

GOPEN ACCESS

Citation: Zhou J, Zhu Y, Ma S, Li Y, Liu K, Xu S, et al. (2021) Bioinformatics analysis identifies DYNC111 as prognosis marker in male patients with liver hepatocellular carcinoma. PLoS ONE 16(10): e0258797. https://doi.org/10.1371/journal. pone.0258797

Editor: Katherine James, Northumbria University, UNITED KINGDOM

Received: April 27, 2021

Accepted: October 5, 2021

Published: October 22, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0258797

Copyright: © 2021 Zhou et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: In this study the original sequence data and clinical information can be downloaded from TCGA database (https://

RESEARCH ARTICLE

Bioinformatics analysis identifies DYNC111 as prognosis marker in male patients with liver hepatocellular carcinoma

Jian Zhou¹°, Yue Zhu²°, Songlin Ma³°, Yi Li¹, Kun Liu⁴, Sihuan Xu², Xin Li², Li Li⁵, Junfang Hu⁶, Yan Liu⁵²*

1 Department of Infectious Diseases, Puren Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, China, 2 Biological Cell Therapy Research Center, Puren Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, China, 3 Department of Gastroenterology, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, 4 Otorhinolaryngology, Puren Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, China, 5 The Ministry of Science and Education, Puren Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, China, 6 Department of Pharmacy, Puren Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, China

So These authors contributed equally to this work.

* liuyan@wust.edu.cn

Abstract

Background

Liver hepatocellular carcinoma (LIHC) is one of the most common malignant tumors. However, the etiology and exact molecular mechanism of LIHC are still not fully understood, which makes it urgent for us to further study the molecular events behind.

Methods

In this study, differences in mRNA expression between LIHC samples and normal adjacent samples were found through analyzing the TCGA database, and key targets were sought. We analyzed 371 LIHC samples and 50 normal adjacent samples according to P <0.01 and logFC>2.5, a total of 1092 genes were identified differentially expressed, including 995 upregulated genes and 97 down-regulated genes. We predicted the interactions of these differentially expressed mRNAs, and used Cyto-Hubba to locate the hub gene-dynein cytoplasmic 1 intermediate chain 1 (DYNC111).

Results

Survival analysis showed that DYNC111 was a prognostic factor for LIHC male patients. Functional enrichment indicated that DYNC111 and differentially expressed interacting proteins were involved in the cell cycle.

Conclusion

In conclusion, this study discovers that DYNC1I1 can be used as a prognostic marker for LIHC male patients.

cancergenome.nih.gov/), and all the data generated by the analysis are included in this article.

Funding: This study was funded by the Hubei Province health and family planning scientific research project (WJ2019M257) and Wuhan Municipal Health scientific research project (WX20C13). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Liver hepatocellular carcinoma (LIHC) is one of the most common malignant tumors of the digestive tract. Globally, the incidence of LIHC ranks the sixth in the incidence of malignant tumors and the fourth in mortality [1]. LIHC seriously affects the lives and health of people. At present, the overall prognosis of LIHC is unsatisfactory. The main reasons include the insidious disease, high degree of malignancy, recurrence and metastasis [2]. Therefore, the identification of LIHC-specific biomarkers can help predict and monitor disease progression, and more importantly, through the implementation of early intervention, cases that may evolve into aggressive diseases can be reduced [3].

The Cancer Genome Atlas Project (TCGA) was jointly launched in 2006 by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). The TCGA database contains genome data for 33 tumor projects and provides original sequencing data to all researchers [4]. TCGA has released many mRNA sequencing data of LIHC cancer patients. This study aimed to determine the mRNA expression differences between LIHC samples and normal adjacent samples through analyzing highthroughput mRNA data downloaded from the TCGA database. We used protein interactions [5] and Cyto-Hubba [6] to find the hub gene-DYNC111. Also, we assessed the prognostic value of DYNC111 and analyzed the possible biological functions of DYNC111, which were expected to provide new insights into the underlying molecular mechanisms of LIHC.

