Identification of *SLITRK6* as a Novel Biomarker in hepatocellular carcinoma by comprehensive bioinformatic analysis

Xudong Liu [a,1], Yajie Liu [a,1], Zhe Liu [b], Yu Zhang [a], Ying Ma [a], Jiangshan Bai [a], Hongmei Yao [a], Yafan Wang [a], Xue Zhao [c], Rui Li [a], Xinqiang Song [a,d], Yuxuan Chen [e], Zhiguo Feng [f,**], Lei Wang [a,d,*]

[a] *College of Life Sciences, Xinyang Normal University, Xinyang, 464000, China*
[b] *Department of Computer Science, City University of Hong Kong, Hong Kong, 999077, China*
[c] *College of International Education, Xinyang Normal University, Xinyang, 464000, China*
[d] *College of Medicine, Xinyang Normal University, Xinyang, 464000, China*
[e] *Department of Recovery Medicine, People's Liberation Army 990 Hospital, Xinyang, Henan, China*
[f] *College of Science, Qiongtai Normal University, Haikou, 571127, China*

ARTICLE INFO

ABSTRACT

Hepatocellular carcinoma (HCC) is the most common primary malignancy of the adult liver and morbidity are increasing in recent years, however, there is still no effective strategy to prevent and diagnose HCC. Therefore, it is urgent to research the effective biomarker to predict clinical outcomes of HCC tumorigenesis. In the current study, differentially expressed genes in HCC and normal tissues were investigated using the Gene Expression Omnibus (GEO) dataset GSE144269 and The Cancer Genome Atlas (TCGA). Gene differential expression analysis and weighted correlation network analysis (WGCNA) methods were used to identify nine and 16 key gene modules from the GEO dataset and TCGA dataset, respectively, in which the green module in the GEO dataset and magenta module in TCGA were significantly correlated with HCC occurrence. Third, the enrichment score of gene function annotation results showed that these two key modules focus on the positive regulation of inflammatory response and cell differentiation, etc. Besides, PPI network analysis, mutation analysis, and survival analysis found that SLITRK6 had high connectivity, and its mutation significantly impacted overall survival. In addition, *SLITRK6* was found to be low expressed in tumor cells. To summarize, SLITRK6 mutation was found to significantly affect the occurrence and prognosis of HCC. *SLITRK6* was confirmed as a new potential gene target for HCC, which may provide a new theoretical basis for personalized diagnosis and chemotherapy of HCC in the future.

1. Introduction

Hepatocellular carcinoma (HCC) is the most common primary malignancy in adults, accounting for approximately 90% of primary liver cancers, and is one of the most common malignancies worldwide [1]. Its high metastatic ability and high recurrence rate lead to a low survival rate and poor overall prognosis [2]. According to statistics, there are nearly one million additional cases of HCC, and the death toll exceeds 1 million related to HCC in the world every year, of which China accounts for 50% of the total cases and deaths [3]. At present, the potentially curative treatments for HCC, such as liver transplantation, tumor resection, or ablation, are limited to tumors [4,5]. Prevention and early diagnosis of hepatocellular carcinoma remain an urgent problem, and there is no effective prevention strategy for hepatocellular carcinoma , it is urgent to research the effective biomarker to predict clinical outcomes of HCC tumorigenesis and propose new strategies for gene-targeted therapy.

In recent years, functional genomics based on bioinformatics has been developing continuously [6]. Numerous tumor biomarkers have been discovered and applied in the clinic, significantly improving tumor

---

* Corresponding author. College of Life Sciences, Xinyang Normal University, Xinyang, 464000, China.
** Corresponding author.
E-mail addresses: ffzzgg@126.com (Z. Feng), wangleibio@126.com (L. Wang).
[1] These authors contributed equally.

prevention and early diagnosis and increasing overall patient survival [7]. In the present, two databases, Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA), have collected data from different chip platforms that have been used to investigate the molecular mechanisms underlying multiple tumorigeneses [8].

Previous studies have proven that HCC phenotypes are strongly associated with specific gene mutations [9]. The *TERT* promoter increases the cancer incidence by being affected by copy number changes that can mutate before cancer development. In HCC, TP53 and CTNNB1 are the next most commonly mutated genes, with 25%–30% of these patients having mutations in these genes [10–12]. In summary, we believe that it is feasible further to reveal the pathogenesis of HCC from gene mutations.

In the current study, we integrated and analyzed HCC RNA-Seq datasets from GEO and TCGA data by combining multiple bioinformatics methods and techniques. Weighted gene co-expression network analysis (WGCNA) was performed on the two groups of datasets respectively to screen out the key modules related to HCC pathogenesis. We combine gene function enrichment analysis, protein-protein interaction (PPI) analysis, and gene survival analysis to detect and analyze key module genes. Finally, *SLITRK6* was proved as a potential tumor prediction biomarker for diagnosing and treating HCC in combination with immunohistochemical validation of the Human Protein Atlas database (HPA). It may provide a reliable basis for determining molecular targets for personalized diagnosis and chemotherapy of HCC in the future.

2. Materials and methods

2.1. Acquisition of liver cancer numbers from TCGA database and GEO database

The design of this study is shown as a flow chart in Fig. 1. We downloaded transcription profiles for GSE144269 and GSE105130 that were acquired from GEO (https://www.ncbi.nlm.nih.gov/gds), and the gene expression RNA-Seq of TCGA-LIHC from the UCSC Xena database (https://xena.ucsc.edu/) [13,14] Next, The principal component analysis (PCA) was used to reduce the dimensionality of the obtained data. The ggfortify (version: X64 4.0.3) and factoextra (version: X64 4.0.3) R packages were selected to reduce the dimensionality of the original data [15].

2.2. Gene differential expression analysis

The DESeq2 package of R software (version: X64 4.0.3) was used to calculate the expression value of the sorted expression matrix of GEO dataset GSE144269 (GSE144269_VOOM_GENEEXPRESSION). The selection criteria of differentially expressed genes (DEGs) were chosen by $|\text{Log}_2(\text{FC})| \geq 1$ and *P-value* < 0.05 [16]. Similarly, the DESeq2 package of R software was used to identify DEGs in the TCGA-LIHC, and the screening criteria were $|\text{Log}_2(\text{FC})| \geq 1$ and *P-value* < 0.01. Two sets of DEGs volcano maps were constructed by ggplot2 package (version: X64 4.0.3) of R software [17].

2.3. Weighted correlation network analysis

DEGs from both gene sets were used in the WGCNA R language
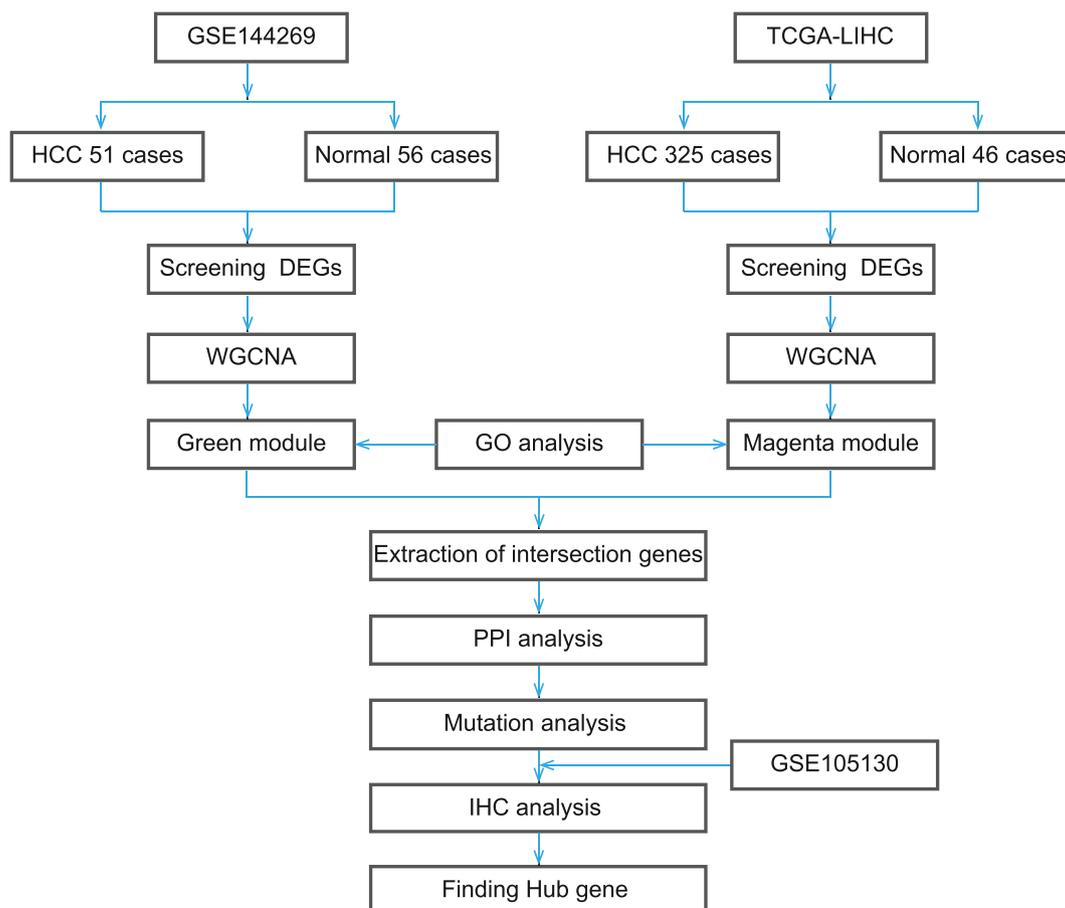
Fig. 1. Flow diagram of this research.

