 (15.8%) were involved in Cellular component, 1176 (47.3%) were involved in Binding of Molecular function, and 289 (14.4%) were involved in Signal transduction pathways (Fig 2). These results are consistent with the functional classification of HCC markers described above.

The top 10 HCCPMs are listed in Table 3. FK506 binding protein 51 (FKBP51, also called FKBP5), an immunophilin, has roles in signaling regulation, tumorigenesis, and chemoresistance [34]. FKBP5 is highly expressed in multiple tissues, including kidney, skeletal muscle, liver, placenta, heart, and peripheral blood. FKBP5 can stimulate the formation of a FKBP5-Hsp90-p23 superchaperone complex, which promotes cell growth in prostate cancer via androgen receptor-mediated signaling [35]. In contrast, downregulation of FKBP5 can promote apoptosis of lymphoblastic leukemia and irradiated melanoma cells through the activation of the transcription factor nuclear factor- κ B (NF- κ B) [34]. Considering that NF- κ B regulates multiple essential functions in hepatocellular injury, liver fibrosis, and HCC [36], in addition to high expression of FKBP5 in the liver [37], we propose that FKBP5 may play a key role in HCC development through NF- κ B signaling.

Table 3. Top 10 candidates for HCCPMs selected based on phosphorylation correlation analysis (S_k) and pathway evaluation (S_{cm}).

Gene ID	Accession Number ^a	Gene Name	S_k	S_{cm}
2289	Q13451	<i>FKBP5</i>	8.00	4.29
7175	P12270	<i>TPR</i>	7.40	4.26
8030	Q16204	<i>CCDC6</i>	6.00	4.69
83667	P58004	<i>SESN2</i>	6.00	4.53
55367	Q9HB75	<i>PIDD1</i>	6.00	4.53
64065	Q96FX8	<i>PERP</i>	6.00	4.53
55256	Q9BV57	<i>ADII</i>	6.00	4.49
27430	Q9NZL9	<i>MAT2B</i>	6.00	4.49
2069	O14944	<i>EREG</i>	6.00	4.42
9252	O75582	<i>RPS6KA5</i>	6.00	4.36

^a UniprotKB/SwissProt Accession Number.

<https://doi.org/10.1371/journal.pone.0234084.t003>

The second-highest candidate, TPR, a nucleoporin, was named based on its translocated promoter region to the kinase domain of the proto-oncogene MET. TPR translocation with MET is associated with an early step during carcinogenesis in gastric carcinoma [38]. TPR also undergoes translocation with NTrk1 (TrkA), a process that is associated with papillary thyroid carcinoma [39].

The third-highest candidate, CCDC6, participates in the function of the PTC1/ret proto-oncogene and is associated with papillary thyroid carcinomas and lung cancers [40, 41]. Phosphorylation of CCDC6, mediated by ataxia-telangiectasia mutated (ATM) kinase, prevents the FBXW7-CCDC6 interaction and FBXW7-mediated CCDC6 degradation during the DNA damage response [42]. Considering the high frequency of gene rearrangements of CCDC6 in several cancers, and the S_k and S_{cm} evaluation in dBMHCC, CCDC6 is a target gene worthy of further study with respect to its association with HCC.

Although acireductone dioxygenase 1 (*ADII*) has not previously been shown to be a HCC-related gene, Chang et al. [43] found that overexpression of ADII regulates the infection of hepatitis C virus (HCV). HCV is well known as a leading etiology of HCC, and thus ADII might serve as a biomarker of HCV-mediated HCC.

Finally, methionine adenosyltransferase 2B (*MAT2B*) regulates HCC development. *MAT2B* interacts with *GIT1* to increase ERK activity and growth in human liver and colon cancer cells [44]. According to these results, dBMHCC predicted several reliable HCCPMs. Based on these HCCPMs, dBMHCC can also provide their corresponding kinases, which could be used as the target for developing protein kinase inhibitors to improve survival and quality of life of cancer patients (7).