Material and methods

Data processing

Raw sequencing data and clinical information were downloaded from the TCGA database (https://cancergenome.nih.gov/). Altogether, a total number of 421 samples were enrolled in this study, including 371 LIHC samples and 50 normal adjacent samples. We used the R language package to process mRNAs sequencing data. The differentially expressed mRNAs between LIHC samples and normal adjacent samples are analyzed by the "DESeq2" package in R. When we were calculating the fold change (FC) of individual mRNA expression, we considered that the differentially expressed mRNAs with P < 0.01 and logFC > 2.5 as significant.

Differential gene analysis

The R package of "DESeq2_1.28.1" [7] was used for Principal Component Analysis (PCA) and MA-plot of significant differentially expressed mRNA, and we used the code "dds <- dds[row-Sums(counts(dds))>1,]" to delete low-expressed genes. The "pheatmap_1.0.12" package [8] and "ggrepel_0.8.2" package [9] in R were used for heatmap and volcano map, respectively. The default parameters were used for all analysis.

Functional enrichment analysis

We used the R package "ClusterProfiler_3.16.1" [10] for gene Ontology (GO) enrichment and used the online database DAVID [11] for gene Kyoto Encyclopedia of Genes and Genomics (KEGG) pathway analysis. GO terms or KEGG pathways with P < 0.05 were considered statistically significant.



log2(fold change)

Fig 1. Identification of expression differences between LIHC samples and normal adjacent samples. A. PCA of 371 LIHC samples and 50 normal adjacent samples expression analysis. Red represented normal adjacent samples, and blue represented LIHC samples; B.MA-plot of the differential mRNA expression analysis; C. Heat map of the differential mRNA expression analysis. Blue represented normal adjacent samples, and red represented LIHC samples; D. Volcano plot of the differential mRNA expression analysis. Red represented differential mRNAs that were highly expressed in LIHC samples.

https://doi.org/10.1371/journal.pone.0258797.g001

Protein-Protein Interaction (PPI) network construction and analysis of modules

STRING is a search tool that can analyze the interaction relationship between genes/proteins (https://string-db.org/). Using STRING to analyze the PPI network can help us understand the relationships between different genes/proteins. Cyto-Hubba software was used to screen for hub genes. We used 11 algorithms of the Cyto-Hubba software to screen out the hub gene: MCC, DMNC, MNC, Degree, EPC, BottleNeck, EcCentricity, Closeness, Radiality, Betweenness, Stress.

Prognosis analysis

We used the LIHC patient data in TCGA and the "Survival" package and "Survminer" package of R language to draw the survival curve of LIHC patients.



Fig 2. Functional enrichment analysis of differentially expressed mRNAs. The most important GO terms for cell component (A), biological process (B) and molecular function (C) of 1092 differential genes were shown in the figures; D. Gene networks identified through KEGG pathway analysis of the differentially expressed mRNAs.

https://doi.org/10.1371/journal.pone.0258797.g002

Results

Identification of differentially expressed mRNAs in LIHC

We retrospectively analyzed the TCGA LIHC dataset. In the present study, a total of 421 samples were enrolled in this study, including 371 LIHC samples and 50 normal adjacent samples, to identify the differentially expressed genes. PCA showed that LIHC samples and normal adjacent samples fell in different areas, and the two groups of samples were effectively



Fig 3. Identify the hub gene. A. Use STRING online database to build a PPI network that interacts with DYNC111; B. Use Cyto-Hubba software to identify the hub gene. DYNC111 was the only hub gene that intersects among the top 100 genes obtained by each of the 11 algorithms.

https://doi.org/10.1371/journal.pone.0258797.g003

distinguished (Fig 1A, 1B) depicted the transcription abundance values of LIHC samples and normal adjacent samples as MA-plot. According to P < 0.01 and logFC > 2.5, a total number of 1092 differentially expressed mRNAs were identified between LIHC samples and normal adjacent samples, including 995 up-regulated and 97 down-regulated mRNAs. We presented the result in the form of Heat map (Fig 1C) and Volcano plot (Fig 1D).

