Analyzing the value of NUSAP1 in hepatocellular carcinoma from clinical and molecular mechanism perspectives: Bioinformatics-based approach.

Wenjie Zhu (20192830201910@whu.edu.cn)
Renmin Hospital of Wuhan University: Wuhan University Renmin Hospital https://orcid.org/0000-0002-9025-6127

Jie Wang
Department of Hepatobiliary Surgery, Renmin Hospital of Wuhan University, Wuhan 430060, China

Zehao Chen
Department of Hepatobiliary Surgery, Renmin Hospital of Wuhan University, Wuhan 430060, China

Jianxin Jiang
Department of Hepatobiliary Surgery, Renmin Hospital of Wuhan University, Wuhan 430060, China

Research

Keywords: Hepatocellular carcinoma, Bioinformatics, NUSAP1, Molecular mechanism, Microarray

Posted Date: November 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-106028/v1

Abstract

Background: Hepatocellular carcinoma (HCC) is one of the most common and highest mortality rates carcinomas worldwide. At present, there are varies of therapeutic methods can be provided for HCC patients. However, there is no one method can inhibit occurrence and development of HCC very well, and prognosis of HCC patients is still very poor. Herein, our study aims to identify a key gene closely related to occurrence, development, poor prognosis of HCC and explore its underlying mechanism from molecular level.

Methods: GSE62232, GSE102079, GSE112790 and GSE121248 genes expression profile datasets were screened from Gene Expression Omnibus (GEO) database. R studio was used to identify DEGs of each dataset. Venn online tool was used to generate a Venn diagram and screen overlapping DEGs of the four datasets. Search Tool for the Retrieval of Interacting Genes (String) online tool was used to draw Protein–Protein Interaction (PPI) network. And the most significant module and the key gene NUSAP1 in PPI network were identified by MCODE and cytoHubba plug-in in Cytoscape software. Oncomine database and Kaplan–Meier Plotter database were used to analyze relationships between expression of NUSAP1 and occurrence, development, prognosis of HCC. The cBioPortal online tool was used to identify co-expression genes of NUSAP1 in HCC patients from TCGA database. Then, KEGG pathway analysis was carried out by DAVID online tool and Cell cycle pathway map was generated by Kyoto Encyclopedia of Genes and Genomes (KEGG) online tool.

Results: A total of 86 overlapping DEGs were screened, which included 55 up-regulated DEGs and 31 down-regulated DEGs. Then the key gene NUSAP1 in the PPI network were screened using cytoHubba plug-in in Cytoscape software. We found NUSAP1 may be associated with occurrence, development and poor prognosis of HCC by analyzing HCC patients in Oncomine database and Kaplan–Meier Plotter. Co-expression genes of NUSAP1 in TCGA database were obtained by cBioPortal. And KEGG pathway analysis was produced using the top 300 co-expression genes of NUSAP1, the result showed most co-expression genes closely related to the expression of NUSAP1 were concentrated in Cell cycle. Thus, we generate a KEGG pathway map of Cell cycle and found that most of these genes were located in S phase and G2/M phase of the Cell cycle and they could regulate the genes in G1 phase, hence, we inferred that NUSAP1 may regulate the progression of HCC by promoting the transition from the G1 phase to the S phase.

Conclusion: NUSAP1 may influence occurrence, development and prognosis of HCC and might be a new molecular marker in HCC.

Background

Hepatocellular carcinoma (HCC) is the sixth most common and the fourth deadliest malignant tumors globally [1]. And it is particularly prevalent in China, Asia and Africa [2]. About 60% new hepatocellular carcinoma cases occur in China every year, and the 5-year survival rate is approximately 12%[3, 4]. In

terms of etiology, the occurrence of hepatocellular carcinoma is related to HBV, HCV, alcohol, aflatoxin, autoimmune diseases, diabetes, obesity and so on [5, 6]. There are many treatment options for HCC now, including surgical resection, orthotopic liver transplantation, ablation, transcatheter arterial chemical embolization (TACE), systemic chemotherapy, etc [7–9]. Despite so many treatment options available, the overall survival (OS) of HCC patients remains poor due to its extremely high rates of postoperative recurrence and metastasis [10]. Therefore, there is an urgent need to study underlying molecular mechanism of HCC occurrence, development and poor prognosis in order to explore some better strategies of prevention, diagnosis and treatment in HCC, which has important clinical significance.

At present, bioinformatics methods and microarrays have been widely used to screen differentially expressed gene (DEGs) in tumors. Herein, we collected four HCC-related mRNA microarray datasets from Gene Expression Omnibus (GEO) and obtained differentially expressed genes (DEGs) between HCC and normal liver tissues by R studio. Then, we obtained overlapping DEGs of the four HCC-related mRNA microarray datasets by Venn online tool and analyzed the relationships among overlapping genes by protein–protein interaction (PPI) network. MCODE and ctyoHubba plug-in in Cytoscape software was respectively used to screen the most significant module in PPI network and key gene in all overlapping genes. Wurmbach liver dataset in Oncomine was used to explore the relatioships between the expression of key gene and occurrence, development of HCC. Kaplan–Meier Plotter was used to analyze the relationship between the expression of key gene and prognosis of HCC. And then, we used cBioPortal to identify co-exprssion genes of key gene, and used Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway map of Cell cycle to analyze the underlying possible mechanism of key gene influencing progression of HCC. And our study may be of significant value to studies in regard to exploring clinical and molecular mechanism of HCC in the future.

Materials And Methods

Microarray data

In our study, four genes expression profile datasets were screened using the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) database of NCBI, including the GSE62232 series[11], the GSE102079 series[12], the GSE112790 series[13] and the GSE121248 series[14]. The GSE62232 dataset included 81 HCC tissue samples and 10 normal liver tissue samples, the GSE102079 dataset included 183 HCC tissue samples and 15 normal liver tissue samples, the GSE112790 dataset included 70 HCC tissue samples and 37 normal liver tissue samples, and the GSE121248 dataset included 152 HCC tissues samples and 105 adjacent normal liver tissue samples. The microarray data from GSE62232, GSE102079, GSE112790 and GSE121248 were based on the GPL570 platform (HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array.

Identification Of DEGs

R studio was used to identify DEGs between HCC and normal liver samples by analyzing GSE62232, GSE102079, GSE112790 and GSE121248 raw data of CEL files. Firstly, we used the RMA package to

normalize all the raw database and the Affy package to assess the quality of samples in each dataset. Then, according to the annotation information in GPL570 platform, probes were changed into the corresponding gene symbols using R studio. At last, the Limma package was used to identify DEGs. The criterion of identifying DEGs was |logFC|>2 and adjusted $p < 0.05$.

Screening Overlapping DEGs

In order to screen overlapping DEGs among GSE62232, GSE102079, GSE112790 and GSE121248, we used the Venn online tool (http://jvenn.toulouse.inra.fr/app/example.html) to draw a Venn diagram and find the overlapping DEGs among GSE62232, GSE102079, GSE112790 and GSE121248.

Construction and module analysis of the Protein–Protein Interaction (PPI)network

The PPI network was generated by the Search Tool for the Retrieval of Interacting Genes (String, http://string-db.org Ver-sion:11.0) online database[15]. The MCODE plug-in in Cytoscape software was used to selected significant modules in the PPI network, and we could find the most important module in all selected significant modules. The criteria of selecting significant modules were degree cutoff = 2, node density cutoff = 0.1, node score cutoff = 0.2, k-core = 2 and maximum depth = 100.