Drug information

To provide HCCPM-related drug information, we collected 10,505 drugs and 16,804 targets from DrugBank 5.0.9 [2]. Among the HCCPMs, 9 genes were the targets of 5 liver cancer drugs, 91 genes were the targets of 124 cancer drugs, and 364 genes were the targets of 1515 non-cancer drugs. Platelet-derived growth factor receptor alpha, PDGFRA, which had the highest evaluation scores ($S_k = 8$, $S_{cm} = 10$) in dBMHCC, was the target of one HCC drug (Regorafenib), five cancer drugs (Imatinib, Sunitinib, XL820, Olaratumab, and Amuvatinib), and four non-cancer drugs (Becaplermin, Pazopanib, Midostaurin, and Ponatinib). Cyclin-dependent kinase 6, CD6, which was highly suggestive as a HCC marker in dBMHCC ($S_k = 6$,

$S_{cm} = 4.21$), was the target of four cancer drugs (Alvocidib, Palbociclib, Ribociclib, and Abemaciclib), and two non-cancer drugs ((2S)-2-({6-[(3-amino-5-chlorophenyl)amino]-9-isopropyl-9H-purin-2-yl}amino)-3-methylbutan-1-ol, and 3,7,3',4'-tetrahydroxyflavone). Receptor-type tyrosine-protein kinase FLT3 can be used for the diagnosis and the determination of optimal therapy for myeloid leukemia [45]. In dBMHCC, FLT3 was predicted as an HCCPM ($S_k = 6.2$, $S_{cm} = 3.697$) and was the target of one liver cancer drug (Sorafenib), three cancer drugs (Sunitinib, Amuvatinib, and Tandutinib), and six non-cancer drugs (ABT-869, Midostaurin, Ponatinib, Brigatinib, Quizartinib, and Nintedanib). In addition, those cancer drugs that target HCCPMs, can be investigated with dBMHCC to determine their feasibility with respect to HCC treatment, which should effectively shorten the time needed for new HCC cancer drug development.

Web interface

dBMHCC provides a user-friendly web interface that integrates HCC phosphor marker-related information and a prediction platform. Databases and Biomarker Prediction are the two main functional units in dBMHCC.

In the Databases unit (Fig 3), the user can use the Link tool to retrieve information from the HCC-611, MKA, GOCU, and Drug databases (Fig 3A). The individual pages for the HCC-611 and SPPKinase databases are shown (Fig 3B and 3C), each of which includes a search function and provides the results from such searches. The other MKA database, PhosMotif, provides Prosite Documentation, Prosite Discriminator, Motif Name, Description, Motif Pattern, Max Repeat, Phosphorylation Site, and Skip Flag of Phosphorylation related Motif information based on a search query. Four GOCU-related results can be generated, HCC Evidence Type and the Tissue, Subcellular Localization, and Phase expression information (Fig 3D–3G). From the Drug database page, the user can use the Search tool to investigate the corresponding drug via the UniProtKB/SwissProt accession number of a protein of interest (Fig 3H).

In the Biomarker Prediction unit (Fig 4), both a Single and Conditional Search are available. The user can use the NCBI Gene ID, UniProtKB/SwissProt Accession Number, or the Gene Symbol for a Single Search (Fig 4A). The quick screen showing the overall information will help the user decide whether to carry out further searches or not (Fig 4B), and the detailed information for the Phosphorylation Correlation Analysis (Fig 4C), HCC Marker Evaluation (Fig 4D), and Drug Information (Fig 4E) pages will be displayed in sequence. For additional detailed information, the user can click on the Motif Entry value from the Phosphorylation Correlation Analysis page to obtain the Matched Pattern in the query protein and detailed information about that motif and its corresponding kinase. Clicking the Kinase Accession Number, the user will get the co-expression information about the target gene and corresponding kinase, such as the tissue, subcellular localization, and phase expression data, to investigate the possibility of phosphorylation. On the HCC Marker Prediction page, the user can click the Pathway Entry value to get detailed information about the HCC-related pathway, including the Pathway Name and all the participating genes. If the gene has been experimentally identified as an HCC-related gene, it will be shown in red. On the Drug Information page, clicking the Drug name, will take the user to the corresponding DrugBank entry for detailed information. To benefit systemic analyses, dBMHCC also provides a Conditional Search. The user can set the range for the Phosphorylation Correlation or HCC Marker Prediction Score or select the Drug classification to obtain the proteins that meet the screening criteria (Fig 4F). The resulting Accession Number, Gene Symbol, Phosphorylation Correlation Score, HCC Marker Prediction Score, and Drug classification data are then listed on the result page (Fig 4G). Clicking an Accession Number from this result page results in a Single Search.



