Original Article

High expression of RASAL1, a hub gene in the progression of liver cancer, suggests a poor prognosis

Ruiwei Shen¹, Xiaona Shao¹, Dawei Chen¹, Chen Wang¹, Ting Lu¹, Dahua Chen¹, Xian Zhu¹, Jieqiong Lin¹, Qunqun Ye¹, Liang Zhao¹, Xingfeng Ge¹, Kai Wang¹, Juan Yi²

¹Department of GI Medicine, Ningbo Medical Centre Lihuili Hospital, Ningbo University, Ningbo 315040, Zhejiang, China; ²Department of Anesthesiology, Ningbo First Hospital, 59 Liuting Street, Ningbo 315000, Zhejiang, China

Received January 25, 2022; Accepted March 21, 2022; Epub April 15, 2022; Published April 30, 2022

Abstract: Objective: Liver cancer (LC) is a frequently occurring lethal malignancy worldwide, yet the molecular mechanisms of carcinogenesis and their development remain uncharacterized. In this study, bioinformatics methods were used to find candidate hub genes for prognosis assessment and clinical treatment of LC. Methods: Differential analysis was carried out based on the evidence of gene expression profiling in LC on The Cancer Genome Atlas (TCGA). The differentially expressed genes (DEGs) were constructed into co-expression networks and divided into modules by virtue of weighted gene co-expression network analysis (WGCNA). Based on the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG), the module genes were subjected to functional enrichment analysis. The LC microarray (GSE105130) in the Gene Expression Omnibus was selected to verify the hub genes’ expression profiles. The validity of the hub genes was verified via survival analysis, as well as expression correlation with the clinicopathological features. Thereafter, gene set variation analysis (GSVA) and single-sample gene set enrichment analysis (GSEA) were applied to investigate the possible biological functions of the hub genes. Results: In total, 3780 DEGs and 17 co-expression modules were obtained. The blue module had the strongest correlation with the tumour stage and the module genes were principally enriched in tumour-associated GO terms, as well as pathways such as Ras protein signal transduction, ERK1/2 cascade, Ras signal pathway, and ECM-receptor interaction. RASAL1, which is highly expressed in LC, was identified as a hub gene for LC progression. Its high expression suggested unfavorable patient prognosis and was correlated with T stage, gender and tumour stage. Further analysis identified that the overexpression of RASAL1 was substantially enriched in cancer-associated gene sets. Conclusion: RASAL1 is a hub gene that influences LC progression, constituting a novel biomarker and molecular target in the future diagnosis and therapy of LC.

Keywords: Weighted gene co-expression network analysis, liver cancer, RASAL1, differential expression analysis

Introduction

Aside from being the sixth most prevalent neoplastic disease, liver cancer (LC) also ranks fourth among the primary causes of cancer deaths worldwide. Multiple adverse factors, such as chronic virus infection, aflatoxin exposure and alcoholism, increase the risk of LC [1]. Due to the large population base and high prevalence of chronic hepatitis B, LC has become the most common cancer type among people under 60 years old in China [1, 2]. Nevertheless, given that early diagnosis of LC is still challenging at present, most patients are already at the late phase of LC during treatment, resulting in a non-optimistic survival of patients (the five-year relative survival rate is merely 10.1%) [3] and seriously compromises the public health of Chinese people. However, as a highly heterogeneous disease, the individual heterogeneity of LC lies to a great extent in its molecular mechanisms [4]. Thus, accurately finding key targets in the treatment of LC has become the focus of initiatives to improve LC clinical efficacy.

With the increasing understanding regarding the molecular drivers of LC, DNA, RNA and proteins can be used as biological markers for cancer prognosis, as well as independent evaluators of clinical efficacy [5], thereby providing unprecedented opportunities for the development of bioinformatics. At present, the Cancer
RASAL1 in the progression of liver cancer

Genome Atlas (TCGA) is the largest database of cancer genes. It stores the gene expression matrix and clinical treatment data of 33 human cancers in 11,000 cases and helps people better understand the molecular mechanism of cancer through high-throughput genome technology, with a view of accelerating the development of cancer target drugs and immunotherapy [6, 7]. In our study, the expression matrix of LC was downloaded from TCGA, differentially expressed genes (DEGs) were screened in tumour samples and their normal counterparts, and hub genes associated with LC progression were mined and identified to discuss their possible biological functions via weighted gene co-expression network analysis (WGCNA), thus offering a theoretical basis in the identification of possible therapeutic targets for LC.