Functional annotation of differentially expressed mRNAs in LIHC

In order to understand the biological roles of the 1092 differentially expressed mRNAs in LIHC, we performed GO and KEGG pathway enrichment analysis using the R package "Clusterprofiler", which showed 20 most significant cellular components, biological process and molecular function terms, and 9 significant KEGG pathways. GO analysis revealed that major terms enriched in the cellular component terms were ion channel complex and transmembrane transporter complex (Fig 2A). The most significantly enriched biological process category were organelle fission and nuclear division (Fig 2B). For the molecular function category, the primary enriched terms were passive transmembrane transporter activity, ion channel activity, and DNA–binding transcription activator activity (Fig 2C). KEGG pathway analysis indicates that the differentially expressed mRNAs played relevant roles in cell cycle, cytokine –cytokine receptor interaction (Fig 2D). The specific details of GO and KEGG pathway enrichment analysis are in the S1–S4 Tables.

Using string and Cyto-Hubba to identify the hub gene in LIHC

We used STRING online database to analyze 1092 differentially expressed mRNAs and construct PPI network (Fig 3A), for details about PPI calculation in STRING, see S5 Table. Then, we downloaded the results and use Cyto-Hubba software for analysis. We analyzed the results of PPI using 11 algorithms of Cyto-Hubba software. We extracted the top 100 hub genes from 11 algorithms, and get the only key gene DYNC111 after taking the intersection (Fig 3B). The detailed scoring of genes in PPI by 11 algorithms was in S6 Table.



Fig 4. Characterization of DYNC111. Genomic localization (A) and protein structure (B) of DYNC111; C. Differential expression of DYNC111 in LIHC samples and normal adjacent samples (P = 3.690758e-05); D. Differential expression of DYNC111 in 50 pairs of LIHC samples and normal adjacent samples (P = 0.0056); E. Differential expression of DYNC111 in LIHC male patients and female patients (P = 0.0026); F. Association between DYNC111 expression and disease-free survival time in the TCGA-LIHC dataset (P = 0.001). G-H. DYNC111 had a significant impact on the survival of LIHC male patients (P = 0.002), but has no significant impact on the survival of LIHC female patients (P = 0.07).

https://doi.org/10.1371/journal.pone.0258797.g004

Characterization of DYNC111

DYNC111 (Dynein Cytoplasmic 1 Intermediate Chain 1) is an important cargo-binding subunit of cytoplasmic dynein. It serves as one of several non-catalytic accessory components of the cytoplasmic dynein 1 complex that are thought to be involved in linking dynein to cargos and to adapter proteins that regulate dynein function. DYNC111 is located at chr7: 95772554-96110322 (Fig 4A). We use online database SWISS-MODEL [12] to examine the protein structure of DYNC111, Fig 4B showed the protein structure of DYNC111, the structure consists of multiple beta folds to form a compact space structure. There was an alpha helix at the N-terminal and the C-terminal, and the alpha helix at the N-terminal is free to the outside. We detected the expression of DYNC111 in LIHC data of the TCGA database, and the results showed that the expression of DYNC111 in LIHC samples was higher than in normal adjacent samples, and the P value = 3.690758e-05, which was statistically significant (Fig 4C). We extracted 50 pairs of LIHC samples in the TCGA data, and the results showed that the expression of DYNC111 in the paired LIHC samples was also higher than in the paired normal adjacent samples, and the P value = 0.0056 (Fig 4D). The expression of DYNC111 in LIHC male patients and female patients was also different (P = 0.0026), and the results showed that its expression in male patients was higher than in female patients (Fig 4E). We further explored whether the elevated expression of DYNC111 level in LIHC affected the survival of patients. LIHC data with gene expression and clinical information from TCGA were used to investigate DYNCII1 prognostic significances. The survival rate of DYNC111 high expression group was significantly lower (P = 0.001, hazard ratio: 0.98, 95% CI: 0.84–1.1, Fig 4F). Our previous results showed that the expression of DYNC111 in LIHC male patients was higher than in female patients. Therefore, we separately investigated the impact of DYNC111 on the survival of LIHC male and female patients. The results found that DYNC111 had a significant impact on the survival of male patients (Fig 4G), while it had not statistically significant impact on the survival of female patients (Fig 4H). These results indicated that high expression of DYNC111 was an adverse factor for the survival of LIHC male patients.