Fig 3. Web interface of search tools and results in the Databases unit of dbMHCC. (A) Web portal of the Databases unit in dbMHCC. dbMHCC contains the HCC-611, MKA, GOCU, and Drug databases and their sub-databases. Users can click Link to download the database information and click Search to investigate the related information for a query protein. (B–H) Web interfaces for the search tools and results from the HCC-611 database (B); the SPPKinase database (C); the type of HCC evidence (D), type of evidence for tissue-specific expression data (E), subcellular localization (F), and phase-specific expression during the cell cycle (G) from the GOCU database; and the Drug database (H).

<https://doi.org/10.1371/journal.pone.0234084.g003>

Conclusion

We have constructed dbMHCC, a comprehensive HCC biomarker database with a reliable prediction system for novel HCC phosphorylated biomarkers. dbMHCC contains 611 HCC significant genes, 234 HCC-related pathways, 17 phosphorylation-related motifs and their 255

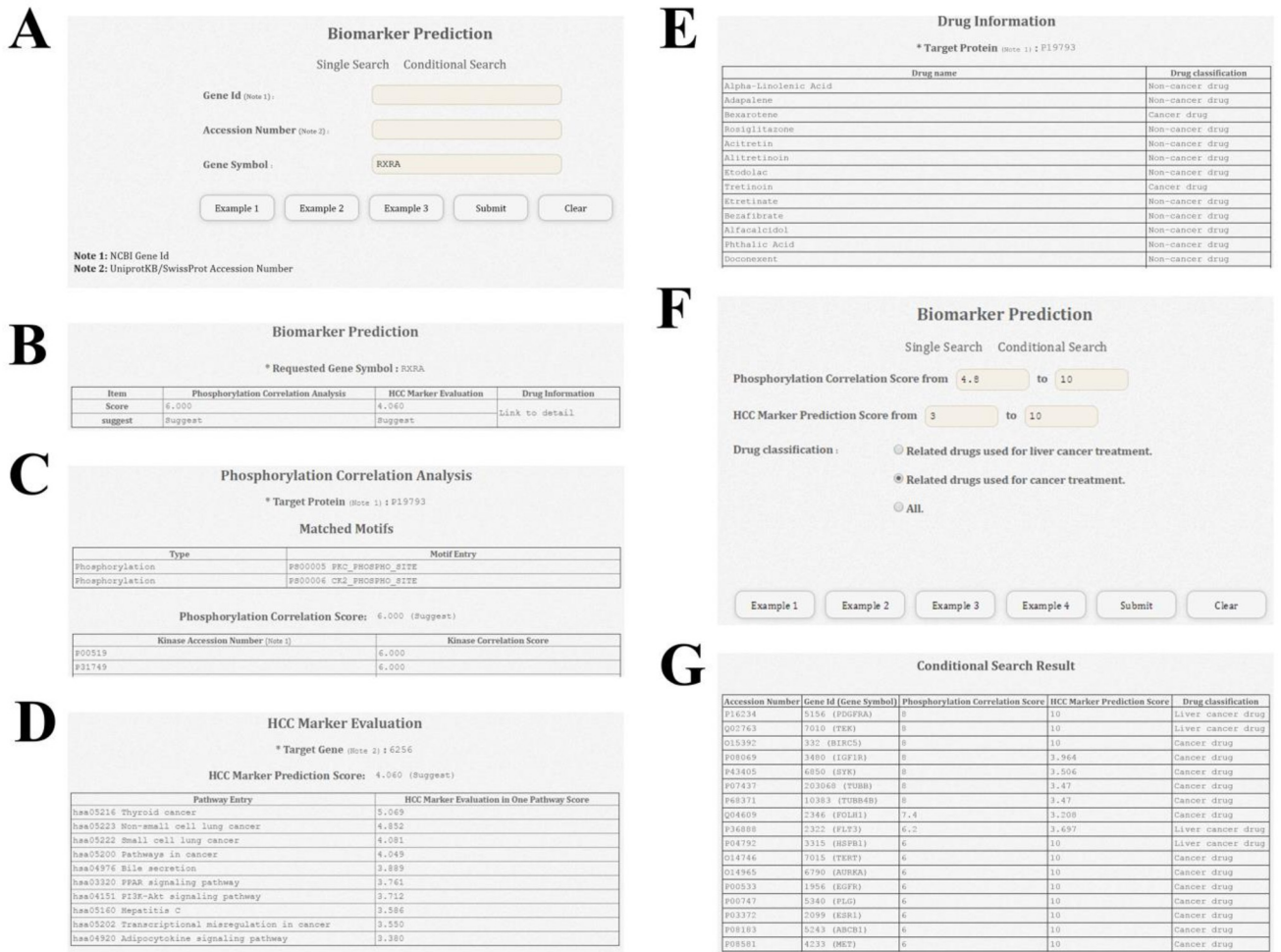


Fig 4. Web interface of the Biomarker Prediction unit of dBMHCC. (A) The dBMHCC website provides a Single Search and Conditional Search for HCCPMs. (B–E) The quick screen of overall information from the Biomarker Prediction (B) is shown first, with detailed information from the Phosphorylation Correlation Analysis (C), HCC Marker Evaluation (D), and Drug Information (E) displayed subsequently in sequence. The Conditional Search for Biomarker Prediction allows the use of up to three screening criteria (F). Based on a Conditional Search, dBMHCC will provide Accession Number, Gene ID, Gene Symbol, Phosphorylation Correlation Score, HCC Marker Prediction Score, and Drug classification of the qualified proteins (G). Clicking on the Accession Number, Gene ID, or Gene Symbol from the Conditional Search Result, will provide detailed information resulting from a Single Search.