Materials and methods

Downloading and pre-processing of data

LC-associated clinical data and mRNA expression profiles were obtained from TCGA (https://portal.gdc.cancer.gov/) for analysis, including 374 tumour samples and 50 healthy ones. The expression of mRNAs was expressed via read count data, and the heatmap of gene expression was drawn using the R package ggplot2.

Screening of DEGs

The R package edgeR was adopted for screening DEGs in the tumour samples and their normal counterparts with the screening conditions of |log2(FC)| > 1 and p-value < 0.05. Following this, the R package ggplot2 was applied for drawing the volcanic plot of the DEGs.

WGCNA and functional enrichment analysis

The co-expression network of DEGs was constructed using the WGCNA software package to search for crucial modules associated with the clinical symptoms of LC. After matching the expression of DEGs with the gene names, the soft threshold power was calculated to obtain scale independence, and the weighted adjacency matrix was subjected to conversion to acquire a topological overlap matrix (TOM) after being established. Average hierarchical clustering was created based on the TOM’s heterogeneity, in which a module had at least 30 genes and a height of 0.25. Genes with analogous expression profiles were assigned to the same gene module. For a more in-depth investigation of the potential biological functions of the key modules, R package ‘clusterProfiler’, ‘enrichplot’ and ‘ggplot2’ were adopted for the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Gene Ontology (GO) enrichment analyses of module genes. Meaningful pathways (P < 0.05) and the first 10 GO terms were screened, among which GO analysis covered the cellular components (CC), molecular functions (MF), and biological processes (BP).

Screening and identification of hub genes

With threshold ≥ 0.1 set as the screening condition of the module correlation edge, the function ‘exportNetworkToCytoscape’ in the R package WGCNA was used to import genes from the key modules into Cytoscape (Version 3.7.1). With the maximal clique centrality (MCC) algorithm, the top 10 crucial genes were selected according to their scores. The histogram of each gene score was drawn using the R package ggplot2. The gene with the highest score was considered the candidate hub gene of this study. In addition, the expression profile of LC (GSE105130) was retrieved from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/), including 25 pairs of tumour tissues from patients with LC and their non-cancerous counterparts. Paired-sample t-test was adopted to analyse the hub genes’ expression levels among different samples, with P < 0.05 considered as statistically significant. Visualization was carried out using the R package ggplot2.

Survival analysis

To further probe into the prognostic significance of hub genes in LC, Kaplan-Meier (K-M) analysis of TCGA was conducted via the survfit function in the R package survival. Low- and high-expression groups were set based on the genes’ median expressions. The 95% confidence interval (95% CI), as well as the hazard ratio (HR), was reported. The function ggsurvplot was used to plot the survival curve.

Correlation between hub gene expression levels and clinicopathological features

We assigned the hub genes in the LC samples in TCGA to low-/high-expression groups based on the median expression. Then, the Chi-square test was applied for association analysis.
between each group and clinical traits (T_stage, M_stage, N_stage, Gender, Age, Survival_Status, Survival_Time, Tumour_Stage), with a statistical significance level of $P < 0.05$.

**Gene set variation analysis (GSVA) and single-sample gene set enrichment analysis (GSEA)**

After assigning the hub genes to high-/low-expression groups based on their median expression, they were analysed by GSEA and GSVA to discuss their biological functions. First, the gene dataset ‘h.all.v7.1.symbols.gmt’ was acquired in MSigDB (https://www.gsea-msigdb.org/gsea/index.jsp). Corresponding gene pathways were found by the msigdbr function in the R package msigdbr, after which they were analysed using the R package GSEA, with FDR < 25% and $P < 0.05$ as screening conditions. In addition, the R package GSVA was used to transform the gene expression profiles into pathway name expression profiles. After the heatmap was drawn using the R packages limma and ggplot2, respectively.

