The American Journal of Applied Sciences
ISSN 2689-0992 Volume 08 - 2026

Proceedings of Women in Academia, Research and Management for Work-life Initiatives for Sustainable Health & Empowering Safety (WARM-WISHES 2026)

Differential Gene Expression and Pathway Analysis in Hepatocellular Carcinoma

Taru Gupta
Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, Uttar Pradesh, India.

Sonia Chadha
Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, Uttar Pradesh, India.

***Corresponding author**
Dr Sonia Chadha
Amity Institute of Biotechnology,
Amity University Uttar Pradesh,
Lucknow Campus,
Lucknow,

Received: 11 May 2026 | Received Revised Version: 24 May 2026 | Accepted: 12 June 2026 | Published: 23 June 2026

DOI: 10.37547/tajas/warm-30

Abstract

Hepatocellular carcinoma (HCC) is the most common primary malignancy of the liver and is one of the leading causes of cancer-related mortality (death) worldwide. Its incidence is particularly high in areas where hepatitis B and C virus infections are endemic, as well as in patients with chronic liver disease and cirrhosis. HCC is often diagnosed at an advanced stage because of the lack of specific early clinical symptoms, resulting in limited treatment options and poor prognosis. Recent advances in genomics, transcriptomics and bioinformatics have provided insight into the molecular mechanisms of HCC initiation and progression. Analysis of differential gene expression and functional enrichment and pathways analysis have helped identify key regulatory genes, hub genes and dysregulated signalling pathways associated with tumour development. In this report, we discuss the gene expression profiles of HCC by means of high-throughput data analysis, particularly on the identification of possible biomarkers for early diagnosis and novel therapeutic targets. We also discuss integration of bioinformatics tools like STRING, Cytoscape and DAVID for construction of interaction networks, functional annotation of genes and visualization of enriched pathways. The findings highlight the necessity of molecular profiling for the understanding of the complexity of HCC and the progress of personalized medicine. Further studies are needed to translate these molecular insights into effective diagnostic, prognostic and therapeutic strategies with improved clinical outcomes in HCC patients.

Keywords: Hepatocellular Carcinoma, Hub Gene, Primary Liver Tumor, DEGs, HCC.

Cite This Article: Gupta, T., & Chadha, S. (2026). Differential Gene Expression and Pathway Analysis in Hepatocellular Carcinoma. The American Journal of Applied Sciences, 357–366. https://doi.org/10.37547/tajas/warm-30

1. Introduction

It is the one of the most common cancer worldwide, Hepatocellular carcinoma is the most frequently and

recurrently occurring primary liver tumor and it is responsible for approximately 500,000 deaths annually, with the majority of cases occurring in Asian and African countries; however, the incidence rate is also increasing continuously in all continents (Llovet et al., 2021; Sung et al., 2021). Early-stage diagnosis of HCC remains a significant challenge due to the lack of specific symptoms. As a result, most patients are diagnosed at a very late stage (Taher et al., 2026). Recent technological advances and automation of RNA sequencing have enabled the scientific community to comprehensively analyse the whole transcriptome in large numbers of patients and within a comparatively short period of time (Wang et al., 2009). The Cancer Genome Atlas (TCGA), a collaborative initiative between the National Cancer Institute and the National Human Genome Research Institute under the National Institutes of Health (NIH), is a rich database containing genomic data on more than 30 different cancer types, including HCC (Weinstein et al., 2013). The expression data (cancer vs. control) from TCGA is routinely exploited by researchers around the world to derive vital insights for better diagnosis, molecular marker discovery, cancer subtypes classification and alternative therapeutic strategies exploration (Tomczak et al., 2015; Weinstein et al., 2013). In this study, we analysed gene expression data of HCC solid tumors versus controls to find differentially expressed genes, followed by pathway analysis. The top twenty hub genes in HCC were correlated with overall survival and disease-free survival to predict the potential prognostic value. Furthermore, we examined dysregulated kinases and discovered potential HCC subtypes, offering a complete description using survival data, gene expression profiles and biological pathway analyses of the differentially expressed genes in each subtype. (Agarwal et al., 2017.; Yamazoe et al., 2026; Li et al., 2026).