Hub gene selection and analysis

CytoHubba plug-in in Cytoscape software was used to screen the top 10 hub genes in PPI network. There were 12 calculating methods in cytoHubba, which included Betweenness, BottleNeck, Closeness, ClusteringCoefficient, Degree, DMNC, EcCentricity, EPC, MCC, MNC, Radiality and Stress. And we used the 12 different calculating methods to get 12 different outcomes of the top 10 hub genes. By analyzing the 12 outcomes, we found the key gene NUSAP1. The Wurmbach liver dataset[16] in Oncomine (http://www.oncom ine.com)[17] was used to analyze key gene NUSAP1 expression levels among normal liver tissues, cirrhosis, hepatocellular carcinoma and liver cell dysplasia. Besides, the relationships between NUSAP1 expression level and different HCC grades, hepatitis C virus infection statue, satellites, vascular invasion or pathological types (detailed) were also analyzed using Wurmbach liver dataset of Oncomine[16, 17]. The Kaplan–Meier Plotter (http://kmplot.com/analysis/) [18] was used to generate Kaplan–Meier curves of overall survival (OS), relapse free survival(RFS)progress free survival (PFS) and disease-specific survival (DSS) of NUSAP1 by analyzing 364, 316, 370, and 362 HCC patients respectively. Then, we used cBioPortal (https://www.cbioportal.org/)[19, 20] to screen the co-expression genes of NUSAP1 in The Cancer Genome Atlas (TCGA) database, and used DAVID (https://david.ncifcrf.gov/home.jsp) [21, 22] to generate KEGG pathway analysis of the co-expression genes of NUSAP1. The pathway map of Cell cycle was generated by Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.kegg.jp/)[23], and the corresponding positions of co-expression genes in the map were marked in red color.

Results

Identification and analysis of DEGs

DEGs of each microarray dataset were identified using R studio (238 in GSE62232, 258 in GSE102079, 169 in GSE112790 and 197 in GSE121248), and the cut-off criterion was |logFC|>2 and adjusted p < 0.05(Supplementary material 1). A total of 86 overlapping DEGs in four microarray datasets were identified using a Venn diagram (Fig. 1 and Supplementary material 2). Then, we used STRING online database to generate a PPI network of 86 overlapping DEGs, and used Cytoscape software for visualization (Fig. 2a); it included 55 up-regulated genes (marked in red) and 31 down-regulated genes (marked in blue). MCODE plug-in in Cytoscape software was used to screen the most significant module in the PPI network of 86 DEGs (Fig. 2b); it included 32 nodes and 474 edges, all of 32 genes in this module were up-regulated genes (marked in red) in HCC tissues.

Screening And Analysis Of Top 10 Hub Genes

CytoHubba plug-in in Cytoscape software was used to find the top 10 hub genes in the PPI network containing 86 DEGs. There were 12 calculating methods in cytoHubba, which included Betweenness, BottleNeck, Closeness, ClusteringCoefficient, Degree, DMNC, EcCentricity, EPC, MCC, MNC, Radiality and Stress. And we used the 12 different calculating methods to get 12 different outcomes of the top 10 hub genes (Fig. 3a-l and Supplementary material 3). By analyzing the 12 different outcomes of the top 10 hub genes, we found that NUSAP1 existed in 8 outcomes (including Betweenness, Closeness, Degree, EPC, MCC, MNC, Radiality and Stress), and NUSAP1 was the highest score in 6 outcomes (including Betweenness, Closeness, Degree, MCC, MNC and Radiality); CDK 1 existed in 9 outcomes (including BottleNeck, Closeness, Degree, EcCentricity, EPC, MCC, MNC, Radiality, Stress), however, CDK1 was the highest score only in 3 outcomes (including EcCentricity, MCC and MNC); RAD51AP1 existed in 8 outcomes(including Closeness, Degree, EcCentricity, EPC, MCC, MNC, Radiality and Stress), and RAD51AP1 was the highest score only in 3 outcomes (including EcCentricity, MCC and MNC). Hence, we suggest NUSAP1 may be a key gene which associated with HCC occurrence, development and poor prognosis.

Analysis of key gene NUSAP1

The key gene NUSAP1 full name and corresponding functions are listed in Table 1. And, we used the Wurmbach liver dataset in Oncomine to analyze expression of NUSAP1 in normal liver, cirrhosis, hepatocellular carcinoma and liver cell dysplasia (Fig. 4a). We found the expression of NUSAP1 in normal tissue, cirrhotic and hepatocyte dysplasia was low, but the expression in hepatocellular carcinoma was significantly high. The result indicates that NUSAP1 may play a significant role in the occurrence of HCC. Meanwhile, we further analyzed the expression of NUSAP1 from different aspects using the Wurmbach liver dataset, including different HCC grades (Fig. 4b), whether infected with hepatitis C virus (Fig. 4c), whether had satellites (Fig. 4d), whether had vascular invasion (Fig. 4e) and different detailed pathological types (Fig. 4f). The results showed that higher expression of NUSAP1 was significantly associated with higher HCC grades, hepatitis virus infection, satellites and vascular invasion. Therefore, we further affirmed that higher expression of NUSAP1 was related to occurrence and development of

HCC. Hepatitis C virus infection cases also had a higher expression of NUSAP1, it might exactly explain that hepatitis virus infection was one risk factor of HCC. Subsequently, in order to further explore the relationships between NUSAP1 and prognosis in HCC, we used Kaplan–Meier Plotter database to analyze overall survival (OS), relapse free survival(RFS), progress free survival (PFS) and disease-specific survival (DSS) of NUSAP1 in HCC patients. As the Kaplan–Meier curve shows, we noted that HCC patients with high NUSAP1 expression were associated with a decrease in OS (Fig. 5a), RFS (Fig. 5b), PFS (Fig. 5c) and DSS (Fig. 5d). Hence, we guess the high expression of NUSAP1 was associated with poor prognosis of HCC.

Table 1
Function of key gene NUSAP1.

Gene symbol	Full name	Function
NUSAP1	Nucleolar and spindle-associated protein 1	NUSAP1 is a nucleolar-spindle-associated protein that plays a role in spindle microtubule organization.

Screening and analysis of co-expression genes of NUSAP1 in HCC

To further explore the underlying mechanism of the influence of NUSAP1 in HCC, we used cBioProtal online data analysis tool to screen co-expression genes of NUSAP1 in TCGA HCC transcriptome data (Pearson scores > 0.3, Spearman scores > 0.3), the top 20 co-expression genes of NUSAP1 were listed in Table 2. Subsequently, we used DAVID online data analysis tool to generate the KEGG pathways of the top 300 co-expression genes of NUSAP1(Supplementary material 4), the top 10 hits were listed in Table 3. The result indicted that the expression of NUSAP1 mainly influence HCC Cell cycle, DNA replication, Oocyte meiosis, Fanconi anemia pathway and so on. Meanwhile, we continued to observe Table 3 and noticed that most co-expression genes closely related to the expression of NUSAP1 were concentrated in Cell cycle. Thus, we generated a Cell cycle pathway map using these genes in Cell cycle (Fig. 5), positions of these genes in the Cell cycle pathway map were marked with red color. The results showed that most of these genes were located in S phase and G2/M phase of the Cell cycle, and they could regulate the genes in G1 phase. Herein, we speculated that NUSAP1 may regulate progression of HCC mainly by promoting the transition from the G1 phase to the S phase. This conclusion was of great clinical significance, if there was a targeted therapy drug that could inhibit the function of NUSAP1 to promote G1 to S phase transformation, it was possible to inhibit the progression of HCC.

