Integrated genomic analysis for prediction of survival for patients with liver cancer using The Cancer Genome Atlas

Yan-Zhou Song, Xu Li, Wei Li, Zhong Wang, Kai Li, Fang-Liang Xie, Feng Zhang

Abstract

AIM
To evaluate the prognostic power of different molecular data in liver cancer.

METHODS
Cox regression screen and least absolute shrinkage and selection operator were performed to select significant prognostic variables. Then the concordance index was calculated to evaluate the prognostic power. For the combination data, based on the clinical cox model, molecular features that better fit the model were combined to calculate the concordance index. Prognostic models were built based on the arithmetic summation of the significant variables. Kaplan-Meier survival curve and log-rank test were performed to compare the survival difference. Then a heatmap was constructed and gene set enrichment analysis was performed for pathway analysis.

Institutional review board statement: The study was reviewed and approved by the Institutional Review Board of Nanjing Medical University.

Conflict-of-interest statement: The authors declare no conflict of interest.

Data sharing statement: The data in this manuscript was accessible through https://portal.gdc.cancer.gov/ or http://gdac.broadinstitute.org/

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Manuscript source: Unsolicited manuscript

Correspondence to: Feng Zhang, MD, Professor, Department of Liver Surgery/Liver Transplantation Center, The First Affiliated Hospital of Nanjing Medical University; 300 Guangzhou Rd, Nanjing 210029, Jiangsu Province, China. doctorzhangfeng@njmu.edu.cn

Received March 21, 2018
Peer-review started: March 22, 2018
First decision: April 24, 2018
Revised: June 13, 2018
Accepted: June 25, 2018
Article in press: June 25, 2018
Published online: July 28, 2018
Liver cancer is the fourth most common digestive cancer, with the second highest cancer mortality rate worldwide. Every year, there are approximately 750,000 new cases, and most of these patients died of liver cancer. Due to tumor heterogeneity and patients’ physical status, the prognosis of liver cancer varied across different patients, with the average 5-year survival rate of 26%.

Prognostic markers can help make better clinical decisions by selecting patients who respond well to some specific treatment. Currently, there are several commonly used clinical markers for liver cancer, such as alphafetoprotein (AFP), patient age, tumor stage and some scoring systems. In addition, some genetic biomarkers are emerging as novel indicators in cancer diagnosis and prognosis, such as GPC3, DKK1, S100A4, S100A14, SOX6, SUOX, xCT, GRK6, et al. The Cancer Genome Atlas (TCGA) is funded by the National Institute of Health (NIH) to describe the genomic alterations across different cancer types. It provides tremendous amount of “omics” data, including sequencing data, miRNA sequencing data, reverse phase protein arrays data, copy number change data and DNA sequencing data. In 2017, the comprehensive genomic characteristics of liver cancer have also been analyzed. It included unsupervised clustering of five molecular platforms to identify the hepatocellular carcinoma patients associated with poor prognosis.

However, there is no consensus on the predictive power of these indicators, especially the molecular markers. Therefore, by utilizing the TCGA data, we aimed to evaluate the prognostic power of liver cancer by molecular markers, and also to assess the predictive power of liver cancer by combining molecular markers and clinical data.

RESULTS
The mRNA data were the most informative prognostic variables in all kinds of omics data in liver cancer, with the highest concordance index (C-index) of 0.61. For the copy number variation, methylation and miRNA data, the combination of molecular data with clinical data could significantly boost the prediction accuracy of the molecular data alone (P < 0.05). On the other hand, the combination of clinical data with methylation, miRNA and mRNA data could significantly boost the prediction accuracy of the clinical data itself (P < 0.05). Based on the significant prognostic variables, different prognostic models were built. In addition, the heatmap analysis, survival analysis, and gene set enrichment analysis validated the practicability of the prognostic models.

CONCLUSION
In all kinds of omics data in liver cancer, the mRNA data might be the most informative prognostic variable. The combination of clinical data with molecular data might be the future direction for cancer prognosis and prediction.

Key words: Liver cancer; Prognosis; Molecular marker; Evaluation; C-index

© The Author(s) 2018. Published by Baishideng Publishing Group Inc. All rights reserved.