Microarray technology has become a routine high-throughput tool for molecular biologists, enabling to track the activity of thousands of genes in mRNA samples in a single run (Allison et al., 2006). These arrays have been used by researchers over the past twenty years to map the various gene-expression profiles associated with hepatocellular carcinoma. They have found a collection of genes that are likely responsible for the initiation and progression of the disease (Wurmbach et al., 2007). The candidate molecules identified in those early studies are now important reference anchors for testing new drug targets and for the development of clinical markers to improve diagnosis, prognosis and treatment for HCC patients. But even with such potential breakthroughs, standard interventions for liver cancer have achieved at best incremental gains, with surgical resection, radiation, conventional chemotherapy and a variety of targeted agents all providing little more than a modest improvement in the five-year survival figure, which remains painfully low, particularly in persons presenting with advanced disease. (Zhang et al., 2017). Aim of the project: To identify and analyse differentially expressed genes (DEGs) in hepatocellular carcinoma (HCC) using transcriptomic data-sets and explore molecular pathways associated with disease progression and prognosis. This includes microarray data collection, DEGs identifications, KEGG Pathway analysis, Protein - Protein interaction network construction and prediction of transcription factor (TF) and miRNA network.

2.Methodology

2.1 Microarray data acquisition

We used publicly available microarray datasets from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) to find differentially expressed genes (DEGs) between normal liver tissues and hepatocellular carcinoma (HCC) samples. We choose three datasets in particular: GSE45267, GSE62232, and GSE84402. The same microarray platform, the [GPL570] Affymetrix Human Genome U133 Plus 2.0 Array, was used to generate all three datasets. This made sure that probe annotation and expression profiling was the same across all three. To evaluate the normalization of expression data, box plots were created and used alongside volcano plots to illustrate significantly upregulated and downregulated DEGs for each specific dataset.

2.2 Identification of Differentially Expressed Genes (DEGs) Related To HCC

Differential expression analysis was performed using GEO2R tool (http://www.ncbi.nlm.nih.gov/geo/geo2r) on expression data obtained from NCBI GEO datasets. Using platform annotation files, the data were normalized and probes were mapped to gene symbols. Genes satisfying the criteria of [log2 fold change (log2FC)] > 1 and adjusted p - value < 0.05 were

considered significantly differentially expressed. We then used a VENNY 2.1.0 tool, to find common DEGs across all three datasets (Oliveros, 2017).

2.3 Functional Annotation and Enrichment Analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 tool (https://david.ncifcrf.gov/), based on DEGs was used to perform gene ontology (GO) functional annotation and Kyoto encyclopaedia of genes and genomes (KEGG) Pathway enrichment analysis (Huang et al., 2009, Sherman et al., 2022). The analysis was carried out separately for upregulated and downregulated gene sets. Gene Ontology (GO) terms were classified into three main categories: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). KEGG pathway analysis was also conducted to identify significantly enriched signalling and metabolic pathway associated with HCC. Only genes with p-value < 0.05 in their enrichment results were included as significantly associated with respective GO terms or KEGG pathways.

2.4 Protein-protein interaction (PPI) Network Construction and Hub Gene Screening

The Common DEGs were used to construct a PPI network using STRING database which is used for Search Tool for the Retrieval of Interacting Genes (STRING: http://string-db.org) (version 12.0) (Szklarczyk et al., 2023). Common DEGs were submitted into STRING with a very high-confidence interaction score threshold (≥0.7) to ensure only significant interactions were included. The resulting interaction data were imported into cytoscape v3.9.1 (https://cytoscape.org/) for visualization and further analysis.

Cytoscape was used for visualization and analysis of the PPI network constructed from DEGs associated with HCC. we used plugins such as CytoHubba designed for the determination of hub genes using topological criteria. Multiple algorithms like Degree, MCC (Maximal Clique Centrality), and MNC (Maximum Neighbourhood Component) were used for centrality-based ranking of nodes in the networks (Cline et al., 2007). The top 10 genes ranked based on their centrality scores were selected as hub genes, representing the most interconnected and potentially crucial regulator in the network.

2.5 miRNA–Hub Gene Network Construction

Network Analyst 3.0 (https://www.networkanalyst.ca/), a web-based platform for comprehensive visual analytics of gene-regulatory network, we constructed gene–miRNA interaction networks from the identified Hub genes in HCC. The list of top 10 hub genes was uploaded into Network Analyst, and regulatory interaction data were retrieved from curated datasets. miRNA-gene interactions, validated targeting miRNAs were sourced from miRTarBase database. The derived regulatory networks were visualized within Network Analyst environment and exported for further analysis which allows interpretation beyond visualization. This analysis helped identify crucial miRNAs that may be contributing to the regulation at HCC.