Table 2
The top 20 co-expression genes of NUSAP1, which were screened by cBioProtal data analysis platform (Pearson score > 0.3, Spearman scores > 0.3).

Gene	Cytoband	Spearman's correlation	p-value	q-value
BUB1B	15q15.1	0.927414494	5.75E-155	1.16E-150
PRC1	15q26.1	0.920889495	1.55E-148	1.56E-144
TOP2A	17q21.2	0.916780924	9.17E-145	6.14E-141
ARHGAP11A	15q13.3	0.916274747	2.59E-144	1.30E-140
KIF23	15q23	0.913767397	4.05E-142	1.63E-138
KIF18B	17q21.31	0.907355696	8.40E-137	2.81E-133
TPX2	20q11.21	0.895222243	1.00E-127	2.88E-124
OIP5	15q15.1	0.892248583	1.14E-125	2.86E-122
LMNB1	5q23.2	0.888274787	5.14E-123	1.15E-119
KIF18A	11p14.1	0.886901834	4.02E-122	8.07E-119
POLQ	3q13.33	0.886469088	7.65E-122	1.40E-118
CKAP2L	2q14.1	0.882338655	3.11E-119	5.21E-116
KIF11	10q23.33	0.880169399	6.68E-118	1.03E-114
HJURP	2q37.1	0.87933053	2.15E-117	3.09E-114
GINS1	20p11.21	0.8789635	3.58E-117	4.79E-114
FOXM1	12p13.33	0.876505662	1.04E-115	1.30E-112
KIF4A	Xq13.1	0.875040502	7.46E-115	8.82E-112
NUF2	1q23.3	0.873623695	4.91E-114	5.48E-111
CDK1	10q21.2	0.872852727	1.36E-113	1.43E-110
PLK4	4q28.1	0.872672685	1.72E-113	1.73E-110

Table 3
KEGG pathway enrichment analysis of the top 300 co-expression genes of NUSAP1.

Gene Set Name	Count	P-Value	Genes
Cell cycle	35	5.43E-36	PCNA, MCM7, BUB1B, TTK, PKMYT1, CDC20, CCNB2, CCNB1, CDC45, ORC6, PTTG1, ORC1, CHEK1, E2F1, E2F2, BUB1, PLK1, CDC7, CDC6, CDC25C, CDC25A, CCNA2, DBF4, RBL1, ESPL1, CCNE2, CCNE1, CDK2, CDK1, MCM3, MCM4, MCM5, MCM6, MCM2, MAD2L1
DNA replication	21	1.35E-28	RFC5, PRIM2, FEN1, RFC3, PCNA, RFC4, RNASEH2A, LIG1, MCM7, PRIM1, POLD3, POLA1, POLA2, POLE2, POLD1, MCM3, MCM4, MCM5, MCM6, DNA2, MCM2
Oocyte meiosis	17	1.91E-12	PLK1, PKMYT1, CDC25C, FBXO43, AURKA, SGO1, CDC20, CCNB2, PTTG1, ESPL1, CCNE2, CCNE1, CDK2, CDK1, FBXO5, BUB1, MAD2L1
Fanconi anemia pathway	13	3.40E-12	FANCI, BLM, RMI2, FANCB, BRCA1, BRCA2, FANCE, FANCG, RAD51, EME1, FANCD2, UBE2T, USP1
Mismatch repair	9	6.06E-10	POLD3, RFC5, RFC3, PCNA, RFC4, MSH2, LIG1, EXO1, POLD1
Homologous recombination	9	3.62E-09	POLD3, BLM, RAD51, EME1, POLD1, XRCC2, XRCC3, RAD54L, BRCA2
Progesterone-mediated oocyte maturation	11	3.72E-07	CCNA2, CCNB2, CCNB1, PLK1, CDK2, CDK1, CDC25C, PKMYT1, BUB1, CDC25A, MAD2L1
MicroRNAs in cancer	13	6.46E-07	DNMT1, CDCA5, KIF23, BRCA1, CDC25C, CDC25A, CCNE2, CCNE1, STMN1, DNMT3B, E2F1, E2F2, EZH2
Pyrimidine metabolism	11	7.05E-07	POLD3, PRIM2, POLA1, RRM1, POLA2, RRM2, PRIM1, POLE2, POLD1, TK1, TYMS
Nucleotide excision repair	8	2.71E-06	POLD3, RFC5, RFC3, PCNA, RFC4, LIG1, POLE2, POLD1

Discussion

Hepatocellular carcinoma (HCC) is the sixth most common and the fourth deadliest malignant tumors globally, it has a serious impact on human health. Herein, to study the relationships between key gene and occurrence, development, prognosis of HCC is absolutely a necessary affair, it may guide us to find a new molecular marker in HCC, the underlying molecular mechanisms of HCC and even a new treatment option at the molecular and genetic level of HCC.

Herein, four gene chip datasets of HCC (GSE62232, GSE102079, GSE112790 and GSE121248) from GEO database were screened for bioinformatic analysis, and 86 common DEGs in the four datasets were identified. We used the 86 common DEGs to make a PPI network by String online tool. Then, the top 10 hub genes in PPI network were screened by cytoHubba plug-in in Cytoscope using 12 different calculating methods, thus, we gain 12 different outcomes of the top hub genes. Subsequently, we screened the key gene NUSAP1 by analyzing the 12 outcomes, which existed in 8 outcomes and was the highest score in 6 outcomes. Therefore, we carried out the following study around the NUSAP1 gene which we regarded as key gene and might related to HCC occurrence, development and poor prognosis.

Nucleolar and spindle-associated protein 1 (NUSAP1) is a nucleolar-spindle-associated protein that plays a role in spindle microtubule organization, which belongs to the NUSP1 family[24, 25]. Before our study, many studies had demonstrated that the overexpression of NUSAP1 was observed in many human neoplasms including colon cancer[26, 27], astrocytoma[28], glioblastoma multiforme[29], renal cell carcinoma[30], prostate cancer[31–33], oral squamous cell carcinoma[34], breast cancer[35, 36], cervical carcinoma[37] and esophageal squamous cell carcinoma[38]. Besides, many studies also showed that overexpression of NUSAP1 was associated with poor survival of colon cancer[26], astrocytoma[28], glioblastoma multiforme[29], renal cell carcinoma[30], prostate cancer[33], breast cancer[36] and esophageal squamous cell carcinoma[38]. However, few studies had demonstrated the relationships between the expression of NUSAP1 and hepatocellular carcinoma. We found two study which were related to NUSAP1 and hepatocellular carcinoma, one study demonstrated that NUSAP1 was a target of miRNA 193a-5p and microRNA 193a-5p can regulate levels of NUSAP1, HCC with low levels of miRNA 193a-5p could increase expression of NUSAP1, and the overexpression of NUSAP1 in HCC samples correlated with shorter survival times of patients[39]. Another study was a transcriptome analysis, by analyzing microarray datasets incorporating cirrhosis and HCC subjects from Gene Expression Omnibus (GEO) database, the author found that NUSAP1 was one of the top 5 significant genes which were associated with onset, progression, prognosis of HCC and exhibited higher expression in HCC compared with normal livers[40]. Herein, to further explore the relationships between the expression of NUSAP1 and HCC, we did the next series of study. Firstly, we attempted to explore the expression of NUSAP1 in normal liver, cirrhosis, hepatocellular carcinoma and liver cell dysplasia, and we found the expression of NUSAP1 significantly higher in hepatocellular carcinoma. Thus, we continued to demonstrate the relationships between NUSAP1 and HCC grades, hepatitis C virus infection statue, satellites and vascular invasion. The results showed expression of NUSAP1 was significant higher in HCC grades, hepatitis virus infection, satellites and vascular invasion. According to these results, we speculated that higher expression of NUSAP1 might be closely associated with occurrence and development of HCC. Hepatitis C virus infection patients also had a higher expression of NUSAP1 compared with the patients without hepatitis C virus infection, it might exactly explain that hepatitis virus infection was one risk factor of HCC. Then, we want further to explore the relationships between the expression of NUSAP1 and overall survival (OS),relapse free survival(RFS)progress free survival (PFS), disease-specific survival (DSS). We found that HCC patients with high expression of NUSAP1 had a lower OS ($P < 0.05$), RFS ($P < 0.05$), PFS ($P < 0.05$) and DSS ($P < 0.05$) compared with the HCC patients with low expression of NUSAP1, which were

analyzed using HCC patients in Kaplan–Meier Plotter database. Herein, we affirmed that high expression of NUSAP1 was closely connected with the poor prognosis of HCC.