Function analysis of DYNC111

We extracted the proteins that interact with DYNC111 in the PPI results (Fig 5A), The specific details of the PPI results were in the Table 1. To reveal the potential biological functions of these interacting proteins, we conducted GO and KEGG analysis. GO function annotation of the interacting proteins was performed using the R package "Clusterprofiler". The most significant GO terms for cellular component, biological process and molecular function were shown in Fig 5B–5D. The KEGG pathway analysis was performed using the DAVID database, and the results of the analysis were shown in Fig 5E. The interacting proteins were mainly enriched in cell cycle, Progesterone–mediated oocyte maturation and Oocyte meiosis. This analysis indicated that DYNC111 was mainly involved in cell cycle. The specific details of GO and KEGG pathway enrichment analysis were in the S7–10 Tables.

Discussion

LIHC is a highly malignant tumor with a high mortality rate. The treatment of LIHC is mainly surgical resection, but due to its high recurrence and metastasis rate, the prognosis of LIHC



Fig 5. Function analysis of DYNC111. The figure showed the cellular component (A), biological process (B) and molecular function (C) of DYNC111 and the proteins that interacted with it (D); E. Possible pathways mediated by DYNC111 and its interacting proteins. The color of the dot represents the P value, and the size of the dot represented the number of counts.

https://doi.org/10.1371/journal.pone.0258797.g005

Gene	PPI
DYNC1I1	KIF20A KIF11 MAD2L1 ZWINT SPTA1 FOLR1 PLK1 SGOL1 NDC80 NUF2 SKA1 SPC24 KIF4A KIF18A RACGAP1 SGOL2 KIF23 SPC25 KIF15 KIF5A ERCC6L KIF2C

Table 1. Proteins that interact with DYNC1I1.

https://doi.org/10.1371/journal.pone.0258797.t001

patients is often far from satisfactory [13, 14]. Due to the insidious onset, rapid progress, high recurrence and metastasis rate of LIHC, the prognosis of patients is poor, the overall 5-year survival remains at 25%-39%, and the recurrence rate of advanced LIHC patients is about 80% [15]. Therefore, a better understanding of the progression of LIHC and new mechanisms of occurrence and progression will help us find new therapeutic targets, formulate more effective treatment strategies, and extend the survival time of LIHC patients.

At present, the first-line targeted drugs for LIHC treatment include levatinib and sorafenib, and the second-line targeted drugs include cabotinib, regorafenib, ramucirumab and nivolumab [16]. In order to improve the prognosis of patients with LIHC and advance development of this field, more researches are needed on molecular targeted therapies that use genome maps and biomarker matching [17, 18]. With the advancement of RNA sequencing and other technologies, the mechanism of the occurrence and progress of LIHC is constantly being explored, and more targets are identified and utilized [19, 20].

In the present study, we find that DYNC111 is differentially expressed in LIHC patients, and it is significantly related to patient survival. Previous studies have shown that DYNC111 can promote the progression and metastasis of gastric cancer [21, 22], colon cancer [23] and glioblastoma [24], but no study has found that DYNC111 has a biological function in LIHC and is a prognostic factor of LIHC. The content of our research makes researchers better understand the biological functions of DYNC111.