<https://doi.org/10.1371/journal.pone.0234084.g004>

corresponding protein kinases, 5955 HCC biomarkers, and 1077 predicted HCCPMs. Among the HCCPMs, 9 genes were the targets of 5 liver cancer drugs, 91 genes were the targets of 124 cancer drugs, and 364 genes were the targets of 1515 non-cancer drugs. The cancer drugs that target HCCPMs should be investigated for their feasibility in HCC treatment, which should effectively shorten the time needed to develop new HCC cancer drugs. In conclusion, dBMHCC is a free and open resource HCC biomarker database, which provides expression profiles, evidence type, and drug information, to support researchers investigating the development of HCC and designing novel HCC-related diagnostics and drug treatment.

Supporting information

S1 Fig. Gene expression information during phases of the cell cycle as provided by the GOCU database. It shows the GOCU-generated information for one entry (A0AVK6). The

phase length, time scale, and expression level were given by Cyclebase (17). The phase length of cell cycle was determined by expression profiles of human periodic genes. The common time scale was chosen to be in percent of the cell division cycle with 100% corresponding to cytokinesis (M/G1-transition). The expression level was presented as \log_2 (Cy5-labeled cDNA from synchronous cells/Cy3-labeled cDNA from asynchronously growing HeLa cells), where (Cy5/Cy3) is the normalized ratio of the background-corrected intensities.

(PDF)

S1 Table. Classification of subcellular localization evidence information obtained from Gene Ontology.

(PDF)

S2 Table. Information about HCC-related genes from the HCC-611 database.

(PDF)

S3 Table. Classification of expression data for tissue-specific expression of various genes.

(PDF)

S4 Table. Subcellular localization information obtained from Gene Ontology.

(PDF)

S5 Table. Subcellular localization information obtained from UniProtKB/SwissProt.