package to perform Weighted gene co-expression network analysis [18]. The power value was calculated by the pink soft-Threshold function of the WGCNA package. Next, we identify the gene modules by using hierarchical clustering with a criterion of at least 30 genes per module (minModuleSize = 30) [19]. As in previous studies, the hclust criterion was used to perform hierarchical clustering. The branches of the hierarchical clustering tree correspond to modules. The clinical data was applied to the WGCNA and used as the basis for the dynamic tree-cutting algorithm used to segment the network module [20]. As we know, the different module signature genes (MEs) are computationally correlated with clinical traits to some extent. The significance of each gene with the selected clinical trait is calculated and quantitatively associated by gene significance (GS), and the module affiliation (MM) of the association between each module's signature genes and the liver cancer gene expression profiles is also calculated. GS within the module on behalf of the correlation coefficient between gene expression level and clinical characteristics, while MM represents the correlation coefficient between gene expression level and gene principal component expression level. We believe that when GS and MM are highly correlated, the most important elements of the module also have a strong association with the selected clinical trait. In this case, we selected the module and used it to construct co-expression networks and identify Hub genes [21].

2.4. Functional annotation of key gene modules based on WGCNA

GO annotation consists of three main parts: biological process (BP), cellular component (CC), and molecular function (MF) [22]. The genes of the key modules were extracted, and the functions of the genes of the key modules of HCC were annotated using DAVID (https://david.abcc.ncifcrf.gov/). The *P-value*<0.05 was considered as significant enrichment [23,24].

2.5. Construction of PPI and screening of Hub genes

Protein-Protein interaction (PPI) network analysis can predict the function of interacting proteins and provide functional relationships of protein interactions [25]. The Search Tool for the Retrieval of Interacting Genes (STRING) database (http://string-db.org) collects and aggregates all publicly available protein-protein interaction information. The genes of key modules were extracted, and the protein and PPI network information data were retrieved using STRING [26]. In addition, to explore the relationship between hub genes, the STRING database and the Cytoscape (http://cytoscape.org/;version 3.7.2) software were used to convert the results. Cytohubba, a plug-in in Cytoscape, was used to select the hub genes with high connectivity in the gene expression network for subsequent analysis based on the top20 score of the Maximum Clique Centrality (MCC) algorithm [27,28].

2.6. Modular gene mutation analysis

Gene mutation is one of the causes of the occurrence and development of cancer, and it is worth exploring the types, mutation sites, and survival effects of a gene mutation for revealing the pathogenesis of hepatocellular carcinoma [29]. We extract the top 20 genes, using a database of 442 cases of hepatocellular carcinoma (TCGA, Firehose Legacy) data in the cBioportal (http://www.cbioportal.org/), to determine the type and frequency of hepatocellular carcinoma (HCC) gene mutations, with mutation rate>1% is the screening criterion [30]. The Cbioportal database online analysis tool was combined to visualize the mutation sites, mutation types, and survival effects of genes with high mutation rates [31].

2.7. Differentiating performance of the prognostic signature

We used the R package of pROC to perform the ROC curve analysis of the hub gene [32].To construct receiver operating characteristic (ROC)

and liver cancer risk scores for normal tissue, compare normal tissue and liver cancer risk scores, and explore the differential diagnostic ability of key genes. The ROC diagnostic curve is used to investigate each feature's prognostic or predictive accuracy under the area under the curve (AUC). AUC>0.5 indicates that the difference has diagnostic significance [33, 34].

2.8. Immunohistochemical verification

The Human protein atlas database (HPA) (https://www.proteinatlas.org), contained a large amount of a particular organization transcriptome and proteomics data. It is composed of an organization and pathology atlas [35]. Immunohistochemistry (IHC) in HPA was used to determine the protein expression level of genes related to survival between HCC and normal tissues, and Immunohistochemical essays are one of the most commonly used methods to detect the location and abundance of protein expression [36,37].

3. Results

3.1. Samples screening and differentially expressed gene identification

The general process of screening for differential genes is shown in Fig. 2A. Principal component analysis (PCA) was used to cluster samples and remove outliers from the first and second components (Fig. 2C and D). There are 51 tumor samples, and 56 normal samples were screened from GSE144269 (Fig. 2B).Next, We used |Log$_2$(FC)|≥1, *P-value*<0.05 as the cutoff criteria.Overall, 8292 DEGs were extracted from GSE144269, including 5322 upregulated genes and 2970 downregulated genes (Fig. 2E). Similarly, 325 cancer samples and 46 paracancer samples were selected from the TCGA-LIHC dataset (Fig. 2B). The cutoff value was |Log$_2$(FC)|≥1, *P-value*<0.01. A total of 8717 DEGs were extracted from TCGA-LIHC, including 6673 upregulated genes and 2044 downregulated genes (Fig. 2F).

3.2. Construction of gene co-expression module

The initial soft threshold method was used to implement our WGCNA method. We first evaluated the reliability of the network, but no outlier samples were detected that needed to be removed. We analyze the network topology of the threshold power from 1 to 30 and determine the scale independence and average connectivity of the relative balance of the WGCNA (Fig. S1). The lowest power fitting index of the scale-free topology is 0.85. Next, power is determined to be 3. Check scale-free topology scale of TCGA-LIHC (R^2 = 0.86, slope = −1.36) and check scale-free topology scale of GSE144269 (R^2 = 0.9, slope = −1.55) (Fig. S2). The distribution approximates the linear scale-free topology, and the fitting degree is high, which proves that the power selection is appropriate. We respectively constructed 8292 DEGs hierarchical clustering trees of the GSE144269 dataset and 8717 DEGs hierarchical clustering trees of the TCGA-LIHC dataset. Then we set mergeCutheight (the threshold for merge modules) to 0.15 to merge similar modules. TCGA-LIHC generates 16 modules, and GSE144269 generates nine modules (Fig. 3A and B). Compared with other modules, the green module in GSE14469 and the magenta module in TCGA-LIHC have a higher correlation with normal state and tumor state (Fig. 3C and D), suggesting that the green module and magenta module may play an important role in the occurrence and development of liver cancer and need to be further analyzed. In order to understand the interaction between the genes contained in the module, we obtained the genes of two key modules.