In order to gain new insight into the molecular functions of DYNC111, we screened the differentially expressed genes that interact with it, analyzed related pathways, and performed GO annotations. Abnormal signal pathways play a vital role in the occurrence and development of LIHC in male patients. We discover that DYNC111 can regulate the cell cycle and is related to some transmembrane transportation, which indicates that DYNC111 may not only affect the progress of LIHC in male patients, but also affect drug transportation.

In summary, we have identified DYNC111 as a potential prognostic predictor of LIHC for male patients. In future research, we will classify patients' samples by sex and use a larger sample size to verify our research results, and the molecular mechanism of DYNC111 in progression of LIHC also needs more in-depth functional researches.

Supporting information

 S1 Table.

 (XLSX)

 S2 Table.

 (XLSX)

 S3 Table.

 (XLSX)

 S4 Table.

 (XLSX)

 S5 Table.

 (XLSX)

S6 Table. (XLSX) **S7 Table.** (XLSX) **S8 Table.** (XLSX) **S9 Table.** (XLSX) **S10 Table.** (XLSX)

Author Contributions

Conceptualization: Yan Liu.

Data curation: Jian Zhou, Yue Zhu, Songlin Ma, Yi Li, Kun Liu, Sihuan Xu, Xin Li, Li Li, Junfang Hu, Yan Liu.

Formal analysis: Jian Zhou, Yue Zhu, Songlin Ma, Yan Liu.

Funding acquisition: Jian Zhou, Yue Zhu.

Investigation: Jian Zhou, Yue Zhu, Songlin Ma, Yi Li, Kun Liu, Sihuan Xu.

Methodology: Jian Zhou, Yue Zhu, Songlin Ma, Yi Li, Kun Liu, Sihuan Xu, Xin Li, Li Li, Junfang Hu.

Project administration: Yan Liu.

Resources: Yi Li, Kun Liu, Sihuan Xu, Yan Liu.

Supervision: Yan Liu.

Validation: Jian Zhou, Yue Zhu, Songlin Ma, Yi Li, Kun Liu, Sihuan Xu, Xin Li, Li Li, Junfang Hu.

Writing - original draft: Jian Zhou, Yue Zhu, Songlin Ma.

Writing - review & editing: Yi Li, Kun Liu, Sihuan Xu, Xin Li, Li Li, Junfang Hu, Yan Liu.

References

- 1. Duran SR, Jaquiss RDB. Hepatocellular Carcinoma[J]. N Engl J Med, 2019, 381(1):e2
- Villanueva A, Hernandez-Gea V, Llovet JM. Medical therapies for hepatocellular carcinoma: a critical view of the evidence[J]. Nat Rev Gastroenterol Hepatol, 2013, 10(1):34–42 https://doi.org/10.1038/nrgastro.2012.199 PMID: 23147664
- 3. Yang JD, Heimbach JK. New advances in the diagnosis and management of hepatocellular carcinoma [J]. Bmj, 2020, 371:m3544 https://doi.org/10.1136/bmj.m3544 PMID: 33106289
- Chandran UR, Medvedeva OP, Barmada MM, et al. TCGA Expedition: A Data Acquisition and Management System for TCGA Data[J]. PLoS One, 2016, 11(10):e0165395 <u>https://doi.org/10.1371/journal.pone.0165395</u> PMID: 27788220
- Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets[J]. Nucleic Acids Res, 2019, 47(D1):D607–d613 https://doi.org/10.1093/nar/gky1131 PMID: 30476243