Although we had demonstrated the relationships between high expression of NUSAP1 and occurrence, development, poor prognostic of HCC, the mechanism of NUSAP1 influencing occurrence, development and poor prognosis of HCC was not very clear until now. We found only one study reported that hepatitis B virus X protein can enhance hepatocarcinogenesis by depressing the targeting of NUSAP1 mRNA by miRNA18b, the specific mechanism was the targeting of NUSAP1 mRNA by the tumor suppressor miRNA18b was controlled by hepatitis B virus X modulated promoter methylation during the host-virus interaction, leading to hepatocarcinogenesis[41]. Hence, in order to further explore the mechanism of NUSAP1 influencing occurrence, development and poor prognosis of HCC, we screened co-expression of NUSAP1 in HCC by cBioProtal online data analysis tool. Then, we used top 300 co-expression genes of NUSAP1 to make KEGG pathway analysis by DAVID online data analysis tool, the result showed that the expression of NUSAP1 mainly affected pathways related to Cell cycle, DNA replication, Oocyte meiosis and Fanconi anemia pathway. Among these pathways, we found that co-expression genes of NUSAP1 were mainly located in Cell cycle pathway. Herein, we continued to generate a KEGG pathway map using these co-expression genes of NUSAP1 in Cell cycle pathway, and the map showed that these genes were mainly located in S phase and G2/M phase of the cell cycle. Thus, we speculated that the mechanism of NUSAP1 might regulate occurrence, development of HCC by promoting the transition from the G1 phase to the S phase.

Through these analyses of multiple databases and tools to NUSAP1, our study demonstrated the high expression of NUSAP1 might be associated with occurrence of HCC by comparing normal liver, cirrhosis, hepatocellular carcinoma and liver cell dysplasia; the high expression of NUSAP1 might be related to development of HCC by comparing different HCC grades, whether had satellites and whether had vascular invasion; the high expression of NUSAP1 might be concerned with poor prognosis of HCC by making Kaplan–Meier curves, including OS, RFS, PFS and DFS. At last, our study speculated the mechanism of NUSAP1 of regulating progression of HCC by generating KEGG pathway map of Cell cycle, which showed that NUSAP1 may facilitate progression of HCC by promoting the transition from the G1 phase to the S phase. And we concluded that NUSAP1 might be of great value to the early diagnosis and treatment of HCC according to our study outcomes.

Conclusion

By integrating and analyzing four microarray datasets, we screened one key gene (NUSAP1) which was closely associated with the occurrence, development and poor prognosis of HCC. Meanwhile, we further studied the mechanism of NUSAP1 in HCC, and found NUSAP1 played a significant role in regulating HCC progression by promoting the transition from the G1 phase to the S phase in cell cycle. Our results were of great clinical significance, it might provide some new ideas about early diagnosis, targeted therapy and prognostic judgement of HCC.

Abbreviations

HCC: hepatocellular carcinoma; GEO: Gene Expression Omnibus; DEGs: differentially expressed genes; PPI: protein–protein interactions; KEGG: Kyoto Encyclopedia of Genes and Genomes; String: Search Tool for the Retrieval of Interacting Genes; TCGA: The Cancer Genome Atlas; NUSAP1: nucleolar and spindle-associated protein 1; OS: overall survival; RFS: relapse free survival; PFS: progress free survival; DSS: disease-specific survival.

Declarations

Acknowledgements

None.

Authors' contributions

ZWJ, WJ and JJX designed the study. ZWJ and CZH collected the literature. ZWJ performed statistical analyses. ZWJ, WJ, CZH and JJX analyzed the data. ZWJ wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by National Natural Science Foundation of China (NO.81871965).

Availability of data and materials

The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA: a*

cancer journal for clinicians 2018, **68**(6):394-424.

2. Forner A, Llovet JM, Bruix J: **Hepatocellular carcinoma**. *Lancet (London, England)* 2012, **379**(9822):1245-1255.
3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J: **Cancer statistics in China, 2015**. *CA: a cancer journal for clinicians* 2016, **66**(2):115-132.
4. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, Xia C, Sun K, Yang Z, Li H *et al*: **Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries**. *The Lancet Global health* 2018, **6**(5):e555-e567.
5. Bruix J, Sherman M: **Management of hepatocellular carcinoma: an update**. *Hepatology (Baltimore, Md)* 2011, **53**(3):1020-1022.
6. Villanueva A: **Hepatocellular Carcinoma**. *The New England journal of medicine* 2019, **380**(15):1450-1462.
7. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, Roberts LR, Heimbach JK: **Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases**. *Hepatology (Baltimore, Md)* 2018, **68**(2):723-750.
8. Cillo U, Noaro G, Vitale A, Neri D, D'Amico F, Gringeri E, Farinati F, Vincenzi V, Vigo M, Zanus G: **Laparoscopic microwave ablation in patients with hepatocellular carcinoma: a prospective cohort study**. *HPB : the official journal of the International Hepato Pancreato Biliary Association* 2014, **16**(11):979-986.
9. Xue T, Le F, Chen R, Xie X, Zhang L, Ge N, Chen Y, Wang Y, Zhang B, Ye S *et al*: **Transarterial chemoembolization for huge hepatocellular carcinoma with diameter over ten centimeters: a large cohort study**. *Medical oncology (Northwood, London, England)* 2015, **32**(3):64.
10. Villanueva A, Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, Cornella H, Liberzon A, Kobayashi M, Kumada H *et al*: **Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma**. *Gastroenterology* 2011, **140**(5):1501-1512.e1502.
11. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F *et al*: **Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets**. *Nature genetics* 2015, **47**(5):505-511.
12. Chiyonobu N, Shimada S, Akiyama Y, Mogushi K, Itoh M, Akahoshi K, Matsumura S, Ogawa K, Ono H, Mitsunori Y *et al*: **Fatty Acid Binding Protein 4 (FABP4) Overexpression in Intratumoral Hepatic Stellate Cells within Hepatocellular Carcinoma with Metabolic Risk Factors**. *The American journal of pathology* 2018, **188**(5):1213-1224.
13. Shimada S, Mogushi K, Akiyama Y, Furuyama T, Watanabe S, Ogura T, Ogawa K, Ono H, Mitsunori Y, Ban D *et al*: **Comprehensive molecular and immunological characterization of hepatocellular carcinoma**. *EBioMedicine* 2019, **40**:457-470.
14. Wang SM, Ooi LL, Hui KM: **Identification and validation of a novel gene signature associated with the recurrence of human hepatocellular carcinoma**. *Clinical cancer research : an official journal of the*

American Association for Cancer Research 2007, **13**(21):6275-6283.

15. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP *et al*: **STRING v10: protein-protein interaction networks, integrated over the tree of life**. *Nucleic acids research* 2015, **43**(Database issue):D447-452.
16. Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J *et al*: **Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma**. *Hepatology (Baltimore, Md)* 2007, **45**(4):938-947.
17. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: **ONCOMINE: a cancer microarray database and integrated data-mining platform**. *Neoplasia (New York, NY)* 2004, **6**(1):1-6.
18. Menyhárt O, Nagy Á, Győrffy B: **Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma**. *Royal Society open science* 2018, **5**(12):181006.
19. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al*: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data**. *Cancer discovery* 2012, **2**(5):401-404.
20. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E *et al*: **Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal**. *Science signaling* 2013, **6**(269):pl1.
21. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC *et al*: **DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists**. *Nucleic acids research* 2007, **35**(Web Server issue):W169-175.
22. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic acids research* 2009, **37**(1):1-13.
23. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes**. *Nucleic acids research* 1999, **27**(1):29-34.
24. Raemaekers T, Ribbeck K, Beaudouin J, Annaert W, Van Camp M, Stockmans I, Smets N, Bouillon R, Ellenberg J, Carmeliet G: **NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization**. *The Journal of cell biology* 2003, **162**(6):1017-1029.
25. Petry S: **Mechanisms of Mitotic Spindle Assembly**. *Annual review of biochemistry* 2016, **85**:659-683.
26. Liu Z, Guan C, Lu C, Liu Y, Ni R, Xiao M, Bian Z: **High NUSAP1 expression predicts poor prognosis in colon cancer**. *Pathol Res Pract* 2018, **214**(7):968-973.
27. Han G, Wei Z, Cui H, Zhang W, Wei X, Lu Z, Bai X: **NUSAP1 gene silencing inhibits cell proliferation, migration and invasion through inhibiting DNMT1 gene expression in human colorectal cancer**. *Exp Cell Res* 2018, **367**(2):216-221.
28. Wu X, Xu B, Yang C, Wang W, Zhong D, Zhao Z, He L, Hu Y, Jiang L, Li J *et al*: **Nucleolar and spindle associated protein 1 promotes the aggressiveness of astrocytoma by activating the Hedgehog signaling pathway**. *J Exp Clin Cancer Res* 2017, **36**(1):127.

29. Qian Z, Li Y, Ma J, Xue Y, Xi Y, Hong L, Dai X, Zhang Y, Ji X, Chen Y *et al*: **Prognostic value of NUSAP1 in progression and expansion of glioblastoma multiforme**. *J Neurooncol* 2018, **140**(2):199-208.
30. Fang L, Zhang M, Chen L, Xiong H, Ge Y, Lu W, Wu X, Heng B, Yu D, Wu S: **Downregulation of nucleolar and spindle-associated protein 1 expression suppresses cell migration, proliferation and invasion in renal cell carcinoma**. *Oncol Rep* 2016, **36**(3):1506-1516.
31. Gulzar ZG, McKenney JK, Brooks JD: **Increased expression of NuSAP in recurrent prostate cancer is mediated by E2F1**. *Oncogene* 2013, **32**(1):70-77.
32. Gordon CA, Gulzar ZG, Brooks JD: **NUSAP1 expression is upregulated by loss of RB1 in prostate cancer cells**. *Prostate* 2015, **75**(5):517-526.
33. Gordon CA, Gong X, Ganesh D, Brooks JD: **NUSAP1 promotes invasion and metastasis of prostate cancer**. *Oncotarget* 2017, **8**(18):29935-29950.
34. Okamoto A, Higo M, Shiiba M, Nakashima D, Koyama T, Miyamoto I, Kasama H, Kasamatsu A, Ogawara K, Yokoe H *et al*: **Down-Regulation of Nucleolar and Spindle-Associated Protein 1 (NUSAP1) Expression Suppresses Tumor and Cell Proliferation and Enhances Anti-Tumor Effect of Paclitaxel in Oral Squamous Cell Carcinoma**. *PLoS One* 2015, **10**(11):e0142252.
35. Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, Kemmner W: **Identification of early molecular markers for breast cancer**. *Mol Cancer* 2011, **10**(1):15.
36. Chen L, Yang L, Qiao F, Hu X, Li S, Yao L, Yang XL, Shao ZM: **High Levels of Nucleolar Spindle-Associated Protein and Reduced Levels of BRCA1 Expression Predict Poor Prognosis in Triple-Negative Breast Cancer**. *PLoS One* 2015, **10**(10):e0140572.
37. Li H, Zhang W, Yan M, Qiu J, Chen J, Sun X, Chen X, Song L, Zhang Y: **Nucleolar and spindle associated protein 1 promotes metastasis of cervical carcinoma cells by activating Wnt/β-catenin signaling**. *J Exp Clin Cancer Res* 2019, **38**(1):33.
38. Guan C, Liu Z, Lu C, Xiao M, Shi H, Ni R, Bian Z: **Nucleolar spindle-associated protein 1 promotes tumorigenesis and predicts poor prognosis in human esophageal squamous cell carcinoma**. *J Cell Biochem* 2019.
39. Roy S, Hooiveld GJ, Seehawer M, Caruso S, Heinzmann F, Schneider AT, Frank AK, Cardenas DV, Sonntag R, Luedde M *et al*: **microRNA 193a-5p Regulates Levels of Nucleolar- and Spindle-Associated Protein 1 to Suppress Hepatocarcinogenesis**. *Gastroenterology* 2018, **155**(6):1951-1966.e1926.
40. Shan S, Chen W, Jia JD: **Transcriptome Analysis Revealed a Highly Connected Gene Module Associated With Cirrhosis to Hepatocellular Carcinoma Development**. *Front Genet* 2019, **10**:305.
41. Yang Z, Li J, Feng G, Wang Y, Yang G, Liu Y, Zhang S, Feng J, Zhang X: **Hepatitis B virus X protein enhances hepatocarcinogenesis by depressing the targeting of NUSAP1 mRNA by miR-18b**. *Cancer Biol Med* 2019, **16**(2):276-287.