3.3. Functional enrichment analysis of key-related modules

The green and magenta modules are most associated with tumor status. There was a high correlation between the number of module
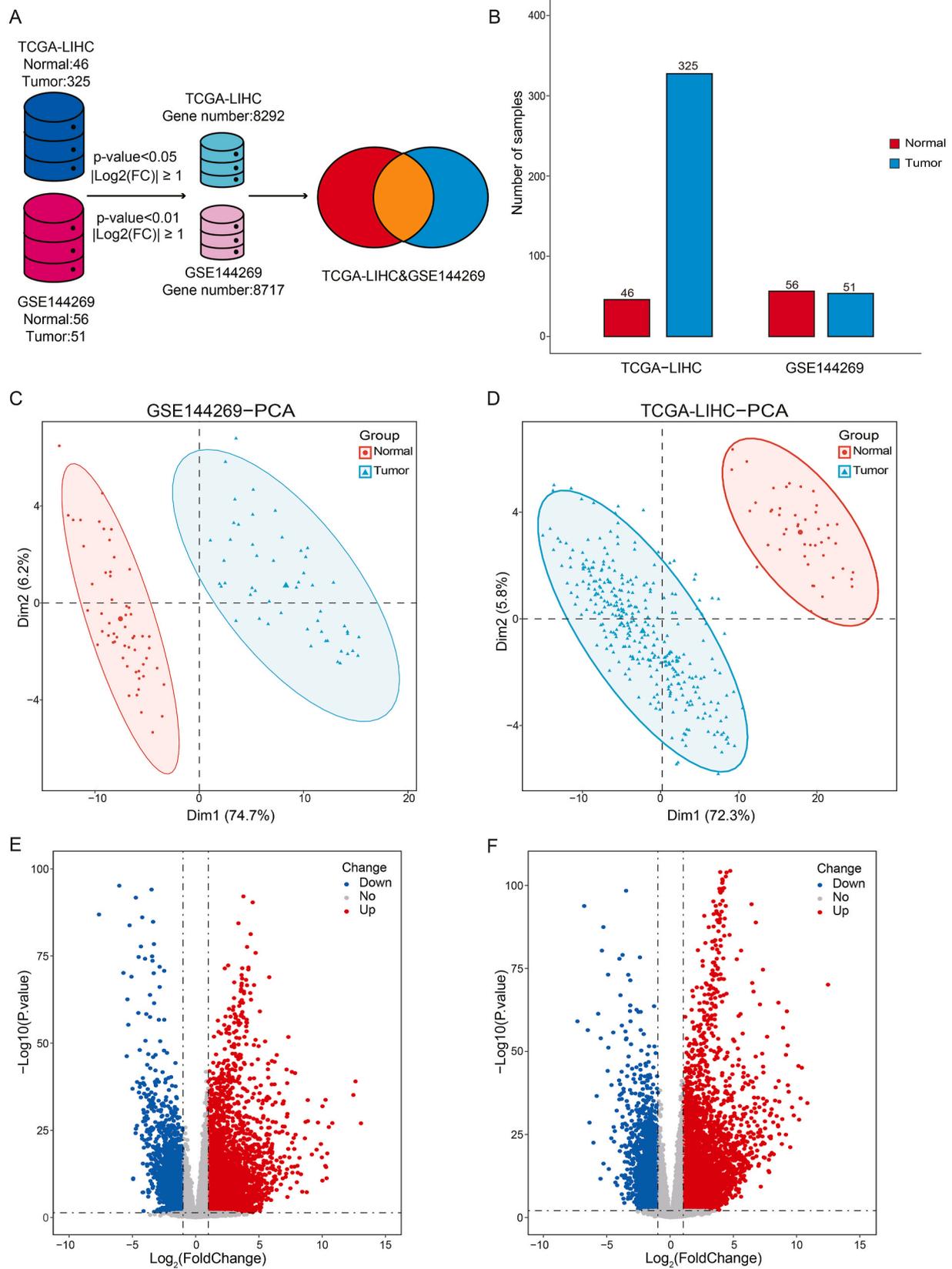
Fig. 2. Screening sample and Identification of DEGs. (A) Flowchart genetic difference analysis. (B) Bar chart of sample size. red: Normal, blue: Tumor. (C, D) Principal component analysis of GSE144269 samples and TCGA-LIHC samples. (E, F) The volcano plots of DGEs for dataset GSE144269 and TCGA-LIHC red: Up; blue: Down; gray: No. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
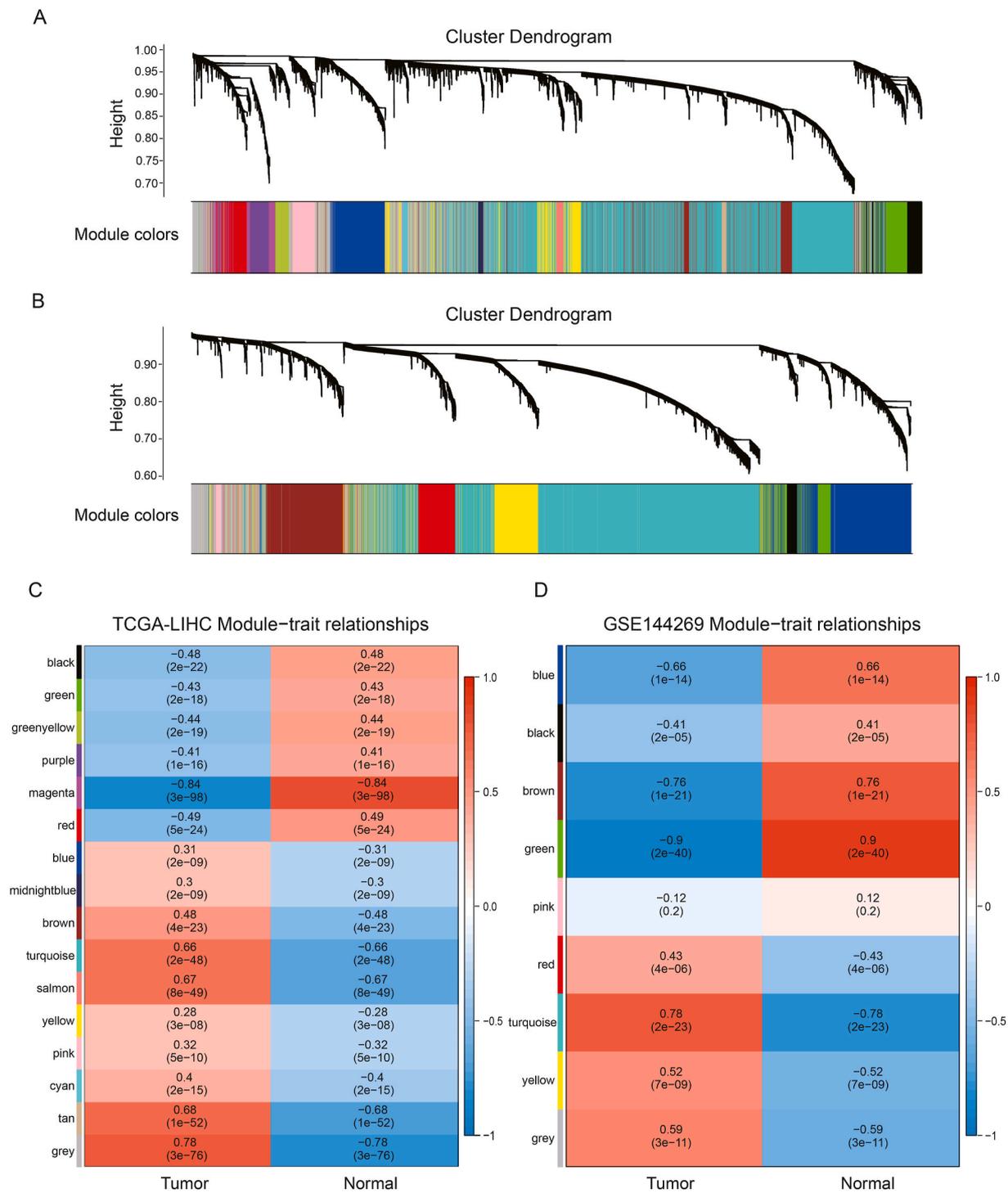
Fig. 3. Identification of modules associated with the tumorigenesis of HCC. (A, B) The cluster dendrogram of genes in TCGA-LIHC and GSE144269. (C, D) Module-trait relationships of TCGA-LIHC and GSE144269.

members in the magenta module and gene significance (cor = 0.93, p = 2.8e-113). There was a high correlation between module member number and gene significance in the green module (cor = 0.93, p = 1e-200), which indicated that these modules were suitable for identifying HUB genes related to normal and tumor status (Fig. 4A and B). To further understand the function of the key modules, we upload all the genes in the green module and magenta module to the David database for GO analysis. The GO analysis results of the green module and magenta module are shown in Fig. 4C and D. The green module is mainly related to the negative regulation of growth, cell differentiation, and

cellular response to erythropoietin (Fig. 4C). The magenta module is mainly related to positive regulation of inflammatory response, recognition of apoptotic cells, and retinoic acid metabolic process (Fig. 4D).