(PDF)

S6 Table. Information about the corresponding kinase and phosphorylation site for a query protein.

(PDF)

S7 Table. Number of corresponding kinases for each phosphorylation-related motif.

(PDF)

S8 Table. Top 13 genes involved in multiple pathways.

(PDF)

S9 Table. The number of pathways in which HCC-related genes are involved.

(PDF)

Author Contributions

Conceptualization: Yen-Wei Chu, Yu-Ting Chen.

Data curation: Ching-Hsuan Chien, Mei-I Sung, Chi-Wei Chen.

Funding acquisition: Yen-Wei Chu, Yu-Ting Chen.

Methodology: Ching-Hsuan Chien, Mei-I Sung, Chi-Wei Chen.

Supervision: Yen-Wei Chu, Yu-Ting Chen.

Writing – original draft: Yen-Wei Chu, Yu-Ting Chen.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*. 2015; 65(2):87–108. Epub 2015/02/06. <https://doi.org/10.3322/caac.21262> PMID: 25651787.
2. Su WH, Chao CC, Yeh SH, Chen DS, Chen PJ, Jou YS. OncoDB.HCC: an integrated oncogenomic database of hepatocellular carcinoma revealed aberrant cancer target genes and loci. *Nucleic Acids*

- Res. 2007; 35(Database issue):D727–31. Epub 2006/11/14. <https://doi.org/10.1093/nar/gkl845> PMID: 17098932; PubMed Central PMCID: PMC1669730.
3. Sun W, Cabrera R. Systemic Treatment of Patients with Advanced, Unresectable Hepatocellular Carcinoma: Emergence of Therapies. *Journal of gastrointestinal cancer*. 2018. Epub 2018/02/18. <https://doi.org/10.1007/s12029-018-0065-8> PMID: 29453759.
 4. Minguez P, Parca L, Diella F, Mende DR, Kumar R, Helmer-Citterich M, et al. Deciphering a global network of functionally associated post-translational modifications. *Mol Syst Biol*. 2012; 8:599. Epub 2012/07/19. <https://doi.org/10.1038/msb.2012.31> PMID: 22806145; PubMed Central PMCID: PMC3421446.
 5. Hubbard MJ, Cohen P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem Sci*. 1993; 18(5):172–7. Epub 1993/05/01. [https://doi.org/10.1016/0968-0004\(93\)90109-z](https://doi.org/10.1016/0968-0004(93)90109-z) PMID: 8392229.
 6. Cotan D, Paz MV, Alcocer-Gomez E, Garrido-Maraver J, Oropesa-Avila M, de la Mata M, et al. AMPK As A Target in Rare Diseases. *Curr Drug Targets*. 2016; 17(8):921–31. Epub 2016/01/14. <https://doi.org/10.2174/1389450117666160112110204> PMID: 26758671.
 7. Eglen R, Reisine T. Drug discovery and the human kinome: recent trends. *Pharmacol Ther*. 2011; 130(2):144–56. Epub 2011/01/25. <https://doi.org/10.1016/j.pharmthera.2011.01.007> PMID: 21256157.
 8. Nguyen K, Jack K, Sun W. Hepatocellular Carcinoma: Past and Future of Molecular Target Therapy. *Diseases (Basel, Switzerland)*. 2015; 4(1). Epub 2015/12/25. <https://doi.org/10.3390/diseases4010001> PMID: 28933381; PubMed Central PMCID: PMC5456309.