**Results**

**Screening of DEGs**

A heatmap was drawn to show the expression levels of 16,006 genes in the LC tumour samples and their normal counterparts in TCGA (Figure 1A). Among these, 3,780 DEGs were screened, with 982 showing up-regulated expression and 2,798 presenting down-regulated expression (Figure 1B), suggesting that LC can cause significant changes in the mRNA expression level in cancer cells.

**WGCNA**

The DEGs retrieved from the TCGA LC expression matrix were used as research subjects to construct a weighted gene co-expression network. Moreover, the hclust function was adopted for the expression matrix calculated by the mean value method to draw the cluster dendrogram, in which clinical traits were input to screen out isolated samples (Figure 2A). The soft threshold power was set to 1-10 to ensure high-scale independence ($R^2 = 0.9$) and low mean connectivity (near 0) when $\beta = 4$ (Figure 2B). Seventeen modules were retained through the dynamic tree cut means (Figure 3A). The module-trait relationships are shown in Figure 3B. The blue module related to tumour stage had the deepest colour ($r = 0.29$, $P < 0.001$), which was regarded as the crucial module for further analyses.

**Enrichment analyses of key modules**

The gene members in the blue module were strongly linked to the gene significance for tumour stage ($r = 0.59$, $P < 0.001$) (Figure 4A). According to the functional enrichment analysis of 476 genes in the module, those in blue colour were enriched in BP-related GO terms, such as Ras protein signal transduction, ERK1/2 cascade, JNK cascade and MAPK cascade. They were also related to CC, such as...
Figure 2. Sample dendrogram and soft threshold calculation. A. Sample dendrogram and trait heatmap: the seven indexes are T.Stage, M.Stage, N.Stage, Gender, Age, Survival_Status, Survival_Time, and Tumor_Stage, respectively. The color intensity is positively correlated with stage progression, male, age, survival and survival time, and the gray area is the missing value. B. Scale independence and mean connectivity of each soft threshold.

Figure 3. Screening key modules in weighted gene co-expression network analysis (WGCNA). A. Gene cluster dendrogram generated by hierarchical clustering based on the dissimilarity measure (equal to one minus the topological overlap matrix, namely, 1 - TOM). Each module is assigned a color. According to the correlation between modules, a total of 17 modules are generated. The two rows below the cluster dendrogram represent the original module and the merged module, respectively. B. Heatmap of module-trait relationships. The darker the module color, the stronger the correlation between the two.
tetra-azide enrichment microdomain, filamentous actin and myosin, as well as MF-associated GO terms, including sensitive ion channel activity, collagen binding and gated channel activity (Figure 4B). KEGG enrichment analysis found that the module genes were correlated with the Ras signal pathway, ECM-receptor interaction, MAPK signal pathway, and others, as shown in the figure. Further analysis revealed a correlation of the blue module genes with carcinogenesis and progression (Figure 4C), suggesting that the blue module genes are bound to the occurrence and development of LC.

Mining of hub genes

The MCC method was used to analyse the centrality of each gene in the module. Up to 401 nodes and 9,276 edges were identified, highlighting the top 10 central genes as follows: RASAL1, PFKP, ITGB4, ELOVL7, ITPR3, TMC4, KIAA1522, CACNB3, C12orf75 and KLC2 (Figure 5A, 5B). Among them, RASAL1 scored the highest in the MCC value. Therefore, it was used as the hub gene in this study for subsequent analyses. The expressions of RASAL1 in the TCGA LC expression matrix and the GSE105130 microassay were discussed. The results showed a noticeable elevation of RASAL1 in the tumour tissues compared with their normal counterparts (Figure 5C).