3.4. PPI analysis and mutation analysis of genes

The 48 overlapping genes were introduced into the STRING database, and Cytoscape and its plugin (cytoHubba) were applied to remove the isolated genes that did not interact with each other. PPI results showed that the protein interaction network constructed by 48 genes
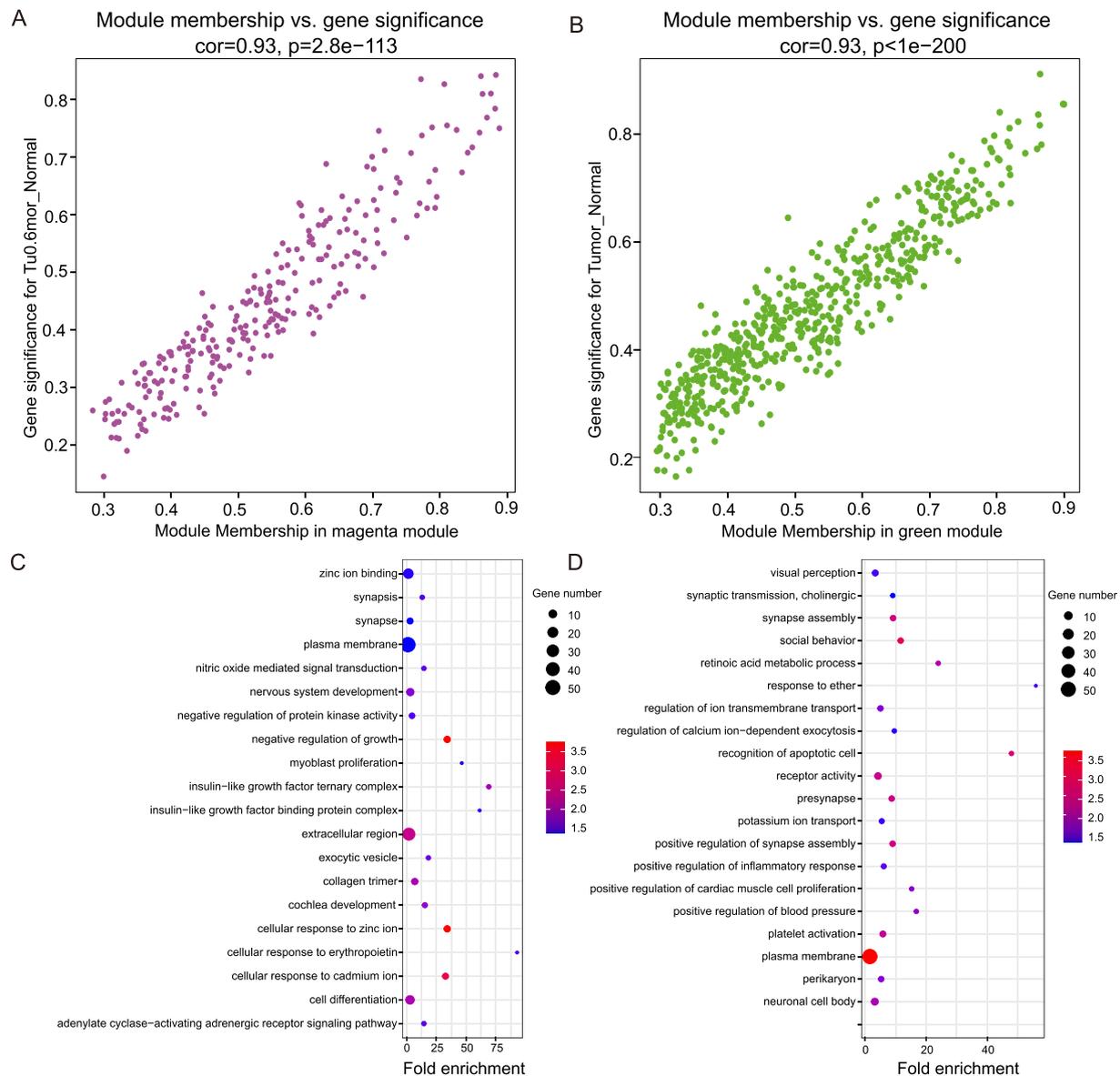
Fig. 4. Enrichment analysis for module eigengenes. (A) The scatter plot between the magenta module membership and GS for tumor. (B) The scatter plot between the green module membership and GS for tumor. (C) GO analysis for the green module. (D) GO analysis for the magenta module. (E) Venn diagrams of the green and magenta module genes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

had 37 nodes and 71 edges. The top 20 genes were screened out by the MCC algorithm for further analysis (Fig. 5A). Combined with the GEPIA database (http://gepia.cancer-pku.cn/detail.php), compared with the normal samples, the expression of these 20 genes were all down-regulated in HCC tissues, among which *IGF1*, *IGF2*, *IGFB3*, *INS-IGF2*, *NPY1R*, and *SYT9* showed the most significant differences (*P-value*<0.05) (Fig. S3). We extracted the top 20 genes and evaluated their mutation frequency using cBioPortal. A total of 12 mutant genes were screened using mutation rate >1% as screening standard. Including *SLITRK6*(5%), *DSCAM*(4%), *GPM6A*(3%), *ID2*(2.7%), *MASP1*(2.7%), *SLC17A8*(2.5%), *ADAM22*(2.5%), *CHRM2*(2.5%), *SYT10*(1.6%), *IGBP3* (1.4%), *NPY1R*(1.4%) and *LGI1*(1.4%). The main types of mutations include inframe mutation, missense mutation, truncating mutation, amplification, and deep deletion(Fig. 5B). We also detected the heat map of their mRNA expression relative to the z-scores of diploid samples and the mutation spectrum of each sample. The mRNA expression heatmap of each sample is shown in Fig. 5C and the mutation spectrum in Fig. 5B. Except for *ID2* and *MASP1*, the mRNA expression levels of the other ten genes were significantly decreased in most samples.

3.5. SLITRK6 mutation survival analysis and expression analysis

The main mutation types of *SLITRK6* include missense and truncating, and the mutation sites are shown in Fig. 6A. Its mutation is concentrated in the LRR_8 (Leucine-rich repeat) region. The prognostic value of 12 gene mutations in HCC was analyzed by a KM curve. The results showed that SLITRK6 (P = 6.202e-3) had a significant effect on the overall survival of HCC. The total survival rate of the altered group was lower than that of the unaltered group (Fig. 6B). The expression level of *SLITRK6* in the GSE105130 database is shown in Fig. 6C. The results showed that the expression of *SLITRK6* in tumor tissues was significantly different from that in normal tissues (p < 2.9e-09), and the expression in tumor tissues was significantly down-regulated (Fig. 6C). ROC analysis of normal and tumor diagnosis prediction of *SLITRK6* on TCGA-LIHC and GSE144269 data sets showed that the AUC value of the GSE144269 database was 0.874, and that of the TCGA-LIHC database was 0.941 (Fig. 6D and E). The diagnostic effect is good.
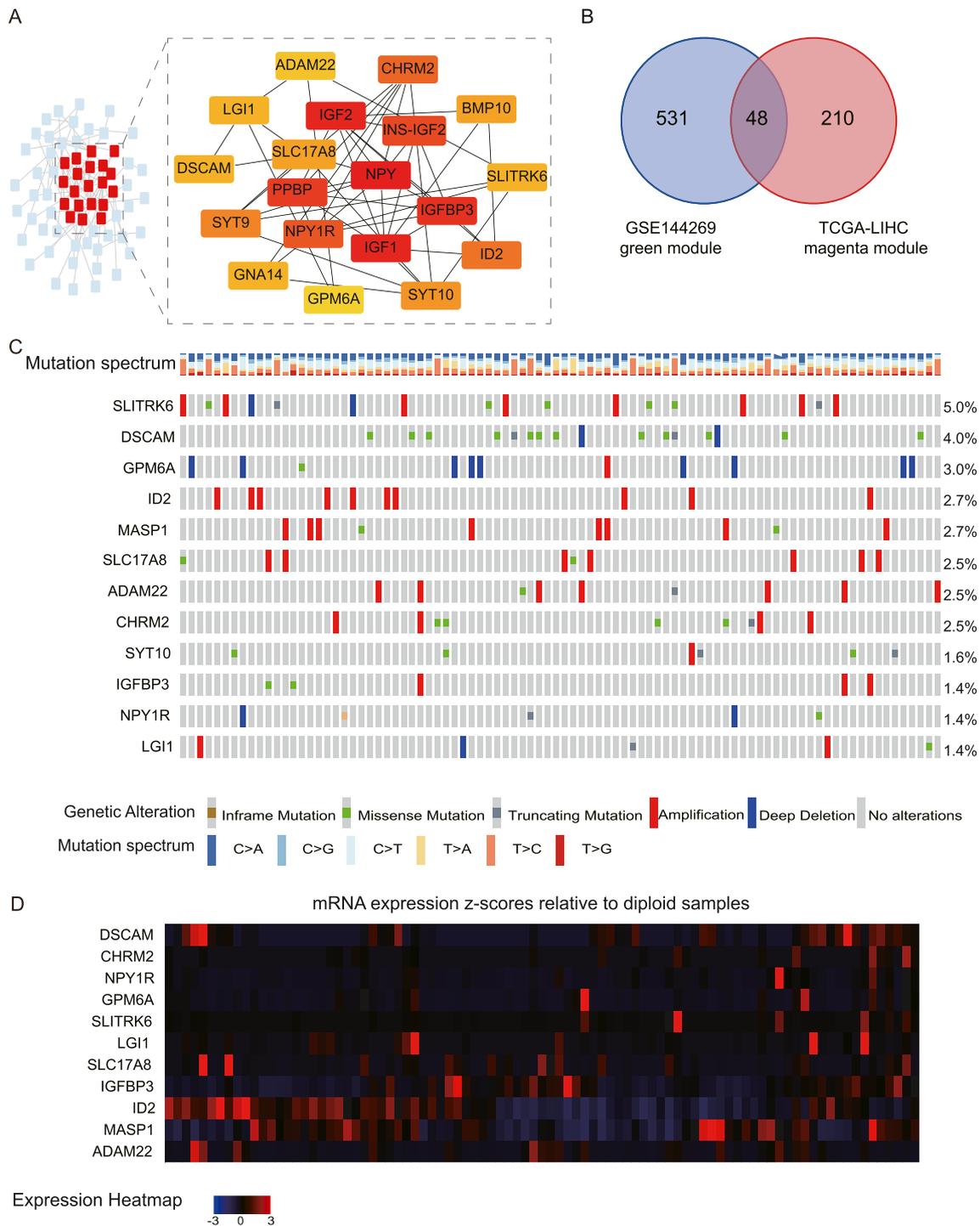
Fig. 5. PPI analysis and mutation analysis of 48 DEGs. (A) The 20 genes with the highest MCC score in the PPI network. (B)The Oncopprint schematic shows the mutation types and mutation rates of *SLITRK6, DSCAM, GPM6A, ID2, MASP1, SLC17A8, ADAM22, CHRM2, SYT10, IGFBP3, NPY1R,* and *LGI1* in 366 samples of mutation and CNA data. (C)mRNA expression z-scores are relative to diploid samples. Including *SLITRK6, DSCAM, GPM6A, ID2, MASP1, SLC17A8, ADAM22, CHRM2, IGFBP3,* and *NPY1R*; *SYT10* temporarily has no data.