Figures

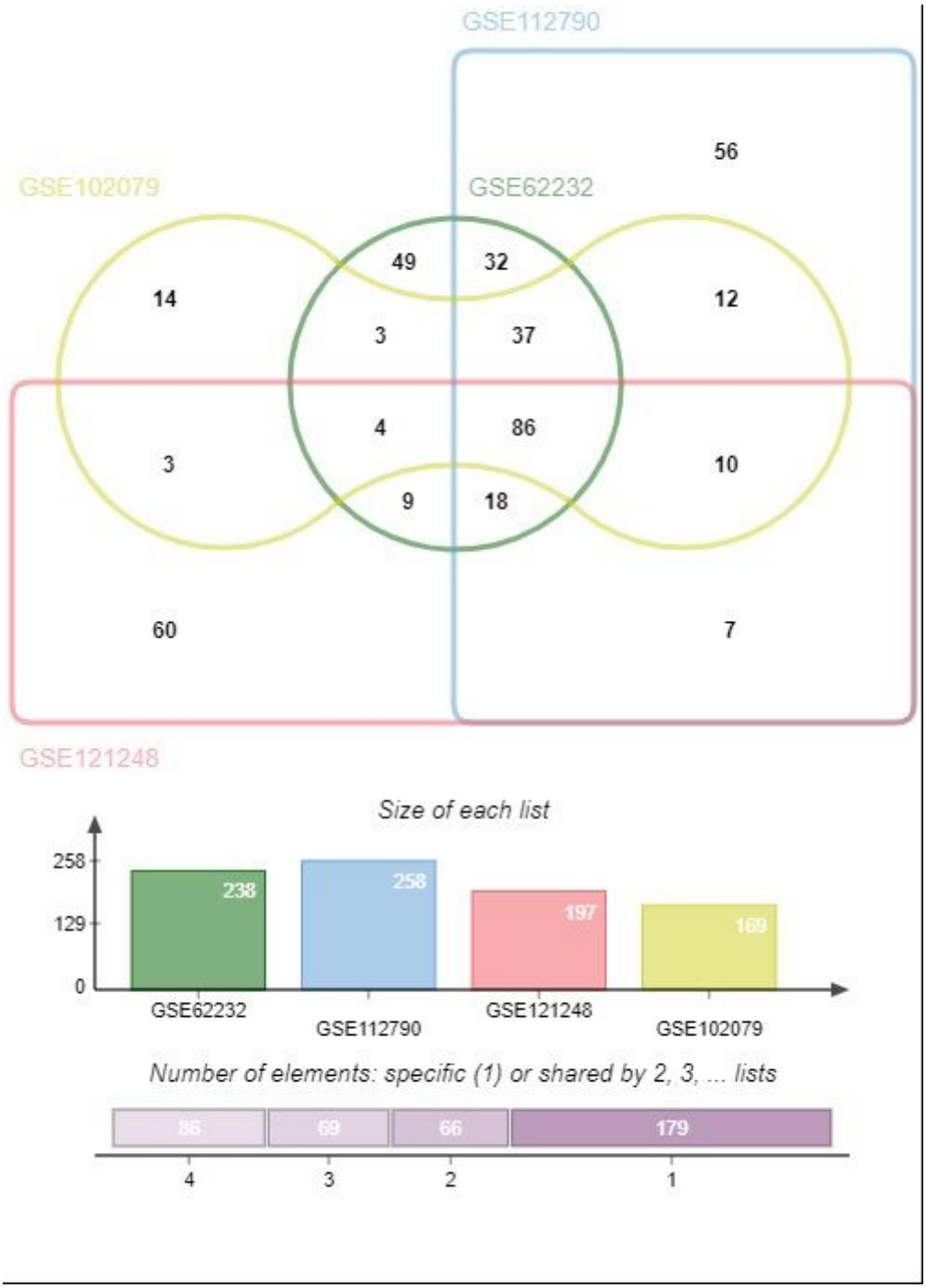

Figure 1

Venn diagram. Identification of overlapping DEGs in four microarray datasets (GSE62232 GSE102079 GSE112790 and GSE121248), the four datasets share 86 overlapping DEGs.

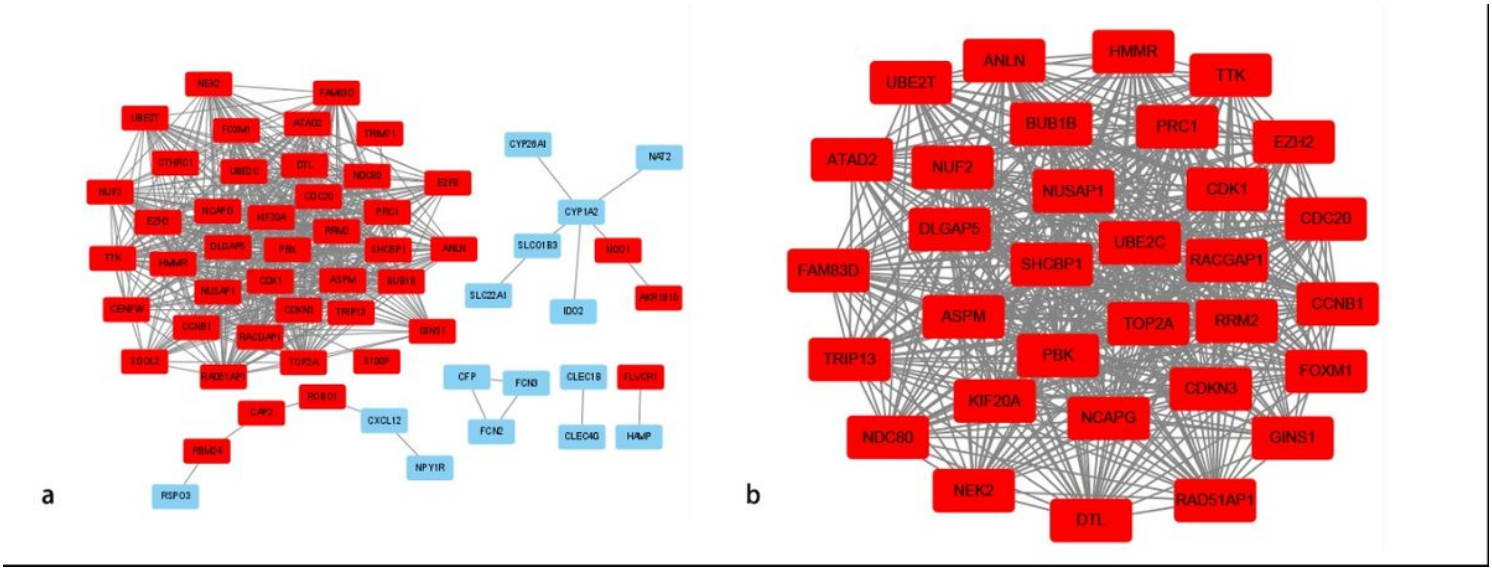

Figure 2

PPI network and the most significant module in PPI network. a. PPI network was constructed using 86 overlapping DEGs (including 55 up-regulated genes and 31 down-regulated genes). Up-regulated and down-regulated genes were respectively marked in red and blue. b. The most significant module of PPI network included 32 nodes and 474 edges. All genes in the most significant module were up-regulated genes and marked in red.