INTRODUCTION
Liver cancer is the fourth most common digestive cancer, with the second highest cancer mortality rate worldwide. Every year, there are approximately 750,000 new cases, and most of these patients died of liver cancer. Due to tumor heterogeneity and patients’ physical status, the prognosis of liver cancer varied across different patients, with the average 5-year survival rate of 26%.

Prognostic markers can help make better clinical decisions by selecting patients who respond well to some specific treatment. Currently, there are several commonly used clinical markers for liver cancer, such as alphafetoprotein (AFP), patient age, tumor stage and some scoring systems. In addition, some genetic biomarkers are emerging as novel indicators in cancer diagnosis and prognosis, such as GPC3, DKK1, S100A4, S100A14, SOX6, SUOX, xCT, GRK6, et al.

The Cancer Genome Atlas (TCGA) is funded by the National Institute of Health (NIH) to describe the genomic alterations across different cancer types. It provides tremendous amount of “omics” data, including sequencing data, miRNA sequencing data, reverse phase protein arrays data, copy number change data and DNA sequencing data. In 2017, the comprehensive genomic characteristics of liver cancer have also been analyzed. It included unsupervised clustering of five molecular platforms to identify the hepatocellular carcinoma patients associated with poor prognosis.

However, there is no consensus on the predictive power of these indicators, especially the molecular markers. Therefore, by utilizing the TCGA data, we aimed to evaluate the prognostic power of liver cancer by molecular markers, and also to assess the predictive power of liver cancer by combining molecular markers and clinical data.

MATERIALS AND METHODS

Data collection and processing
The clinical data and level 3 molecular data, including copy number variation (CNV), methylation, mRNA, miRNA and protein data of the liver cancer patients were downloaded from the TCGA repository (http://gdac.broadinstitute.org/). For the clinical data, we included the patient age and tumor stage, which were the easiest accessible information. For the CNV data, we adopted the Affymetrix genome-wide human SNP array 6.0 platform to detect the copy number variation. After we got the level 3 segmentation file of the copy number data, the GISTIC version 2.0.22 was executed to detect significant regions of amplification and deletion. For the methylation data, the Illumina DNA methylation 450 platform level 3 data was utilized. With respect to the miRNA and mRNA data, the Illumina HiSeq miRNASeq and Illumina HiSeq mRNASeq level 3 data were adopted. For the protein data, we applied the reverse phase protein array (RPPA) data platform.

Concordance index calculation
To evaluate the prognostic power among different omics data, we built a core set of samples. The patients in the core sample set were able to provide complete information from clinical data and different molecular platforms, including CNV, methylation, mRNA, miRNA and protein data. The concordance index (c-index) was calculated according to the method suggested by Yuan et al. Briefly, training group patients (80% of total patients) and testing group patients (20% of total patients) were selected by the stratified random sampling method. The prognostic models were evaluated according to the method suggested by Yuan et al. The predicted survival of the core sample set was obtained by the model, and the heatmap analysis, survival analysis, and gene set enrichment analysis validated the practicability of the prognostic models.
To establish the prognostic models, we performed the cox regression analysis and LASSO for all the cancer patients (including both training group and testing group) with complete data. After the significant prognostic variables were identified, we selected the significant variables to build the scoring model. To simplify the process and make it easier to apply in clinic, if the variable value in one patient was higher than the median value, we deemed it to be positive, with the score 1 or -1. Otherwise, the score was 0. The score of 1 or -1 was determined by the corresponding coefficient in the cox regression analysis. If the coefficient was positive, the score was 1. If the coefficient was negative, the score was -1. Normally, the higher the score was, the higher the risk of poor prognosis. When we divided the patients into the high risk group and low risk group based on the prognostic score, the Kaplan-Meier survival curve was performed to evaluate the overall survival. At the same time, the log-rank test was applied to evaluate the survival difference.

**Heatmap construction and gene set enrichment analysis**

With respect to the gene expression analysis, the R package "limma" was utilized to detect the differentially expressed genes. Afterwards, the top 100 highly expressed genes and top 100 lowly expressed genes were selected to build the heatmap using the R package "gplots". For the pathway analysis, we performed gene set enrichment analysis (GSEA) proposed by the Broad Institute (http://broadinstitute.org/gsea/downloads.jsp), with the gene sets downloaded from the MSigDB collections (software.broadinstitute.org/gsea/msigdb/collections.jsp).