RASAL1 expression and survival analysis of patients with LC

Based on the K-M analysis, the five-year survival rates of LC patients with high and low RASAL1 expression were 63.04% and 67.37%,
RASAL1 in the progression of liver cancer

Figure 5. Mining and identification of hub genes. A. The Maximal Clique Centrality (MCC) method was used to mine hub genes, and the genes with the top 10 MCC values were colored in red and yellow. B. MCC values of hub genes; C. Comparison of RASAL1 expression level between tumor samples and normal samples in TCGA and GSE105130 microarray. Note: *P < 0.05; ****P < 0.0001.

Figure 6. Survival curve of liver cancer patients with high and low RASAL1 expression. The red represents RASAL1 high expression group and the blue represents RASAL1 low expression group.

respectively, indicating a worse prognosis in LC patients with high RASAL1 expression (HR = 1.150 (0.640-1.580), P < 0.05; Figure 6).

Correlation between RASAL1 expression and clinicopathological features

The Chi-square test was used to analyse the correlation of RASAL1 expression with the clinicopathological features of LC patients. The results identified a connection between RASAL1 expression and T stage, gender and tumour stage of LC patients (Figure 7), which indicated that a high expression of RASAL1 accounted for a higher proportion in LC patients with higher T stage (TIII, TIV) and clinical stage II-IV LC, as well as female patients (P < 0.05).
RASAL1 in the progression of liver cancer

The GSEA results (Figure 8) showed that high RASAL1 expression was significantly enriched in gene sets like HALLMARK_TGF_BETA_SIGNALING, HALLMARK_WNT_BETA_CATENIN_SIGNALING, HALLMARK_IL6_JAK_STAT3_SIGNALING, HALLMARK_APOPTOSIS, HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSFORMATION, and HALLMARK_PEROxisome, indicating that RASAL1 was involved in tumour-related BP in LC. Furthermore, GSVA revealed that RASAL1 was obviously enriched in high-expression gene sets, such as HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSFORMATION, HALLMARK_MTORC1_SIGNALING, and HALLMARK_PS3_PATHWAY (Figure 9), which further attest that RASAL1 is related to the malignant transformation of cancer cells.

Discussion

Chronic liver injury traverses the entire development process of LC. Sustained damage leads to a high turnover rate of liver cells, which drives obvious changes in the liver microenvironment, leading to the formation of a proto-oncogenic environment and promoting gene transformation and malignant transformation of liver cells [8]. An increasing number of studies have shown that aberrantly expressed genes may be the key factors in accelerating the malignant progression of LC [9, 10]. Thus, finding genes with different gene expression levels carries profound implications for exploring biomarkers for diagnosing and treating LC [11-13].

In this study, the DEGs in the LC expression matrix of the TCGA database were screened through differential expression analysis, including 982 up-regulated DEGs and 2798 down-regulated DEGs. Then, based on the above DEGs, a co-expression network was set up through WGCNA, and 17 gene modules were segmented and retained. WGCNA is a unique method for identifying gene modules with ana-

Figure 7. Correlation between RASAL1 expression and clinicopathological features. Note: *P<0.05; **P<0.01; ****P<0.0001.

Figure 8. Gene set enrichment analysis (GSEA) of RASAL1. RASAL1 high expression groups in (A) “HALLMARK_TGF_BETA_SIGNALING”, “HALLMARK_WNT_BETA_CATENIN_SIGNALING”, “HALLMARK_IL6_JAK_STAT3_SIGNALING”; (B) “HALLMARK_APOPTOSIS”, “HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSFORMATION”, “HALLMARK_PEROxisome”; (C) “HALLMARK_INFLAMMATORY_REACTION”, “HALLMARK_ALLOGRAFT_REJECTION”, “HALLMARK_TNFA_SIGNALING_VIA Enrichment in the “_NFKB” gene set.
logical expression modes and finding the central participants in the modules, namely, module genes [14, 15]. At present, WGCNA is applied to identify potential biomarkers in various malignancies, such as breast cancer and LC [16, 17], providing candidate genes that are highly correlated with the clinical traits for tumours.