 9. Hsu CN, Lai JM, Liu CH, Tseng HH, Lin CY, Lin KT, et al. Detection of the inferred interaction network in hepatocellular carcinoma from EHCO (Encyclopedia of Hepatocellular Carcinoma genes Online). *BMC Bioinformatics*. 2007; 8:66. Epub 2007/03/01. <https://doi.org/10.1186/1471-2105-8-66> PMID: 17326819; PubMed Central PMCID: PMC1828168.
 10. Lee L, Wang K, Li G, Xie Z, Wang Y, Xu J, et al. Liverome: a curated database of liver cancer-related gene signatures with self-contained context information. *BMC Genomics*. 2011; 12 Suppl 3:S3. Epub 2012/03/06. <https://doi.org/10.1186/1471-2164-12-s3-s3> PMID: 22369201; PubMed Central PMCID: PMC3333186.
 11. Ouyang J, Sun Y, Li W, Zhang W, Wang D, Liu X, et al. dbPHCC: a database of prognostic biomarkers for hepatocellular carcinoma that provides online prognostic modeling. *Biochim Biophys Acta*. 2016; 1860(11 Pt B):2688–95. Epub 2016/03/05. <https://doi.org/10.1016/j.bbagen.2016.02.017> PMID: 26940364.
 12. Sigrist CJ, de Castro E, Cerutti L, Cuče BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res*. 2013; 41(Database issue):D344–7. Epub 2012/11/20. <https://doi.org/10.1093/nar/gks1067> PMID: 23161676; PubMed Central PMCID: PMC3531220.
 13. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*. 2015; 43(Database issue):D512–20. Epub 2014/12/18. <https://doi.org/10.1093/nar/gku1267> PMID: 25514926; PubMed Central PMCID: PMC4383998.
 14. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res*. 2012; 40(Database issue):D261–70. Epub 2011/12/03. <https://doi.org/10.1093/nar/gkr1122> PMID: 22135298; PubMed Central PMCID: PMC3245126.
 15. Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, et al. Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res*. 2011; 39(Database issue):D261–7. Epub 2010/11/11. <https://doi.org/10.1093/nar/gkq1104> PMID: 21062810; PubMed Central PMCID: PMC3013696.
 16. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004; 32(Database issue):D258–61. Epub 2003/12/19. <https://doi.org/10.1093/nar/gkh036> PMID: 14681407; PubMed Central PMCID: PMC308770.
 17. Gauthier NP, Jensen LJ, Wernersson R, Brunak S, Jensen TS. Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results. *Nucleic Acids Res*. 2010; 38(Database issue):D699–702. Epub 2009/11/26. <https://doi.org/10.1093/nar/gkp1044> PMID: 19934261; PubMed Central PMCID: PMC2808877.
 18. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. *Gastroenterology*. 2017; 152(4):745–61. Epub 2017/01/04. <https://doi.org/10.1053/j.gastro.2016.11.048> PMID: 28043904.
 19. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 2012; 40(Database issue):D109–14. Epub 2011/11/15. <https://doi.org/10.1093/nar/gkr988> PMID: 22080510; PubMed Central PMCID: PMC3245020.