**RESULTS**

**Characteristics of the TCGA samples and molecular platform information**

Data of the liver cancer patients were downloaded from the TCGA repository (http://gdc.broadinstitute.org/). The platform information and sample size for each data type are shown in Table 1. The mean age of the liver cancer patients was 59 years old. They were 68% male and 32% female. In all of the included patients, 50% were in stage I, 25% were in stage II, 24% were in stage III, and the other 1% were in stage IV. To evaluate the prognostic power among the different omics data, we also built a core sample set of 171 patients, which only included patients with complete data of all five omics platforms, namely, copy number variation (CNV), methylation, mRNA, miRNA and protein level.

**Prognostic power of clinical data and different types of molecular data**

The concordance index (C-index) calculated from clinical data and each type of molecular data are shown in Figure 2. The average C-index of clinical data, CNV, methylation, mRNA, and protein level was 0.56, 0.51, 0.57, 0.58, 0.61 and 0.57, respectively. For the molecular data alone, the mRNA data seemed to be the most informative predictors among all molecular data, with the highest C-index of 0.61. The C-index calculated from the

Table 1 Characteristics of the Cancer Genome Atlas samples and molecular platform information

|               | CNV         | Methylation | mRNA     | miRNA     | Protein |
|---------------|-------------|-------------|----------|-----------|---------|
| Platform info | SNP, 6      | 450K        | HiSeq    | HiSeq     | RPPA    |
| Sample size   | 371         | 377         | 371      | 372       | 184     |
a similar method to the copy number variation model, we built the prognostic model based on the methylation data (Figure 4A). Patients with different survival could be distinguished by the prognostic model, as shown in the Kaplan-Meier survival curve (Figure 4B). With respect to the mechanisms leading to the different outcomes, limma and GSEA were performed to show the different gene expression pattern and enriched pathways between the high risk patients and low risk patients (Figure 4C and D). The E2F targets, G2M checkpoint, Myc targets V1, spermatogenesis and PI3K/AKT/mTOR pathway signaling were among the top five enriched pathways in the high risk group of patients (Figure 4D).

Establishment of the prognostic model based on the miRNA data
In the establishment of the miRNA model, there were 372 patients with complete miRNA data and clinical data in the analysis. The cox regression and LASSO were performed to screen the significant prognostic miRNAs. Results showed that miR-3690, miR-561 and miR-621 were independent prognostic factors for the liver cancer patients. Based on these three markers, the prognostic model was built, as shown in Figure 5A. Depending on the scores calculated from the miRNA model, the patients were divided into the high risk group and low risk group. As shown in the Kaplan-Meier survival curve, there was a significant survival difference between these two group patients (Figure 5B). According to the heatmap, it shows distinct gene expression patterns between these group patients (Figure 5C). In addition, GSEA showed that WNT/beta-catenin, G2M checkpoint, allograft rejection, mitotic spindle, and inflammatory response were among the top five enriched pathways in the high risk group of patients (Figure 5D).

Establishment of the prognostic model based on the mRNA data
In this section, we included 371 patients with complete mRNA data and clinical data to build the prognostic model based on the mRNA data. The concordance index (C-index) value on the left Y-axis indicated the prognostic power of each data type. The \( P \) values on the top half of the figure represented the comparisons between the molecular data alone and the combination data. The \( P \) values on the lower half of the figure represented the comparisons between the clinical data alone and the combination data.

Establishment of the prognostic model based on the copy number data
To establish the prognostic model based on the copy number data, we included all 371 patients with complete copy number variation data and clinical data. After the cox regression screen and LASSO, the 10p15.1 and 15q26.3 were identified to be the independent prognostic factors for the liver cancer. Afterwards, the prognostic model based on the significant copy number variation was built, which was able to divide the patients into the high risk group and low risk group (Figure 3A). Kaplan-Meier survival curve showed that there was significant difference with respect to the overall survival between these two group patients (Figure 3B). Afterwards, the limma and GSEA were performed to evaluate the different gene expression and pathways between these two groups. The heatmap showed that there was a distinct gene expression pattern between the high risk group and low risk group (Figure 3C). The GSEA showed that the spermatogenesis, WNT/beta-catenin, E2F targets, mitotic spindle and G2M checkpoint were the top five enriched pathways in the high risk group patients (Figure 3D).
model. After the cox regression screen and LASSO, CCDC21, GTF3C2, and DBF4 were selected to build the prognostic model, which divided the patients into the high risk group and low risk group based on each patient’s prognostic score (Figure 6A). According to the Kaplan-Meier survival curve and log-rank test, there was significant difference of survival between the high risk patients and low risk patients. The heatmap showing the different gene expression patterns of the high risk group and low risk group. It showed that the gene expression patterns of high risk group and low risk group were obviously distinct. D: The top enriched pathways in the high risk group, as indicated by gene set enrichment analysis. It showed that spermatogenesis, WNT/beta-catenin, E2F targets, mitotic spindle and G2M checkpoint were the top five enriched pathways in the high risk group patients.