Based on the results of this study, the blue module was the key that was most related to tumour staging. According to the KEGG and GO enrichment analyses of 476 genes in this module, these genes were associated with the biological functions of Ras protein signal transduction, ERK1/2 cascade, as well as JNK and MAPK cascades. In addition, genes in this blue module were greatly enriched in the MAPK signalling pathway, ECM-receptor interaction, Ras signalling pathway and others. The first pathway is a vital regulator in physiological processes, such as cell growth and differentiation. The MAPK family contains JNK, ERK1/2 and p38 MAPK cascades, while the abnormal regulation of RAS/MAPK/ERK is implicated in the metastasis, invasion and proliferation of various malignant tumours [18, 19]. These studies also support the impact of the genes in the blue module on the occurrence and development of LC.

MCC is a topology analysis method [18, 19]. In the present study, the candidate hub genes (RASAL1, PFKP, ITGB4, ELOVL7, ITPR3, TMC4, KIAA1522, CACNB3, C12orf75 and KLC2) in the blue module were chosen via the MCC algorithm. Among them, RASAL1 was identified as the hub gene by sequencing according to the MCC value. Comparing the expression matrix between the LC samples and the normal counterparts in TCGA and GSE105130, it was found that RASAL1 was up-regulated in LC, and patients with high RASAL1 expression had poor prognoses. Moreover, RASAL1 expression was related to the advanced T stage and clinical stage of LC patients. It has been pointed out that the staging of LC, including clinical staging, TNM staging and BCLC staging, is the key factor in predicting patient survival and selecting different clinical treatment schemes [20]. Accordingly, it is of profound importance to find an appropriate biomarker for choosing feasible clinical treatment schemes for LC.

RASAL1, which is located at 12q23-q24, is a calcium-dependent RAS protein activity regulation gene whose gene promoter methylation level is substantially increased in gastric cancer, colorectal cancer and other gastrointestinal tumour tissues, promoting tumour invasion and growth [21, 22]. In addition, RASAL1 is shown to be a cancer suppressor in multiple neoplastic diseases. When RASAL1 is inactivated due to the hypermethylation and mutation of tumours, its low expression level can activate the RAS/MEK/ERK signalling pathway and promote cell apoptosis by inducing cell cycle arrest to inhibit cell proliferation [23, 24]. It is also worth mentioning that RASAL1 overexpression is associated with distant metastasis and tumour invasion [25]. Chang et al. [26] found that RASAL1 is up-regulated in ovarian cancer (OC) tissues and cells, and its high expression is related to the deterioration of the
survival rate. Moreover, it participates in the regulation of the MAPK signaling pathway, and RASAL1 silencing inhibits the proliferation and invasion of OC cells. The results of GSEA and GSVA on RASAL1 also demonstrated that RASAL1 is associated with the activation of LC-related signal pathways, such as TGF-β and epithelial mesenchymal transition [27, 28].

Due to experimental condition limitations, the specific role of RASAL1 in LC was not discussed in detail and verified by experiments. This will be the main direction of our follow-up research. Nevertheless, our study uncovered the key role of RASAL1 in LC for the first time, including its contribution to the progression of LC. The findings of this study serve as significant guides in the prognosis evaluation of LC patients and targeted therapy.

Acknowledgements

This work was supported by the Ningbo Natural Science Foundation, research on the molecular mechanism of RASAL1 participating in the occurrence and development of liver cancer by regulating Ras signaling pathway, approval numbers: 2019A610222.

Disclosure of conflict of interest

None.

Address correspondence to: Juan Yi, Department of Anesthesiology, Ningbo First Hospital, 59 Liuting Street, Ningbo 315000, Zhejiang, China. Tel: +86-0574-87085588; E-mail: yijuan73@126.com

References


[18] Teng P, Xu X, Ni C, Yan H, Sun Q, Zhang E and Ni Y. Identification of key genes in calcific aor-
RASAL1 in the progression of liver cancer

[28] Yu K, Li Q, Shi G and Li N. Involvement of epithelial-mesenchymal transition in liver fibrosis. Saudi J Gastroenterol 2018; 24: 5.