3. Results

3.1 Data Normalization and DEGs Visualization

Boxplots were employed to evaluate data normalization for each GEO2R analysis, and volcano plots were utilized to visualize the distribution of significantly upregulated and downregulated genes (DEGs) identified between HCC and normal liver tissue across datasets GSE 84402, GSE45267, GSE62232. (Figures 1,2 and 3)

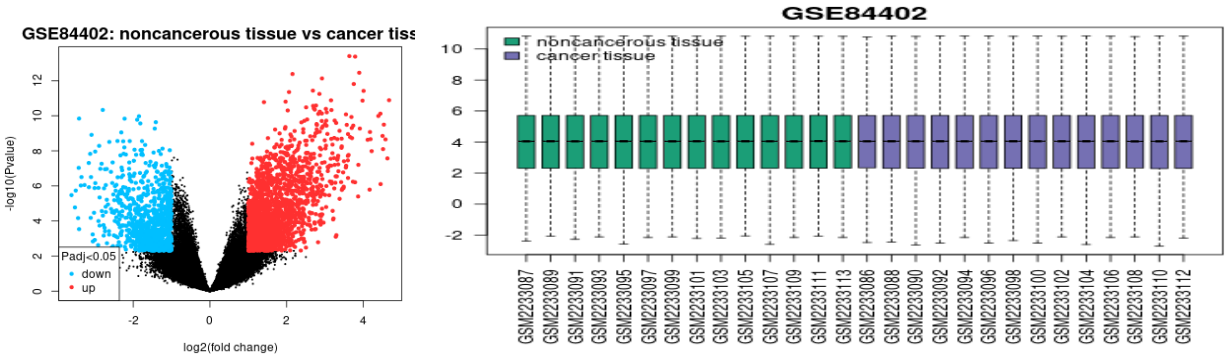

Figure 1: Volcano plot and box plot of dataset GSE8440

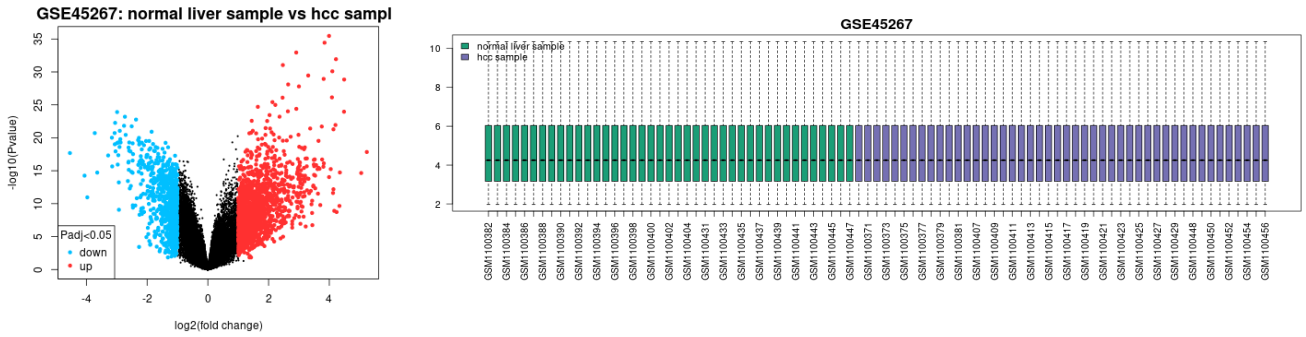

Figure 2: Volcano plot and box plot of dataset GSE45267

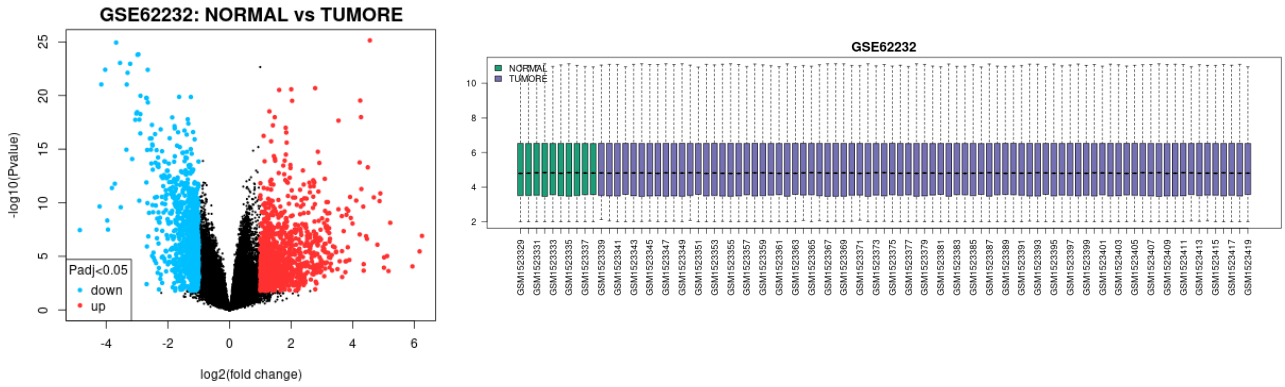

Figure 3: Volcano Plot AND Box Plot of the dataset GSE62232

3.2 Identification of Common DEGs Across Datasets

Venn diagrams were utilized to identify common DEGs by showcasing the overlapping genes among GSE45267, GSE84402 and GSE62232. As a result, CHI3L1 was identified as an up regulated gene and was the only common up regulated gene, while A2M, ABCC10, ABCC4, ABCF1, ABHD12, ABHD14B, and others were found to be down regulated, with a total of 1546 common down regulated genes (Figure 4).