**DISCUSSION**

Several months after the publication of liver cancer TCGA, we systemically evaluated the prognostic power of different omics data of liver cancer[8]. We also explored whether additive prognostic power could be gained by the combination of molecular data with clinical data. In addition, based on the significant prognostic variables identified from the cox regression and LASSO analysis in different omics data, we also built the prognostic models.

In the evaluation of different omics data, we did not include the mutation data from DNA sequencing platform, since the mutation gene and hot-spot mutation site varied a lot among different patients, and it is difficult to divide the patients into a high risk group and low risk group. In addition, it might also need some complicated bioinformatical algorithm to evaluate the prognostic power.

Our results showed that the mRNA data alone seemed to be the most informative prognostic variable (C-index = 0.61) among all the molecular platforms. Consistently, by analyzing breast cancer, glioblastoma multiforme, myeloid leukemia and lung squamous cell carcinoma data, Zhao demonstrated that molecular variables evaluated at the transcription level could reflect patient survival more effectively than those evaluated at the DNA/epigenetic level. There have been several studies focusing on the mRNA prognostic models of liver cancer. For example, Chen *et al.*[10] combined three mRNA...
markers to build the prognostic model and demonstrated it to be useful. Nault et al[11] built a 5-gene score associated with overall survival of liver cancer patients after resection. Zhang et al[12] defined a hepatic stellate cell 122-gene signature to identify liver cancer patients with poor prognosis. Although with lower prognostic power, some prognostic models based on the CNV, methylation, miRNA and protein data have also widely attracted attention[13-16].

Boulesteix et al[17] claimed that little attention was paid to the assessment of the added prognostic power of a molecular signature, given that the clinical predictors or an established model were available. Thus in this study, we showed that the combination of molecular data with clinical data could significantly boost the prediction accuracy of the molecular data alone in the CNV, methylation and miRNA data. On the other hand, the combination of clinical data with methylation, miRNA and mRNA data could significantly boost the prediction accuracy of the clinical data itself. Consistently, Yuan et al[9] summarized that incorporating molecular features with clinical data yielded significantly increased predictions (FDR < 0.05) for kidney cancer, glioblastoma multiforme and ovarian cancer, but the quantitative gains were limited. Some complicated bioinformatical algorithms might be needed to further improve the prognostic power of the combination data.

With respect to the prognostic models based on different omics data, we did not include the protein model, since no significant prognostic variable passed through the cox screen and LASSO analysis. We suppose that one reason is the relative sample size of the patients with protein data. The other reason is due to the complexity of the proteome data. In the copy number model, the 10p15.1 and 15q26.3 were included in our study. Previous studies showed that a telomerase repressor gene might be located on 10p15.1, thus relating it to the prognosis of liver cancer[18]. Meanwhile, genomic copy number variations of 15q26.3 were deemed to be predictive markers for the systemic recurrence of breast cancer.

Figure 4 Establishment of the prognostic model based on the methylation data. A: The high risk group and low risk group based on the prognostic score. The patients with the score of 0 were considered as high risk, and patients with the score of -1 and -2 were considered as low risk; B: The Kaplan-Meier survival curves of the high risk group and low risk group, which showed that there was a significant difference of survival between the high risk patients and low risk patients; C: The heatmap showing the different gene expression patterns of the high risk group and low risk group. It showed that the gene expression patterns of high risk group and low risk group were obviously distinct; D: The top enriched pathways in the high risk group, as indicated by the gene set enrichment analysis. It showed that the E2F targets, G2M checkpoint, Myc targets V1, spermatogenesis and PI3K/AKT/mTOR pathway signaling were among the top five enriched pathways in the high risk group of patients.
In the analysis of methylation data, REL and MCM2 were demonstrated to be independent prognostic factors. Previous studies showed that REL was involved in apoptosis, inflammation, immune response, and oncogenic processes. Constitutive activation of the Rel/NF-κB pathway could lead to oncogenesis by driving proliferation, enhancing cell survival, or promoting angiogenesis or metastasis[20]. There were also studies reporting that MCM2 was closely related to tumor grade and overall survival in renal cancer[21]. However, this is the first study to demonstrate the methylation of these two markers as prognostic factors in liver cancer.