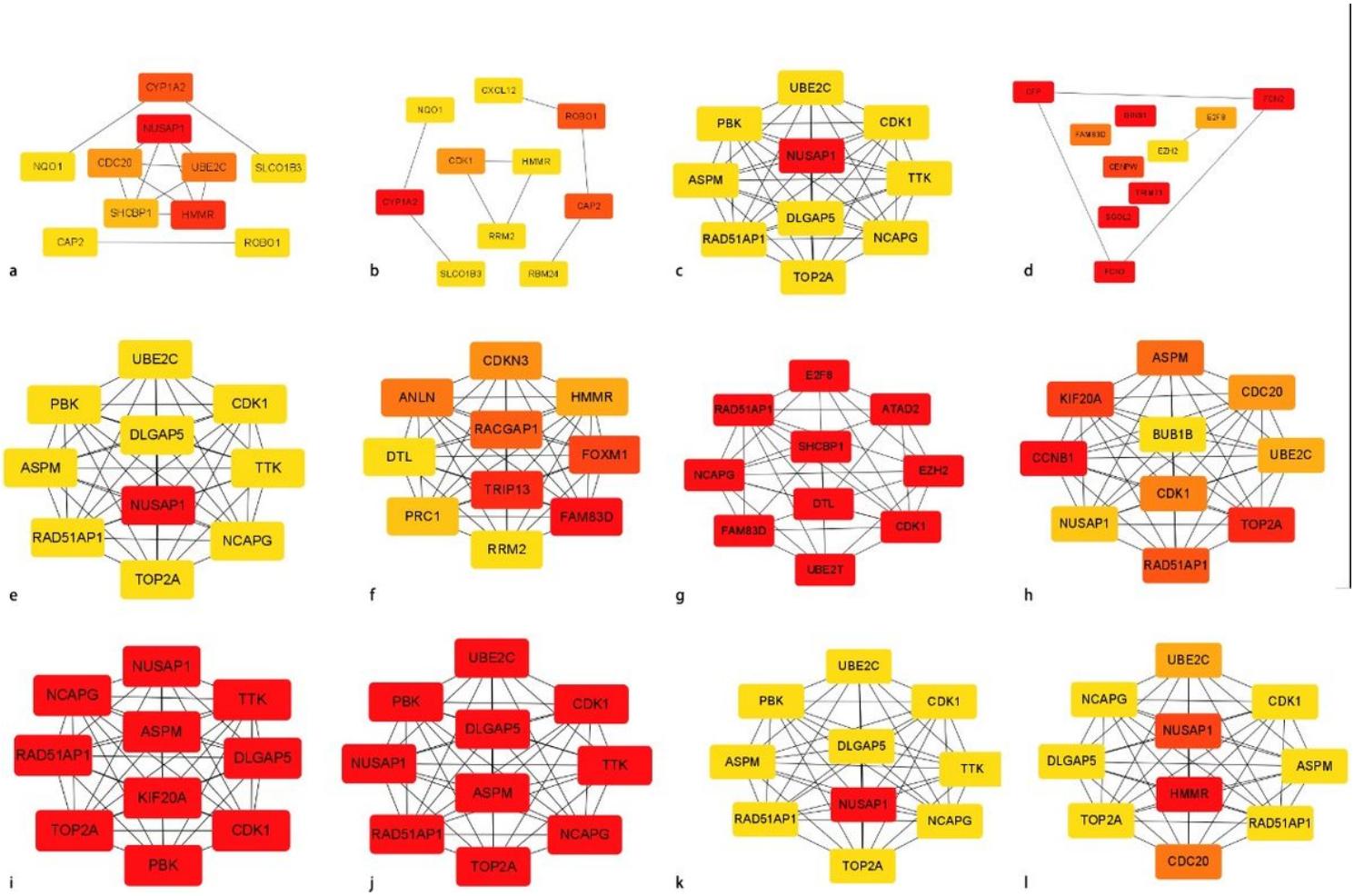

Figure 3

Top 10 hub genes in the PPI network, which were calculated by method Betweenness (a), BottleNeck (b), Closeness (c), ClusteringCoefficient (d), Degree (e), DMNC (f), EcCentricity (g), EPC (h), MCC (i), MNC (j), Radiality (k) and Stress (l). The darker the color, the higher the score.

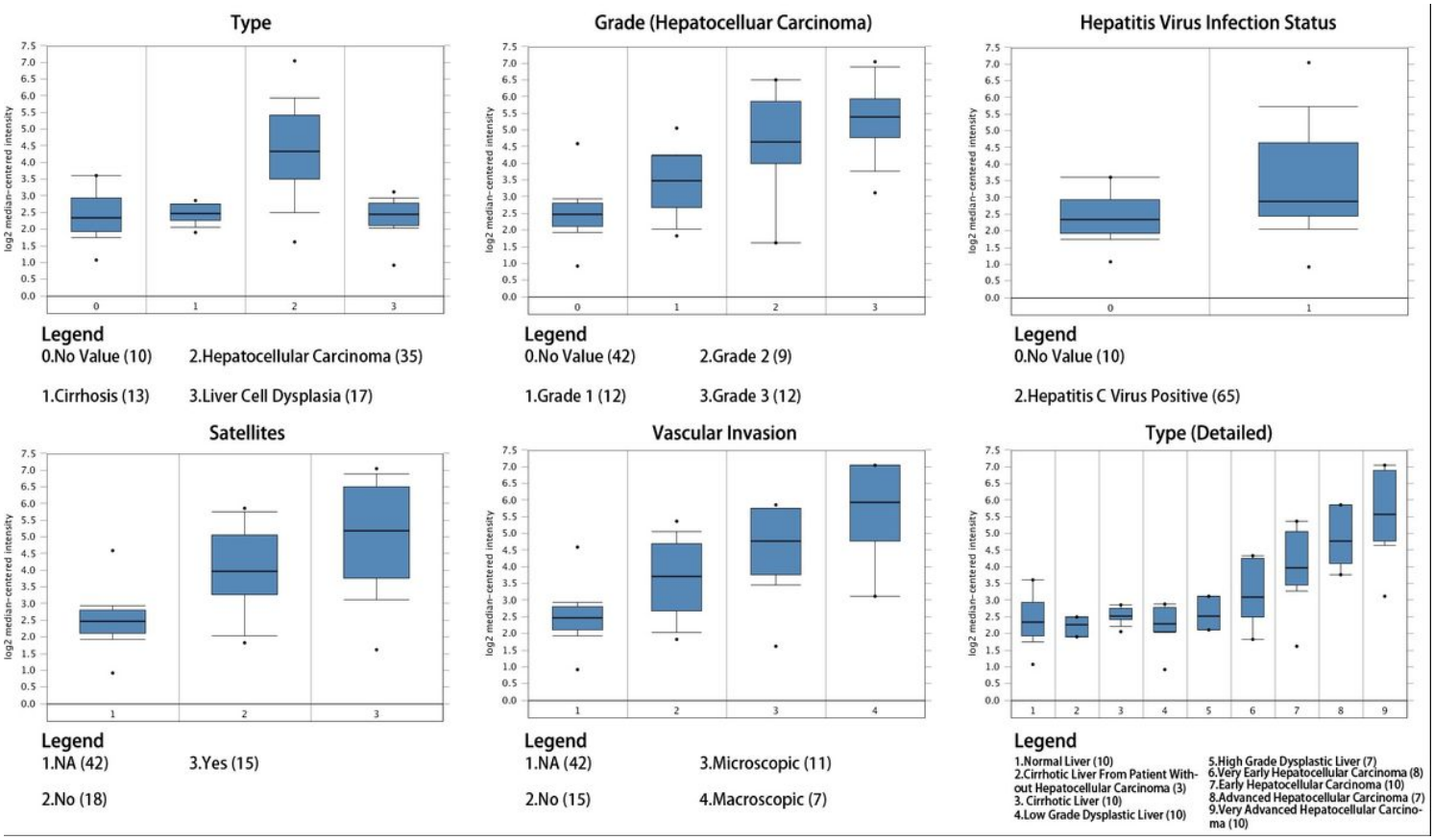

Figure 4

a. Expression of NUSAP1 in normal liver, cirrhosis, hepatocellular carcinoma and liver cell dysplasia in the Wurmbach liver dataset. b-f. Association between the expression of NUSAP1 and tumor grades (b), hepatitis virus infection status (c), satellites (d), vascular invasion (e) and type (detailed) in the Wurmbach liver dataset.