20. Fujibuchi W, Goto S, Migimatsu H, Uchiyama I, Ogiwara A, Akiyama Y, et al. DBGET/LinkDB: an integrated database retrieval system. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*. 1998;683–94. Epub 1998/08/11. PMID: [9697222](#).
21. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018; 46(D1):D1074–d82. Epub 2017/11/11. <https://doi.org/10.1093/nar/gkx1037> PMID: [29126136](#); PubMed Central PMCID: [PMC5753335](#).
22. Upadhy D, Ogata M, Reneker LW. MAPK1 is required for establishing the pattern of cell proliferation and for cell survival during lens development. *Development (Cambridge, England)*. 2013; 140(7):1573–82. <https://doi.org/10.1242/dev.081042> PMID: [23482492](#).
23. Delire B, Starkel P. The Ras/MAPK pathway and hepatocarcinoma: pathogenesis and therapeutic implications. *European journal of clinical investigation*. 2015; 45(6):609–23. Epub 2015/04/03. <https://doi.org/10.1111/eci.12441> PMID: [25832714](#).
24. Wu Y, Liu H, Ding H. GPC-3 in hepatocellular carcinoma: current perspectives. *Journal of hepatocellular carcinoma*. 2016; 3:63–7. Epub 2016/11/24. <https://doi.org/10.2147/JHC.S116513> PMID: [27878117](#); PubMed Central PMCID: [PMC5108608](#).
25. Zhang H, Chen GG, Zhang Z, Chun S, Leung BC, Lai PB. Induction of autophagy in hepatocellular carcinoma cells by SB203580 requires activation of AMPK and DAPK but not p38 MAPK. *Apoptosis: an international journal on programmed cell death*. 2012; 17(4):325–34. Epub 2011/12/16. <https://doi.org/10.1007/s10495-011-0685-y> PMID: [22170404](#).
26. Chi HC, Chen SL, Tsai CY, Chuang WY, Huang YH, Tsai MM, et al. Thyroid hormone suppresses hepatocarcinogenesis via DAPK2 and SQSTM1-dependent selective autophagy. *Autophagy*. 2016; 12(12):2271–85. Epub 2016/09/23. <https://doi.org/10.1080/15548627.2016.1230583> PMID: [27653365](#); PubMed Central PMCID: [PMC5173272](#).
27. Chen Y, Tseng SH. Targeting tropomyosin-receptor kinase fused gene in cancer. *Anticancer Res*. 2014; 34(4):1595–600. Epub 2014/04/03. PMID: [24692687](#).
28. Greco A, Mariani C, Miranda C, Lupas A, Pagliardini S, Pomati M, et al. The DNA rearrangement that generates the TRK-T3 oncogene involves a novel gene on chromosome 3 whose product has a potential coiled-coil domain. *Mol Cell Biol*. 1995; 15(11):6118–27. Epub 1995/11/01. <https://doi.org/10.1128/mcb.15.11.6118> PMID: [7565764](#); PubMed Central PMCID: [PMC230863](#).
29. Armstrong F, Duplantier MM, Trempat P, Hieblot C, Lamant L, Espinos E, et al. Differential effects of X-ALK fusion proteins on proliferation, transformation, and invasion properties of NIH3T3 cells. *Oncogene*. 2004; 23(36):6071–82. Epub 2004/06/23. <https://doi.org/10.1038/sj.onc.1207813> PMID: [15208656](#).
30. Xuan J, Zhang Y, Zhang X, Hu F. Matrix metalloproteinase-1 expression in breast cancer and cancer-adjacent tissues by immunohistochemical staining. *Biomedical reports*. 2015; 3(3):395–7. Epub 2015/07/03. <https://doi.org/10.3892/br.2015.420> PMID: [26137243](#); PubMed Central PMCID: [PMC4467224](#).
31. Wiczorek E, Wasowicz W, Gromadzinska J, Reszka E. Functional polymorphisms in the matrix metalloproteinase genes and their association with bladder cancer risk and recurrence: a mini-review. *International journal of urology: official journal of the Japanese Urological Association*. 2014; 21(8):744–52. Epub 2014/03/19. <https://doi.org/10.1111/iju.12431> PMID: [24635493](#).
32. Liu M, Hu Y, Zhang MF, Luo KJ, Xie XY, Wen J, et al. MMP1 promotes tumor growth and metastasis in esophageal squamous cell carcinoma. *Cancer Lett*. 2016; 377(1):97–104. Epub 2016/05/01. <https://doi.org/10.1016/j.canlet.2016.04.034> PMID: [27130665](#).
33. Dey S, Ghosh N, Saha D, Kesh K, Gupta A, Swarnakar S. Matrix metalloproteinase-1 (MMP-1) Promoter polymorphisms are well linked with lower stomach tumor formation in eastern Indian population. *PloS one*. 2014; 9(2):e88040. Epub 2014/02/08. <https://doi.org/10.1371/journal.pone.0088040> PMID: [24505369](#); PubMed Central PMCID: [PMC3914871](#).
34. Li L, Lou Z, Wang L. The role of FKBP5 in cancer aetiology and chemoresistance. *Br J Cancer*. 2011; 104(1):19–23. Epub 2010/12/02. <https://doi.org/10.1038/sj.bjc.6606014> PMID: [21119664](#); PubMed Central PMCID: [PMC3039800](#).
35. Ni L, Yang CS, Gioeli D, Frierson H, Toft DO, Paschal BM. FKBP51 promotes assembly of the Hsp90 chaperone complex and regulates androgen receptor signaling in prostate cancer cells. *Mol Cell Biol*. 2010; 30(5):1243–53. Epub 2010/01/06. <https://doi.org/10.1128/MCB.01891-08> PMID: [20048054](#); PubMed Central PMCID: [PMC2820886](#).
36. Luedde T, Schwabe RF. NF-kappaB in the liver—linking injury, fibrosis and hepatocellular carcinoma. *Nature reviews Gastroenterology & hepatology*. 2011; 8(2):108–18. Epub 2011/02/05. <https://doi.org/10.1038/nrgastro.2010.213> PMID: [21293511](#); PubMed Central PMCID: [PMC3295539](#).
37. Baughman G, Wiederrecht GJ, Chang F, Martin MM, Bourgeois S. Tissue distribution and abundance of human FKBP51, and FK506-binding protein that can mediate calcineurin inhibition. *Biochem Biophys*