miRNA alterations were reported to participate in the initiation and progression of human cancer[22]. In our study, miR-3690, miR-561, and miR-621 were included in the prognostic model. This is the first time that we identified miR-3690 as the prognostic variable in cancer prognosis. On the other hand, Qian et al.[23] reported that miR-561 inhibited cellular proliferation and invasion by targeting c-Myc in gastric cancer. MiR-621 was supposed to sensitize breast cancer to chemotherapy by inhibiting FBXO11 and increasing p53 activity[24]. Thus, it was not difficult to comprehend that they were all involved in the prognosis of liver cancer.

With respect to the mRNA model, CCDC21, GTF3C2 and DBF4 were demonstrated to be independent prognostic factors in liver cancer. CCDC21, also called Cep85, is an antagonist of Nek2A that suppresses centrosome disjunction[25]. GTF3C2 is essential for RNA polymerase III-mediated transcription. Until now, there are few studies relating CCDC21 and GTF3C2 to cancer progression or prognosis. With respect to DBF4, the CDC7-DBF4 kinase, which was correlated with p53 inactivation, was reported to be overexpressed in multiple cancers[26]. Thus, it is the first study to correlate these markers with patient survival in liver cancer.

Interestingly, most prognostic variables identified in our study seemed to be novel markers in cancer study. Probably due to the strict screen procedure, a lot of
traditional prognostic markers were ignored, such as P53 and FOXM1. However, on the other hand, it also demonstrates that our prognostic markers might be more robust and more stable after the strict screening procedure. In summary, the major goal of our study was to evaluate the prognostic power of different omics data in liver cancer. For the first time, we showed that the mRNA data was the most informative prognostic variables in all kinds of omics data in liver cancer. In addition, we also revealed that the combination of clinical data with molecular data might be the future direction for cancer prognosis and prediction. Larger sample size and more mature bioinformatic algorithms were needed to predict the cancer prognosis more precisely in the future.

**ARTICLE HIGHLIGHTS**

**Research background**
Liver cancer is the fourth most common digestive cancer worldwide. Prognostic markers can help to make better clinical decision by selecting patients who respond well to some specific treatment. Besides traditional clinical markers, genetic biomarkers are emerging as novel indicators in cancer diagnosis and prognosis. The Cancer Genome Atlas (TCGA) is funded by the National Institute of Health (NIH) to describe the genomic alterations across cancer types. It provides tremendous amount of “omics” data, including mRNA sequencing, miRNA sequencing, reverse phase protein arrays, copy number change and DNA sequencing.

**Research motivation**
Although there seems to be great potential value of the clinical and genetic markers, there is no consensus on the predictive power of these indicators, especially the molecular markers.

**Research objectives**
By utilizing the TCGA data, we aimed to evaluate the prognostic power of liver cancer by molecular markers, and also to assess the predictive power of liver cancer by combining molecular markers and clinical data.

**Research methods**
Cox regression screen and least absolute shrinkage and selection operator (LASSO) were performed to select significant prognostic variables. Then the concordance index was calculated to evaluate the prognostic power. For the combination data, based on the clinical cox model, molecular features that better fit the model were combined to calculate the concordance index. Prognostic models were built based on the arithmetic summation of the significant variables. Kaplan-Meier survival curve and log-rank test were performed to compare the survival difference. Then the heatmap was constructed and gene set enrichment
REFERENCES

1 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018; 68: 7-30 [PMID: 29313949 DOI: 10.3322/caac.21442]

2 Maluccio M, Covey A. Recent progress in understanding, diagnosing, and treating hepatocellular carcinoma. CA Cancer J Clin 2012; 62: 394-399 [PMID: 23070000 DOI: 10.3322/caac.21161]