3.6. Immunohistochemical analysis

Immunohistochemical staining obtained from the human protein map database also confirmed the expression level of SLITRK6 in HCC. Notably, the protein level of *SLITRK6* was highly expressed in tumor tissues (Fig. 7).

4. Discussion

Hepatocellular carcinoma (HCC) is one of the most important malignant tumors of the liver. Although new and advanced treatments have emerged in recent years, the survival rate of liver cancer is still very low, and the prognosis is very poor [38]. Therefore, better biomarkers are needed to determine the specific prognosis and progression of HCC. Gene mutation is one of the causes of the occurrence and development of cancer [29]. In the present study, the type and site of gene mutation and
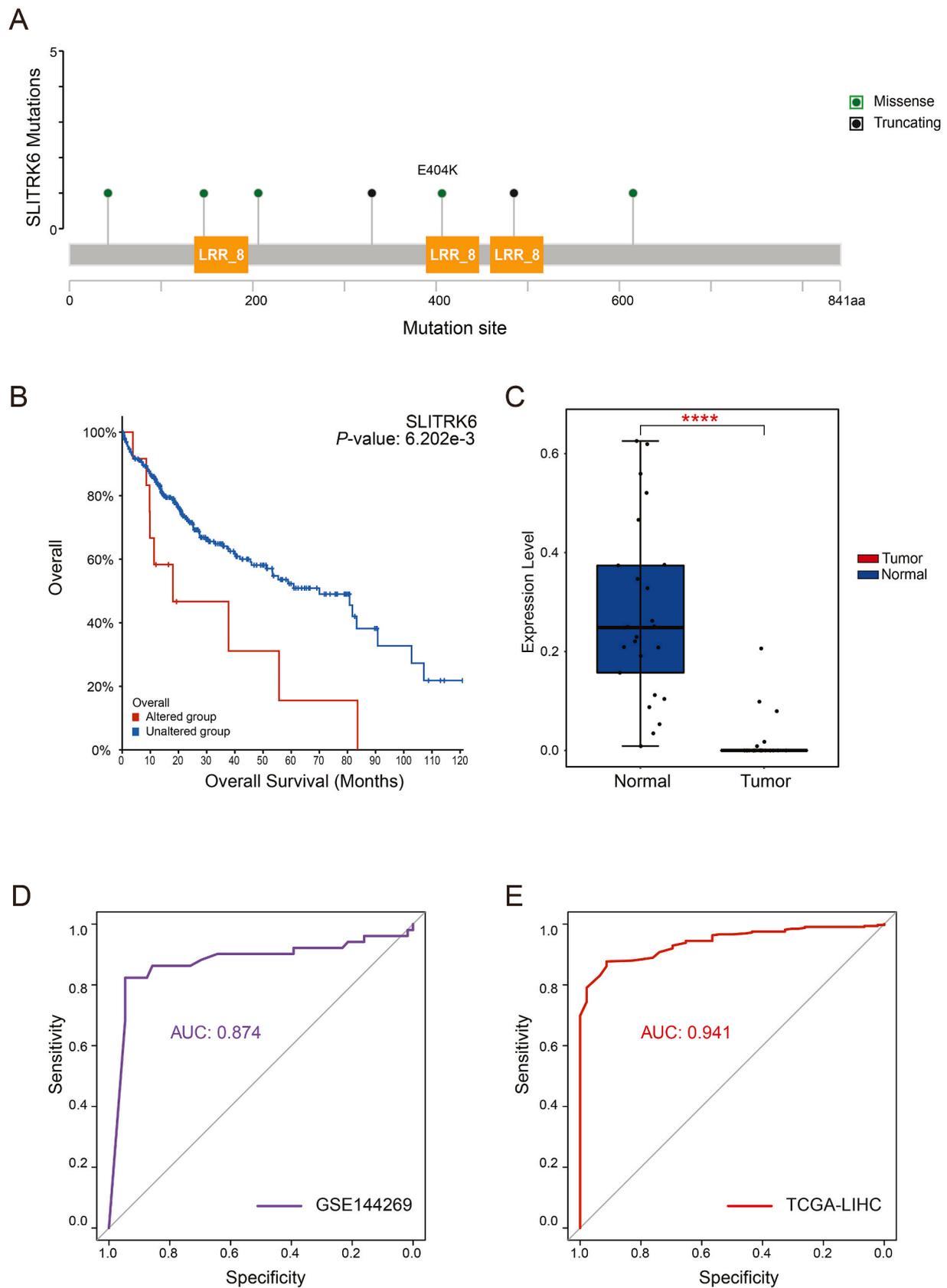
Fig. 6. Overall survival analysis and gene expression of *SLITRK6*. (A) Mutation sites and types of *SLITRK6;* LRR_8:Leucine rich repeat. (B) Mutation survival analysis of *SLITRK6*; *P-value* = 6.202e-3. (C) Expression boxplot of *SLITRK6* in GSE105130 dataset; *P-value* < 0.05. (D) The ROC analysis of the GSE144269 dataset for normal and tumor diagnosis prediction by *SLITRK6*; AUC = 0.874. (E) The ROC analysis of TCGA-LIHC dataset for normal and tumor diagnosis prediction by *SLITRK6*; AUC = 0.941.