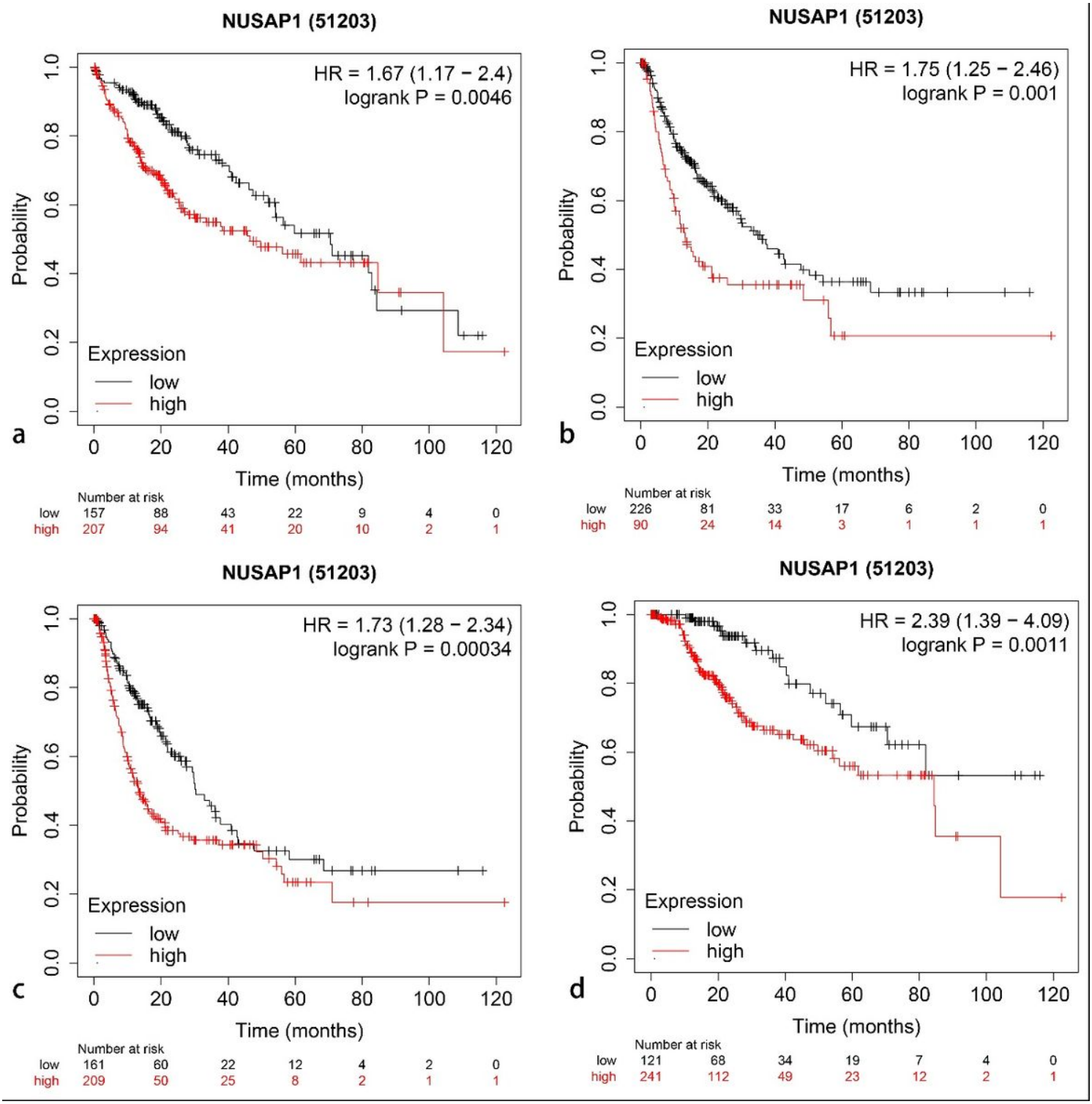

Figure 5

a. Overall Survival (OS) analysis of NUSAP1. b. Relapse Free Survival (RFS) analysis of NUSAP1. c. Progress Free Survival (PFS) analysis of NUSAP1. d. Disease-specific Survival（DSS）analysis of NUSAP1. (All of the four figures were generated by Kaplan–Meier Plotter online tool, $p < 0.05$ was considered statistically significant).

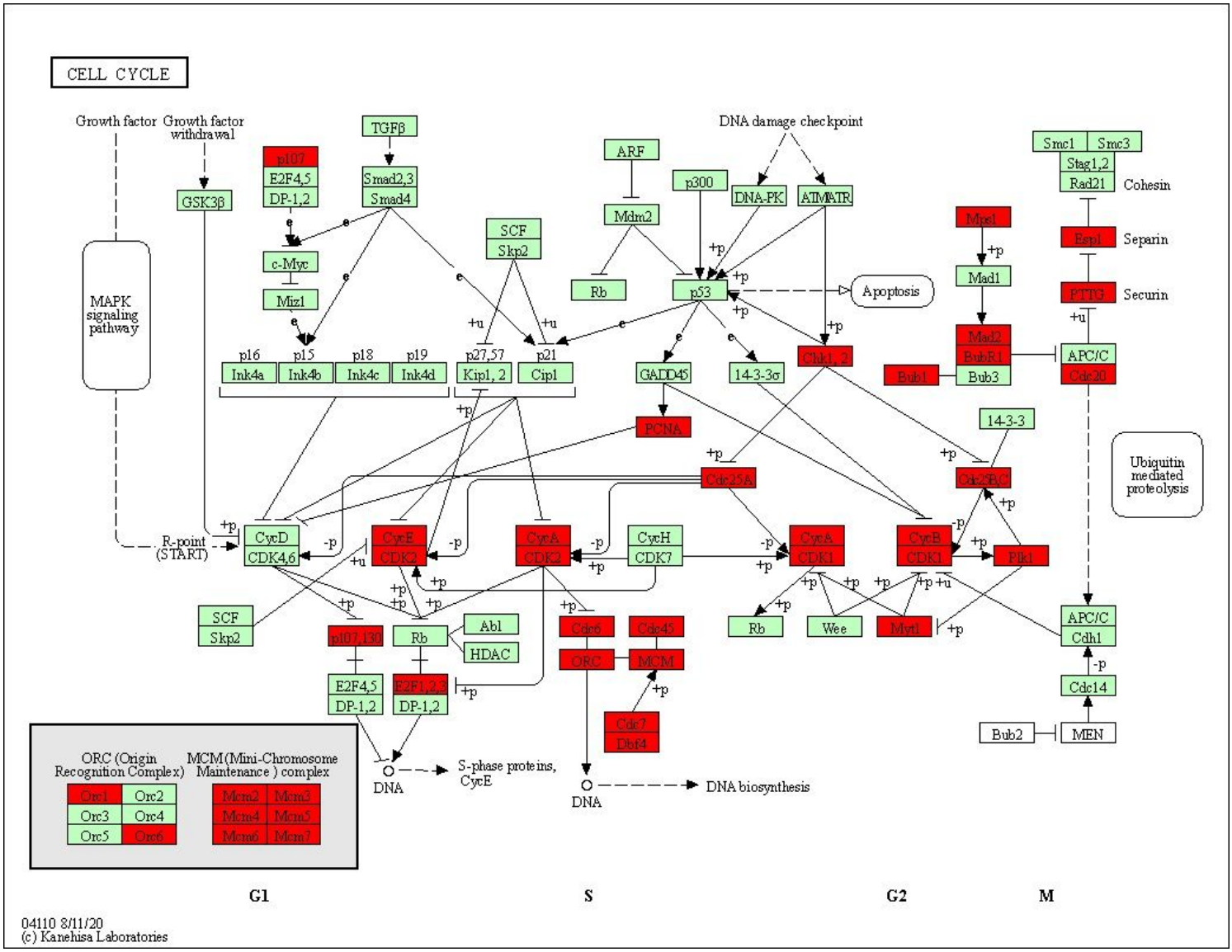

Figure 6

Positional relationships of the co-expression genes of NUSAP1 in the Cell cycle. The positions of the related genes in the Cell cycle pathway were marked with red color, these genes were mainly existing in S phase and G2/ M phase of the Cell cycle.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial1.xlsx
- Supplementarymaterial2.xlsx
- Supplementarymaterial3.xlsx
- Supplementarymaterial4.xlsx