3 Xu XS, Qu K, Liu C, Zhang YL, Liu J, Song YZ, Zhang P, Liu SN, Chang HL. Highlights for α-fetoprotein in determining prognosis and treatment monitoring for hepatocellular carcinoma. World J Gastroenterol 2012; 18: 7242-7250 [PMID: 23326129 DOI: 10.3748/wjg.v18.i48.7242]

4 Borzio M, Dioni Gi, Rossini A, Marignani M, Sacco R, De Sio I, Bertolini E, Francia G, Giacomini A, Parisi G, Vicari S, Toldi A, Salmi A, Bocci S, Mita M, Formari F. External validation of the ITALIICA prognostic system for patients with hepatocellular carcinoma: A multicenter cohort study. Hepatology 2018; 67: 2215-2225 [PMID: 29165831 DOI: 10.1002/hep.29662]

5 Zucman-Rossi J, Villanueva A, Nault JC, Livoti VM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology 2015; 149: 1226-1239.e4 [PMID: 26995527 DOI: 10.1053/j.gast.2015.05.061]

6 Scaglione B, Kazemi M, Pozzato G, Dapra B, Farra R, Grassi M, Zanconati F, Grassi G. Novel hepatocellular carcinoma molecules with prognostic and therapeutic potentials. World J Gastroenterol 2014; 20: 1268-1288 [PMID: 24574801 DOI: 10.3748/wjg.v20.i5.1268]

7 Tomczak K, Czerwińska P, Wiśniewska M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015; 19: A66-A77 [PMID: 25691825 DOI: 10.5114/ wo.2014.47136]

8 Cancer Genome Atlas Research Network; Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell 2017; 169: 1327-1341.e23 [PMID: 28622513 DOI: 10.1016/j.cell.2017.05.046]

9 Yuan Y, Van Allen EM, Orombi 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, Liang H. Assessing the clinical utility of cancer genomic and proteomic data across tumor types. Nat Biotechnol 2014; 32: 644-652 [PMID: 24952901 DOI: 10.1038/nbt.2940]

10 Chen SS, Yu KK, Ling QX, Huang C, Li N, Zheng JM, Bao SX, Cheng Q, Zhu MQ, Chen MQ. The combination of three molecular markers can be a valuable predictive tool for the prognosis of hepatocellular carcinoma patients. Sci Rep 2016; 6: 24582 [PMID: 27094145 DOI: 10.1038/srep24582]

11 Nault JC, De Reyniès A, Villanueva A, Calderaro J, Rebourcoul S, Couchy G, Decens F, Poznański J, Leforestier A, Laurent-Puig P, Livoti VM, Zucman-Rossi J. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. Gastroenterology 2013; 145: 176-187 [PMID: 23567350 DOI: 10.1053/j.gastro.2013.03.051]

12 Zhang DY, Goosens N, Guo J, Tsai MC, Chow HH, Altukhaynak C, Sangiovanni A, Iavarone M, Colombo M, Kobayashi M, Kumada H, Villanueva A, Livoti VM, Hoshida Y, Friedman SL. A hepatic stellate cell gene expression signature associated with outcomes in hepatitis C cirrhosis and hepatocellular carcinoma after curative resection. Gut 2016; 65: 1754-1764 [PMID: 26045137 DOI: 10.1136/gutjnl-2015-309655]

13 Kawaguchi K, Honda M, Yamashita T, Okada H, Shirasaki T, Nishikawa M, Nio K, Araki K, Sakai Y, Yamashita T, Mizukoshi E, Kaneko S, Tagawa T. DNA Copy Number Variation Is Associated with Poor Outcome in Liver Cancer. Am J Pathol 2016; 186: 2055-2067 [PMID: 27315779 DOI: 10.1016/j.ajpath.2016.04.011]

14 Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Friedman SL. The combination of clinical data with molecular data might be the future direction for cancer prognosis and prediction. The Cancer Genome Atlas Research Network. Nat Rev Cancer 2014; 14: 672-686 [PMID: 24946162 DOI: 10.1038/nrclinonc.2014.229]

15 Boulesteix AL, Sauerbrei W. Added predictive value of high-throughput molecular data to clinical data and its validation. Brief Bioinform 2011; 12: 215-229 [PMID: 21425078 DOI: 10.1093/bib/bbr085]