A

SLITRK6
Normal tissue
antibody :HPA014491
staining: Not detected
intensity: Negative
quantity: None

B

SLITRK6
Tumor tissue
antibody :HPA014491
staining: High
intensity: Strong
quantity: 75%-25%
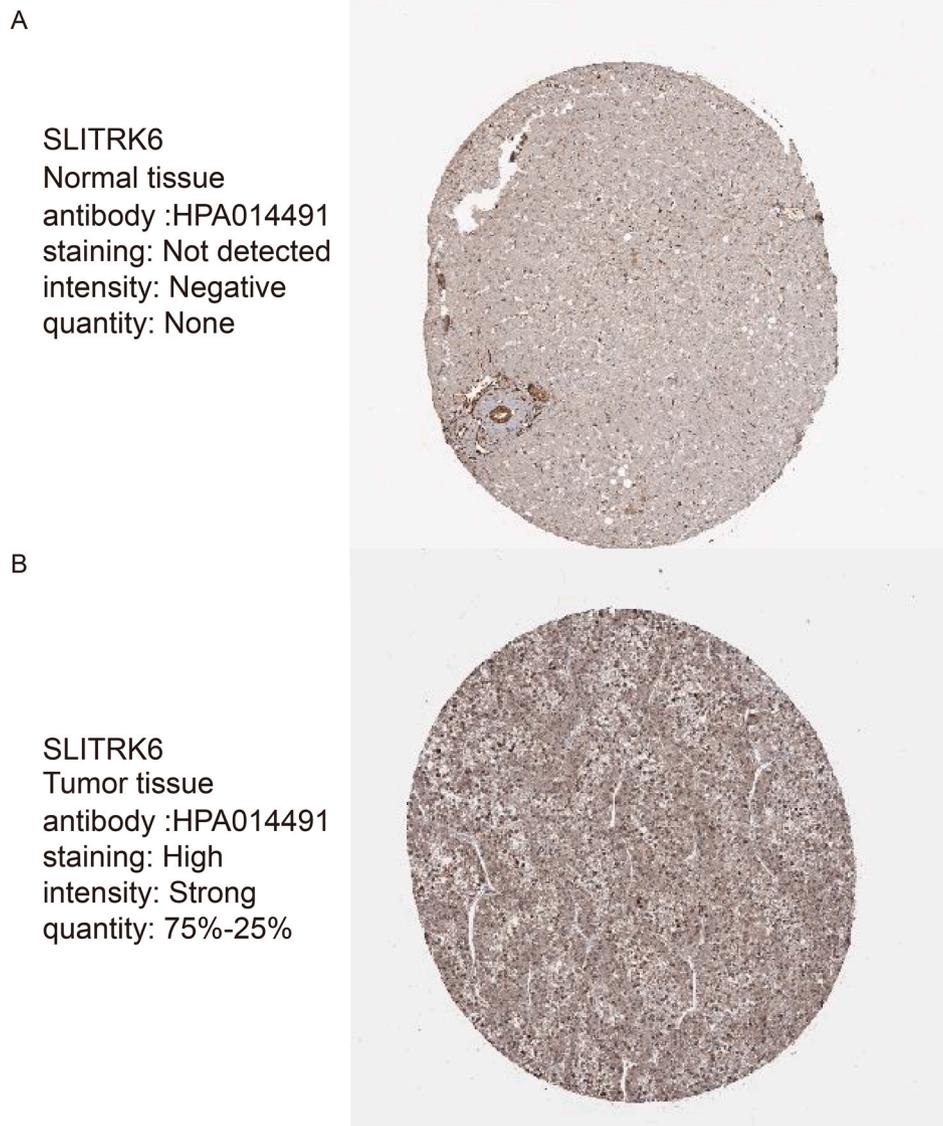
Figure 7. Immunohistochemistry of *SLITRK6* in HCC and para-cancer tissues from the HPA database. (A) Normal antibody :HPA014491. (B) **Para-cancer** antibody :HPA014491.

its effect on survival were discussed, and it was further confirmed that gene mutation plays an important role in the diagnosis and prognosis of liver cancer.

Firstly, We used PCA to conduct dimensionality reduction analysis on the original data, removing outliers in all samples. Next, 46 normal and 325 cancer samples were screened out in the TCA-LIHC data set, and 56 normal samples and 51 cancer samples were screened out of the GSE144269 samples. The differentially expressed genes were verified by R packet DESeq2.

Next, WGCNA analysis identified two important modules in the TCGA-LIHC and GSE144269 dataset, associated with HCC pathogenicity. Compared to other methods, WGCNA has several significant advantages. Because the WGCNA analysis made the transition from single-gene studies to multi-gene studies and explored the relationship between clinical characteristics and co-expression modules, its results were more reliable and had greater biological significance [39].

To study the interaction between various genes and their mechanisms, we extract DEGs in two groups of data, respectively constructed two gene co-expression networks, and combining with clinical data; building blocks the correlation between genetic and clinical characteristics of heat maps. Among them, the magenta module of TCGA-LIHC

and the green module of GSE144269 are closely related to the clinical characteristics of the tumor. The module with the highest correlation with clinical characteristics was considered to be the key module to explore the main causes of disease progression [40].

Functional enrichment analysis showed that the magenta module was significantly related to apoptosis and inflammatory response. In GSE144269, the green module was mainly related to the negative regulation of growth and cell differentiation. These results suggested that abnormal cell differentiation and growth inhibition may be the potential mechanism of HCC.

By extracting 48 intersecting genes from two highly correlated modules, we analyzed and summarized the two sets of data. In addition. the protein interaction network of 48 genes was constructed by combining the STRING database, and we found that they had high interaction relationship. By introducing Cytoscape software and its cytoHubba plugin, the top 20 Hub genes with the highest MCC scores were screened by MCC algorithm, which may play a vital role in regulating the protein network. We can study these genes in more detail.

Combined with the GEPIA database, we found that compared with normal samples, the expression levels of these 20 genes were down-regulated in cancer samples, among which *IGF1*, *IGF2*, *IGFBP3*, *INS-*

IGF2, NPY1R, and *SYT9* were down-regulated most significantly (*P-value*<0.05). To study whether the mutations of these Hub genes are associated with the pathogenesis of cancer, we screened 12 highly mutated genes using sample data from the cBioportal database, including *SLIRK6, DSCAM, GPM6A, ID2, MASP1, SLC17A8, ADAM22, CHRM2, SYT10, IGFBP3, NPY1R,* and *LGI1*. A gene with a mutation rate greater than 1% is defined as a highly mutated gene. The main types of mutations include inframe mutation, missense mutation, truncation mutation, amplification, and deep deletion.

Using survival analysis, we further examined the effect of SLITRK6 hypermutation on overall survival. The results showed that the *SLITRK6* mutation had a significant effect on the overall survival of patients (*P-value* = 6.202e-3). The mutation sites of *SLITRK6* were mainly concentrated in the leucine repeat enrichment region, and the mutation types were mainly missense and truncating.

More interestingly, we verified the expression level of *SLITRK6* in combination with another set of data GSE105130, and found that the expression level was significantly down-regulated in tumor samples (*P-value*<0.05). To verify the diagnostic accuracy of *SLITRK6*, ROC analysis results showed that the AUC of TCGA-LIHC is 0.941 and that of GSE144269 is 0.874, indicating that the AUC>0.5 of *SLITRK6* in GSE144269 and TCGA-LIHC had good diagnostic significance.

Finally, we performed an immunohistochemical analysis of *SLITRK6* in the HPA database, and *SLITRK6* was found to be highly expressed in tumor tissue.

SLITRK6, one of six members of the *SlitRK* family of transmembrane proteins, has been found to be closely involved in cell adhesion, cell differentiation, stem cell characterization, cancer cell migration, and invasion [41,42]. Studies have shown that *SLITRK6* is moderately negatively correlated with tumor malignancy. Extensive immunohistochemical studies have shown that SLITRK6 is expressed in a variety of epithelial tumors, including bladder cancer, lung cancer, breast cancer and glioblastoma [43]. *SLITRK6* is a complete membrane protein, low expression in most tissues, but the high expression in bladder cancer [44]. Although *SLITRK6* is considered to be a promoter of tumorigenesis, studies have shown that the expression of SLITRK6 is significantly down-regulated in smokers and chronic obstructive pulmonary (COPD) patients disease compared with healthy non-smokers [45]. In the study of apigenin regulating the pro-inflammatory activation of TNFα in MD-MB-468 cells of Triple-Negative Breast Cancer (TNBC). Several TNFα deferentially upregulated transcripts were reduced by apigenin, including *CXCL10, C3, PGLYRP4, IL22RA2, KMO, IL7R, ROS1, CFB, IKBKe, SLITRK6* (a checkpoint target), and *MMP13* [46]. The *SLITRK1, SLITRK2, SLITRK3, SLITRK4, SLITRK5,* and *SLITRK6* genes have been detected in the human brain, and the expression profile of each *SLITRK* is unique. The results showed that all SLITRK genes are differentially expressed in brain tumors, including astrocytoma, oligodendroglioma, glioblastoma, medulloblastoma, and supratentorial primitive neuroectodermal tumor (PNET). Interestingly, *SLITRK6* expression is highly selective in several different types of tumor tissues [47]. *SLITRK* family proteins control neurite outgrowth and regulate synaptic development [48]. *SLITRK6* is located in the region of chromosome 13q31.1. Research shows that chromosome 13q31.1 will produce microdeletion. These types of deletions may cause different genetic effects on genotypes and/or phenotypes [49]. While studying an autosomal recessive syndrome characterized by high myopia and sensorineural deafness found that three mutant *SLITRK6* proteins all showed defects in cell surface localization. Histological studies of retinal development after birth in mice with *SLITRK6* deficiency showed delayed synapses in *SLITRK6*-deficient mice [50].