- Otasek D, Morris JH, Bouças J, et al. Cytoscape Automation: empowering workflow-based network analysis[J]. Genome Biol, 2019, 20(1):185 <u>https://doi.org/10.1186/s13059-019-1758-4</u> PMID: 31477170
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. Genome Biol, 2014, 15(12):550 <u>https://doi.org/10.1186/s13059-014-0550-8</u> PMID: 25516281
- 8. R K. Pheatmap: Pretty Heatmaps[J]. 2015,
- 9. HW. ggplot2: Elegant Graphics for Data Analysis.[J]. Springer, 2009,
- Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. Omics, 2012 16(5):284–287 https://doi.org/10.1089/omi.2011.0118 PMID: 22455463
- Dennis G Jr., Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery[J]. Genome Biol, 2003, 4(5):P3 PMID: 12734009
- Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes[J]. Nucleic Acids Res, 2018, 46(W1):W296–w303 <u>https://doi.org/10.1093/nar/gky427</u> PMID: 29788355
- Yang JD, Hainaut P, Gores GJ, et al. A global view of hepatocellular carcinoma: trends, risk, prevention and management[J]. Nat Rev Gastroenterol Hepatol, 2019, 16(10):589–604 <u>https://doi.org/10.1038/</u> s41575-019-0186-y PMID: 31439937
- Huang Z, Zhou JK, Peng Y, et al. The role of long noncoding RNAs in hepatocellular carcinoma[J]. Mol Cancer, 2020, 19(1):77 https://doi.org/10.1186/s12943-020-01188-4 PMID: 32295598
- **15.** Zuo Q, He J, Zhang S, et al. PGC1 α suppresses metastasis of HCC by inhibiting Warburg effect via PPAR γ -dependent WNT/ β -catenin/PDK1 axis[J]. Hepatology, 2020,
- Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma[J]. Nat Rev Dis Primers, 2016, 2:16018 https://doi.org/10.1038/nrdp.2016.18 PMID: 27158749
- Lim H, Ramjeesingh R, Liu D, al. Optimizing Survival by Changing the Landscape of Targeted Therapy for Intermediate and Advanced Hepatocellular Carcinoma: A Systematic Review[J]. J Natl Cancer Inst, 2020
- 18. Finn RS, Zhu AX. Evolution of Systemic Therapy for Hepatocellular Carcinoma[J]. Hepatology, 2020,
- Kudo M, Ueshima K, Yokosuka O, et al. Sorafenib plus low-dose cisplatin and fluorouracil hepatic arterial infusion chemotherapy versus sorafenib alone in patients with advanced hepatocellular carcinoma (SILIUS): a randomised, open label, phase 3 trial[J]. Lancet Gastroenterol Hepatol, 2018, 3(6):424–432 https://doi.org/10.1016/S2468-1253(18)30078-5 PMID: 29631810
- 20. Pillai A, Ahn J, Kulik L. Integrating Genomics Into Clinical Practice in Hepatocellular Carcinoma: The Challenges Ahead[J]. Am J Gastroenterol, 2020,
- Gong LB, Wen T, Li Z, et al. DYNC111 Promotes the Proliferation and Migration of Gastric Cancer by Up-Regulating IL-6 Expression[J]. Front Oncol, 2019, 9:491 <u>https://doi.org/10.3389/fonc.2019.00491</u> PMID: 31249807
- 22. Gong L, Wen T, Li Z, et al. TNPO2 operates downstream of DYNC111 and promotes gastric cancer cell proliferation and inhibits apoptosis[J]. Cancer Med, 2019, 8(17):7299–7312 <u>https://doi.org/10.1002/cam4.2582 PMID: 31605449</u>
- Yin Z, Yan X, Wang Q, et al. Detecting Prognosis Risk Biomarkers for Colon Cancer Through Multi-Omics-Based Prognostic Analysis and Target Regulation Simulation Modeling[J]. Front Genet, 2020, 11:524 https://doi.org/10.3389/fgene.2020.00524 PMID: 32528533
- Sakthikumar S, Roy A, Haseeb L, et al. Whole-genome sequencing of glioblastoma reveals enrichment of non-coding constraint mutations in known and novel genes[J]. Genome Biol, 2020, 21(1):127. https://doi.org/10.1186/s13059-020-02035-x PMID: 32513296