- Res Commun. 1997; 232(2):437–43. Epub 1997/03/17. <https://doi.org/10.1006/bbrc.1997.6307> PMID: 9125197.
38. Soman NR, Correa P, Ruiz BA, Wogan GN. The TPR-MET oncogenic rearrangement is present and expressed in human gastric carcinoma and precursor lesions. *Proc Natl Acad Sci U S A*. 1991; 88(11):4892–6. Epub 1991/06/01. <https://doi.org/10.1073/pnas.88.11.4892> PMID: 2052572; PubMed Central PMCID: PMC51773.
 39. Xu S, Powers MA. Nuclear pore proteins and cancer. *Semin Cell Dev Biol*. 2009; 20(5):620–30. Epub 2009/07/07. <https://doi.org/10.1016/j.semcdb.2009.03.003> PMID: 19577736; PubMed Central PMCID: PMC2706781.
 40. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, et al. RET, ROS1 and ALK fusions in lung cancer. *Nat Med*. 2012; 18(3):378–81. Epub 2012/02/14. <https://doi.org/10.1038/nm.2658> PMID: 22327623.
 41. Schwaller J, Anastasiadou E, Cain D, Kutok J, Wojiski S, Williams IR, et al. H4(D10S170), a gene frequently rearranged in papillary thyroid carcinoma, is fused to the platelet-derived growth factor receptor beta gene in atypical chronic myeloid leukemia with t(5;10)(q33;q22). *Blood*. 2001; 97(12):3910–8. Epub 2001/06/05. <https://doi.org/10.1182/blood.v97.12.3910> PMID: 11389034.
 42. Zhao J, Tang J, Men W, Ren K. FBXW7-mediated degradation of CCDC6 is impaired by ATM during DNA damage response in lung cancer cells. *FEBS Lett*. 2012; 586(24):4257–63. Epub 2012/10/31. <https://doi.org/10.1016/j.febslet.2012.10.029> PMID: 23108047.
 43. Chang ML, Huang YH, Cheng JC, Yeh CT. Interaction between hepatic membrane type 1 matrix metalloproteinase and acireductone dioxygenase 1 regulates hepatitis C virus infection. *J Viral Hepat*. 2016; 23(4):256–66. Epub 2015/11/06. <https://doi.org/10.1111/jvh.12486> PMID: 26537061.
 44. Peng H, Li TW, Yang H, Moyer MP, Mato JM, Lu SC. Methionine adenosyltransferase 2B-GIT1 complex serves as a scaffold to regulate Ras/Raf/MEK1/2 activity in human liver and colon cancer cells. *The American journal of pathology*. 2015; 185(4):1135–44. Epub 2015/03/22. <https://doi.org/10.1016/j.ajpath.2014.12.016> PMID: 25794709; PubMed Central PMCID: PMC4380842.
 45. Sexauer AN, Tasian SK. Targeting FLT3 Signaling in Childhood Acute Myeloid Leukemia. *Frontiers in pediatrics*. 2017; 5:248. Epub 2017/12/07. <https://doi.org/10.3389/fped.2017.00248> PMID: 29209600; PubMed Central PMCID: PMC5702014.