In recent years, many studies have confirmed the effect of gene mutation on tumor progression and prognosis. *MUC16* encodes cancer antigen 125 (CA-125), which is often mutated in gastric cancer (GC). Studies have shown that MUC16 mutation may be related to higher tumor mutation load, better survival outcome, immune response, and cell cycle pathway [51]. It has been reported that tumor mutation load

(TMB) and immune infiltration can predict the response to immunotherapy in several types of tumors. In cutaneous melanoma, TMB is positively correlated with prognosis [52]. FMS-like tyrosine kinase 3 (*FLT3*) mutations are commonly present in acute myeloid leukemia (AML) and are associated with poor prognosis [53]. Combined with the results of our study, it is further demonstrated that the mutation of *SLITRK6* has an impact on the occurrence and development of HCC, which may be a new potential biomarker for HCC.

Certainly, there are many deficiencies in our study; we conducted comprehensive bioinformatics analysis based mainly on data from online databases, lacking complete in vivo and in vitro functional validation. At the same time, different doses of anti-cancer drugs also have various effects on the oxidative stress, inflammation, apoptosis, and viability of tumor cells [54–58]. The diagnostic treatment of tumors is challenging but still has a bright future.

5. Conclusions

Here, we combined multiple bioinformatics research methods to screen out the *SLITRK6* mutation that had a greater influence on the survival of HCC. The results of this study systematically confirmed the influence of *SLITRK6* mutation on HCC, which theoretically has a good diagnostic effect. It provides a new idea for the diagnosis, prognosis, and targeted therapy of HCC.

Author contributions

The study conception and design were performed by Lei Wang and Zhiguo Feng. Data collection and analysis were performed by Xudong Liu, Yajie Liu, Yu Zhang, Ying Ma, Jiangshan Bai, Hongmei Yao, Yafan Wang, Xue Zhao, Rui Li, Yuxuan Chen and Xinqiang Song. The first draft of the manuscript was written by Xudong Liu, Yajie Liu and Zhe Liu. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability statement

Publicly available datasets were analyzed in this study. GSE144269 and GSE105130 can be found in GEO (https://www.ncbi.nlm.nih.gov/gds) and the gene expression RNA-Seq of TCGA-LIHC from the UCSC Xena database (https://xena.ucsc.edu/).

Ethics approval and consent to participate

Not applied.

Consent for publication

Not applied.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2021.101157.

References

[1] A. Forner, M. Reig, J. Bruix, Hepatocellular carcinoma, Lancet 391 (10127) (2018) 1301–1314.

[2] O. Ahmed, L. Liu, A. Gayed, A. Baadh, M. Patel, J. Tasse, U. Turba, B. Arslan, The changing face of hepatocellular carcinoma: forecasting prevalence of nonalcoholic steatohepatitis and hepatitis C cirrhosis, J Clin Exp Hepatol 9 (1) (2019) 50–55.

[3] F. Xu, G. Zha, Y. Wu, W. Cai, J. Ao, Overexpressing lncRNA SNHG16 inhibited HCC proliferation and chemoresistance by functionally sponging hsa-miR-93, OncoTargets Ther. 11 (2018) 8855–8863.

[4] Z. Chen, H. Xie, M. Hu, T. Huang, Y. Hu, N. Sang, Y. Zhao, Recent progress in treatment of hepatocellular carcinoma, Am J Cancer Res 10 (9) (2020) 2993–3036.

[5] M.S. Roh, Recent progress in the treatment of hepatocellular carcinoma, Curr. Opin. Oncol. 2 (4) (1990) 725–730.

[6] D.E. Biancur, K.S. Kapner, K. Yamamoto, R.S. Banh, J.E. Neggers, A.S.W. Sohn, W. Wu, R.T. Manguso, A. Brown, D.E. Root, A.J. Aguirre, A.C. Kimmelman, Functional genomics identifies metabolic vulnerabilities in pancreatic cancer, Cell Metabol. 33 (1) (2021) 199–210, e8.

[7] Q. Zheng, S. Min, Q. Zhou, Identification of potential diagnostic and prognostic biomarkers for LUAD based on TCGA and GEO databases, Biosci. Rep. 41 (6) (2021).

[8] X. Li, Z. Li, H. Zhu, X. Yu, Autophagy regulatory genes MET and RIPK2 play a prognostic role in pancreatic ductal adenocarcinoma: a bioinformatic analysis based on GEO and TCGA, BioMed Res. Int. 2020 (2020) 8537381.

[9] J. Calderaro, M. Ziol, V. Paradis, J. Zucman-Rossi, Molecular and histological correlations in liver cancer, J. Hepatol. 71 (3) (2019) 616–630.

[10] J. Zucman-Rossi, A. Villanueva, J.C. Nault, J.M. Llovet, Genetic landscape and biomarkers of hepatocellular carcinoma, Gastroenterology 149 (5) (2015) 1226–1239 e4.

[11] T. Couri, A. Pillai, Goals and targets for personalized therapy for HCC, Hepatol Int 13 (2) (2019) 125–137.

[12] J. Long, A. Wang, Y. Bai, J. Lin, X. Yang, D. Wang, X. Yang, Y. Jiang, H. Zhao, Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma, EBioMedicine 42 (2019) 363–374.

[13] G. Ma, M. Liu, K. Du, X. Zhong, S. Gong, L. Jiao, M. Wei, Differential expression of mRNAs in the brain tissues of patients with alzheimer's disease based on GEO expression profile and its clinical significance, BioMed Res. Int. 2019 (2019) 8179145.

[14] Q. Wang, S. Zhao, L. Gan, Z. Zhuang, Bioinformatics analysis of prognostic value of PITX1 gene in breast cancer, Biosci. Rep. 40 (9) (2020).

[15] Y. Hao, M. Kacal, A.T. Ouchida, B. Zhang, E. Norberg, H. Vakifahmetoglu-Norberg, Targetome analysis of chaperone-mediated autophagy in cancer cells, Autophagy 15 (9) (2019) 1558–1571.

[16] Z. Li, C. Jiang, Y. Yuan, TCGA based integrated genomic analyses of ceRNA network and novel subtypes revealing potential biomarkers for the prognosis and target therapy of tongue squamous cell carcinoma, PLoS One 14 (5) (2019), e0216834.

[17] Y. Wan, X. Zhang, H. Leng, W. Yin, W. Zeng, C. Zhang, Identifying hub genes of papillary thyroid carcinoma in the TCGA and GEO database using bioinformatics analysis, PeerJ 8 (2020), e9120.

[18] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.

[19] Z. Tian, W. He, J. Tang, X. Liao, Q. Yang, Y. Wu, G. Wu, Identification of important modules and biomarkers in breast cancer based on WGCNA, OncoTargets Ther. 13 (2020) 6805–6817.

[20] P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R, Bioinformatics 24 (5) (2008) 719–720.

[21] M. Gao, W. Kong, Z. Huang, Z. Xie, Identification of key genes related to lung squamous cell carcinoma using bioinformatics analysis, Int. J. Mol. Sci. 21 (8) (2020).

[22] G. Yang, Y. Zhang, J. Yang, Identification of potentially functional CircRNA-miRNA-mRNA regulatory network in gastric carcinoma using bioinformatics analysis, Med. Sci. Mon. Int. Med. J. Exp. Clin. Res. 25 (2019) 8777–8796.

[23] Y. Chi, H. Wang, F. Wang, M. Ding, PHTF2 regulates lipids metabolism in gastric cancer, Aging 12 (8) (2020) 6600–6610.

[24] K. Nie, L. Shi, Y. Wen, J. Pan, P. Li, Z. Zheng, F. Liu, Identification of hub genes correlated with the pathogenesis and prognosis of gastric cancer via bioinformatics methods, Minerva Med. 111 (3) (2020) 213–225.

[25] S.K. Miryala, A. Anbarasu, S. Ramaiah, Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools, Gene 642 (2018) 84–94.

[26] M. Ni, X. Liu, J. Wu, D. Zhang, J. Tian, T. Wang, S. Liu, Z. Meng, K. Wang, X. Duan, W. Zhou, X. Zhang, Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis, Front. Genet. 9 (2018) 469.

[27] X. Zhang, H. Feng, Z. Li, D. Li, S. Liu, H. Huang, M. Li, Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis, OncoTargets Ther. 11 (2018) 6001–6021.

[28] Z. Wei, Y. Liu, S. Qiao, X. Li, Q. Li, J. Zhao, J. Hu, Z. Wei, A. Shan, X. Sun, B. Xu, Identification of the potential therapeutic target gene UBE2C in human hepatocellular carcinoma: an investigation based on GEO and TCGA databases, Oncol Lett 17 (6) (2019) 5409–5418.