16 Leuraud P, Aguirre-Cruz L, Huang-Xuan K, Crinière E, Tanguy ML, Golmard JL, Delattre JY, Sanson M. Telomerase reactivation in malignant gliomas and loss of heterozygosity on 10p15.1. Neurology 2003; 60: 1820-1822 [PMID: 12796538 DOI: 10.1212/01.WNL.000065894.89082.5E]

17 Hwang KT, Han W, Cho J, Lee JW, Ko E, Kim EK, Jung SY, Jeong EM, Bae JY, Kang JJ, Yang SJ, Kim SW, Noh DY. Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer. Int J Cancer 2008; 123: 1807-1815 [PMID: 18649361 DOI: 10.1002/ijc.23672]

18 Gilmore T, Gapuzan ME, Kalatzidis D, Starczynowski D. Rel/NF-κB/κappa B signal transduction in the generation and treatment of human cancer. Cancer Lett 2002; 181: 1-9 [PMID: 12430173 DOI: 10.1016/S0304-3835(01)00795-9]

19 Dudderidge TJ, Stoeber K, Loddo M, Atkinson G, Fanshawe T, Griffiths DF, Williams GH. MCM2, Geminin, and KI67 define proliferative state and are prognostic markers in renal cell carcinoma. J Proteome Res 2014; 13: 4833-4846 [PMID: 24946162 DOI: 10.1021/pr500229n]

20 Boulesteix AL, Sauerbrei W. Added predictive value of high-throughput molecular data to clinical data and its validation. Brief Bioinform 2011; 12: 215-229 [PMID: 21425078 DOI: 10.1093/bib/bbr085]

21 Leuraud P, Aguirre-Cruz L, Huang-Xuan K, Crinière E, Tanguy ML, Golmard JL, Delattre JY, Sanson M. Telomerase reactivation in malignant gliomas and loss of heterozygosity on 10p15.1. Neurology 2003; 60: 1820-1822 [PMID: 12796538 DOI: 10.1212/01.WNL.000065894.89082.5E]

22 Hwang KT, Han W, Cho J, Lee JW, Ko E, Kim EK, Jung SY, Jeong EM, Bae JY, Kang JJ, Yang SJ, Kim SW, Noh DY. Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer. Int J Cancer 2008; 123: 1807-1815 [PMID: 18649361 DOI: 10.1002/ijc.23672]

23 Gilmore T, Gapuzan ME, Kalatzidis D, Starczynowski D. Rel/NF-κB/κappa B signal transduction in the generation and treatment of human cancer. Cancer Lett 2002; 181: 1-9 [PMID: 12430173 DOI: 10.1016/S0304-3835(01)00795-9]

24 Dudderidge TJ, Stoeber K, Loddo M, Atkinson G, Fanshawe T, Griffiths DF, Williams GH. MCM2, Geminin, and KI67 define proliferative state and are prognostic markers in renal cell carcinoma. J Proteome Res 2014; 13: 4833-4846 [PMID: 24946162 DOI: 10.1021/pr500229n]

25 Boulesteix AL, Sauerbrei W. Added predictive value of high-throughput molecular data to clinical data and its validation. Brief Bioinform 2011; 12: 215-229 [PMID: 21425078 DOI: 10.1093/bib/bbr085]
Shao ZM, Wu ZH, Wu J. MiRNA-621 sensitizes breast cancer to chemotherapy by suppressing FBXO11 and enhancing p53 activity. *Oncogene* 2016; 35: 448-458 [PMID: 25867061 DOI: 10.1038/onc.2015.96]

Chen C, Tian F, Lu L, Wang Y, Xiao Z, Yu C, Yu X. Characterization of Cep85 - a new antagonist of Nek2A that is involved in the regulation of centrosome disjunction. *J Cell Sci* 2015; 128: 3290-3303 [PMID: 26220856 DOI: 10.1242/jcs.171637]

Bonte D, Lindvall C, Liu H, Dykema K, Farge K, Weinreich M. Cdc7-Dbf4 kinase overexpression in multiple cancers and tumor cell lines is correlated with p53 inactivation. *Neoplasia* 2008; 10: 920-931 [PMID: 18714392 DOI: 10.1593/neo.08216]