[29] Y. Lin, R. Liang, Y. Qiu, Y. Lv, J. Zhang, G. Qin, C. Yuan, Z. Liu, Y. Li, D. Zou, Y. Mao, Expression and gene regulation network of RBM8A in hepatocellular carcinoma based on data mining, Aging 11 (2) (2019) 423–447.

[30] P. Wu, Z.J. Heins, J.T. Muller, L. Katsnelson, I. de Bruijn, A.A. Abeshouse, N. Schultz, D. Fenyo, J. Gao, Integration and analysis of CPTAC proteomics data in the context of cancer genomics in the cBioPortal, Mol. Cell. Proteomics 18 (9) (2019) 1893–1898.

[31] J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (269) (2013) pl1.

[32] G.M. Liu, H.D. Zeng, C.Y. Zhang, J.W. Xu, Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma, Cancer Cell Int. 19 (2019) 138.

[33] Q. Xie, W. Ou-Yang, M. Zhang, H. Wang, Q. Yue, Decreased expression of NUSAP1 predicts poor overall survival in cervical cancer, J. Cancer 11 (10) (2020) 2852–2863.

[34] Q. Fang, H. Chen, Development of a novel autophagy-related prognostic signature and nomogram for hepatocellular carcinoma, Front Oncol 10 (2020) 591356.

[35] X. Liu, J. Wang, M. Chen, S. Liu, X. Yu, F. Wen, Combining data from TCGA and GEO databases and reverse transcription quantitative PCR validation to identify gene prognostic markers in lung cancer, OncoTargets Ther. 12 (2019) 709–720.

[36] C.Y. Li, J.H. Cai, J.J.P. Tsai, C.C.N. Wang, Identification of hub genes associated with development of head and neck squamous cell carcinoma by integrated bioinformatics analysis, Front Oncol 10 (2020) 681.

[37] Y. Zhang, X. Liu, L. Liu, J. Li, Q. Hu, R. Sun, Expression and prognostic significance of m6A-related genes in lung adenocarcinoma, Med. Sci. Mon. Int. Med. J. Exp. Clin. Res. 26 (2020), e919644.

[38] A. Villanueva, Hepatocellular carcinoma, N. Engl. J. Med. 380 (15) (2019) 1450–1462.

[39] X. Yin, J. Wang, J. Zhang, Identification of biomarkers of chromophobe renal cell carcinoma by weighted gene co-expression network analysis, Cancer Cell Int. 18 (2018) 206.

[40] Y. Cai, J. Mei, Z. Xiao, B. Xu, X. Jiang, Y. Zhang, Y. Zhu, Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico, Hereditas 156 (2019) 20.

[41] K. Urh, M. Zlajpah, N. Zidar, E. Bostjancic, Identification and validation of new cancer stem cell-related genes and their regulatory microRNAs in colorectal cancerogenesis, Biomedicines 9 (2) (2021).

[42] C.J. Sandberg, E.O. Vik-Mo, J. Behnan, E. Helseth, I.A. Langmoen, Transcriptional profiling of adult neural stem-like cells from the human brain, PLoS One 9 (12) (2014), e114739.

[43] K. Morrison, P.M. Challita-Eid, A. Raitano, Z. An, P. Yang, J.D. Abad, W. Liu, D. R. Lortie, J.T. Snyder, L. Capo, A. Verlinsky, H. Avina, F. Donate, I.B. Joseph, D. S. Pereira, K. Morrison, D.R. Stover, Development of ASG-15ME, a novel antibody-drug conjugate targeting SLITRK6, a new urothelial cancer biomarker, Mol. Cancer Therapeut. 15 (6) (2016) 1301–1310.

[44] T. Sanford, S. Porten, M.V. Meng, Molecular analysis of upper tract and bladder urothelial carcinoma: results from a microarray comparison, PLoS One 10 (8) (2015), e0137141.

[45] S. Mostafaei, A. Kazemnejad, S. Azimzadeh Jamalkandi, S. Amirhashchi, S. C. Donnelly, M.E. Armstrong, M. Doroudian, Identification of novel genes in human airway epithelial cells associated with chronic obstructive pulmonary disease (COPD) using machine-based learning algorithms, Sci. Rep. 8 (1) (2018) 15775.

[46] D. Bauer, E. Mazzio, A. Hilliard, E.T. Oriaku, K.F.A. Soliman, Effect of apigenin on whole transcriptome profile of TNFalpha-activated MDA-MB-468 triple negative breast cancer cells, Oncol Lett 19 (3) (2020) 2123–2132.

[47] J. Aruga, N. Yokota, K. Mikoshiba, Human SLITRK family genes: genomic organization and expression profiling in normal brain and brain tumor tissue, Gene 315 (2003) 87–94.

[48] T. Morlet, M.R. Rabinowitz, L.R. Looney, T. Riegner, L.A. Greenwood, E. A. Sherman, N. Achilly, A. Zhu, E. Yoo, R.C. O'Reilly, R.N. Jinks, E.G. Puffenberger, A. Heaps, H. Morton, K.A. Strauss, A homozygous SLITRK6 nonsense mutation is associated with progressive auditory neuropathy in humans, Laryngoscope 124 (3) (2014) E95–E103.

[49] Y. Jia, H. Zhao, D. Shi, W. Peng, L. Xie, W. Wang, F. Jiang, H. Zhang, X. Wang, Genetic effects of a 13q31.1 microdeletion detected by noninvasive prenatal testing (NIPT), Int. J. Clin. Exp. Pathol. 7 (10) (2014) 7003–7011.

[50] M. Tekin, B.A. Chioza, Y. Matsumoto, O. Diaz-Horta, H.E. Cross, D. Duman, H. Kokotas, H.L. Moore-Barton, K. Sakoori, M. Ota, Y.S. Odaka, J. Foster 2nd, F. B. Cengiz, S. Tokgoz-Yilmaz, O. Tekeli, M. Grigoriadou, M.B. Petersen, A. Sreekantan-Nair, K. Gurtz, X.J. Xia, A. Pandya, M.A. Patton, J.I. Young, J. Aruga, A.H. Crosby, SLITRK6 mutations cause myopia and deafness in humans and mice, J. Clin. Invest. 123 (5) (2013) 2094–2102.

[51] X. Li, B. Pasche, W. Zhang, K. Chen, Association of MUC16 mutation with tumor mutation load and outcomes in patients with gastric cancer, JAMA Oncol 4 (12) (2018) 1691–1698.

[52] K. Kang, F. Xie, J. Mao, Y. Bai, X. Wang, Significance of tumor mutation burden in immune infiltration and prognosis in cutaneous melanoma, Front Oncol 10 (2020) 573141.

[53] S. Chen, Y. Chen, Z. Zhu, H. Tan, J. Lu, P. Qin, L. Xu, Identification of the key genes and microRNAs in adult acute myeloid leukemia with FLT3 mutation by bioinformatics analysis, Int. J. Med. Sci. 17 (9) (2020) 1269–1280.

[54] F. Kar, S. Kacar, C. Hacioglu, G. Kanbak, V. Sahinturk, Concanavalin A induces apoptosis in a dose-dependent manner by modulating thiol/disulfide homeostasis in C6 glioblastoma cells, J. Biochem. Mol. Toxicol. 35 (5) (2021), e22742.

[55] C. Hacioglu, F. Kar, S. Kacar, V. Sahinturk, G. Kanbak, High concentrations of boric acid trigger concentration-dependent oxidative stress, apoptotic pathways and morphological alterations in DU-145 human prostate cancer cell line, Biol. Trace Elem. Res. 193 (2) (2020) 400–409.

[56] F. Kar, C. Hacioglu, S. Kacar, V. Sahinturk, G. Kanbak, Betaine suppresses cell proliferation by increasing oxidative stress-mediated apoptosis and inflammation in DU-145 human prostate cancer cell line, Cell Stress Chaperones 24 (5) (2019) 871–881.

[57] S. Kacar, F. Kar, C. Hacioglu, G. Kanbak, V. Sahinturk, The effects of L-NAME on DU145 human prostate cancer cell line: a cytotoxicity-based study, Hum. Exp. Toxicol. 39 (2) (2020) 182–193.

[58] C. Hacioglu, S. Kacar, F. Kar, G. Kanbak, V. Sahinturk, Concentration-Dependent effects of zinc sulfate on DU-145 human prostate cancer cell line: oxidative, apoptotic, inflammatory, and morphological analyzes, Biol. Trace Elem. Res. 195 (2) (2020) 436–444.