Figure 4: Venn diagram illustrating the shared up regulated and down regulated genes between the three datasets GSE45267, GSE62232, and GSE84402

3.3 Functional Annotation and Enrichment Analysis of DEGs in HCC

The outcomes gained from the functional annotation DAVID are compiled in the table below. The enriched terms cover Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) as well as KEGG pathways. These visualizations highlight critical biological roles and molecular mechanism altered in HCC. The common DEGs were found to be associated establishment of chromosome localization, chromosome organization, regulation of chromosome segregation, ribonucleoprotein complex biogenesis, and regulation of telomerase RNA localization to Cajal bodies (Figure 5). The DEGs were enriched in cellular components involving spliceosomal complex, catalytic step 2 spliceosome, U2-type spliceosomal complex, and ribonucleoprotein complexes.

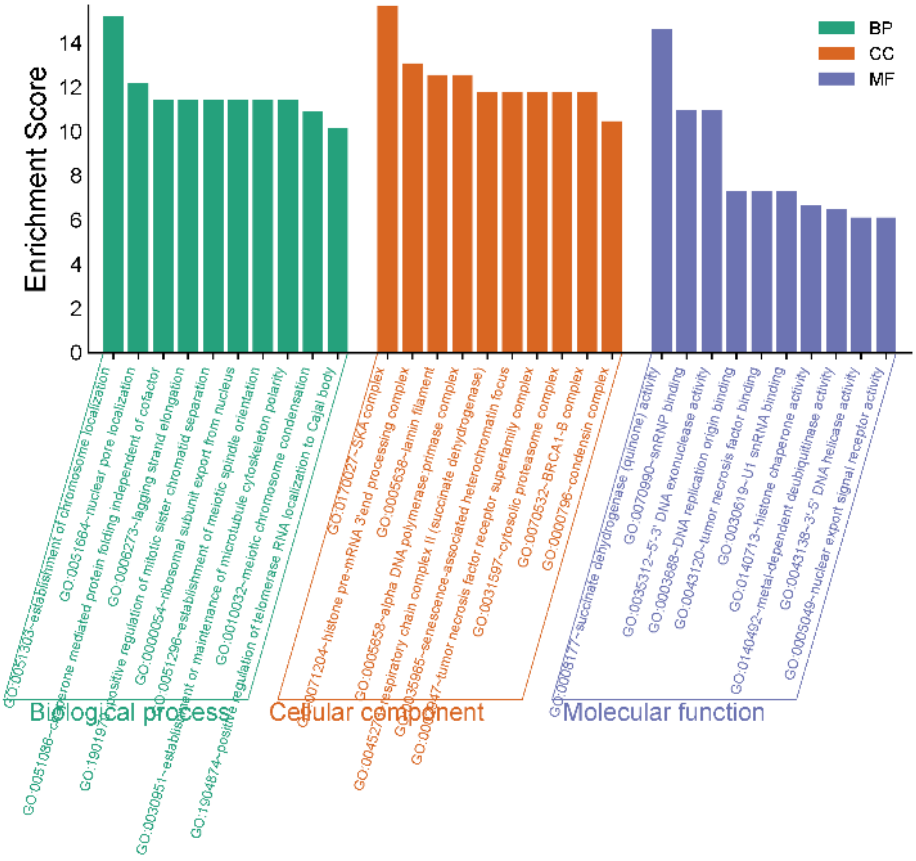

Figure 5: GO functional enrichment analysis of significant DEGs The Enriched GO term is represented along the horizontal axis. The negative Log10 (P-value) is represented along the vertical axis

3.4 Construction of PPI network and identification of Hub genes

A PPI network was created using STRING analysis with the common DEGs which further showed very strong interactions among them indicating their role as functionally interconnected teammates in HCC progression. (Figure 6) The STRING PPI Network was visualized by Cytoscape, where interactions were highlighted, and the important 10 hub genes were determined through CytoHubba based on high degree of interconnectivity illustrating importance in HCC molecular alterations. (Figure 7)

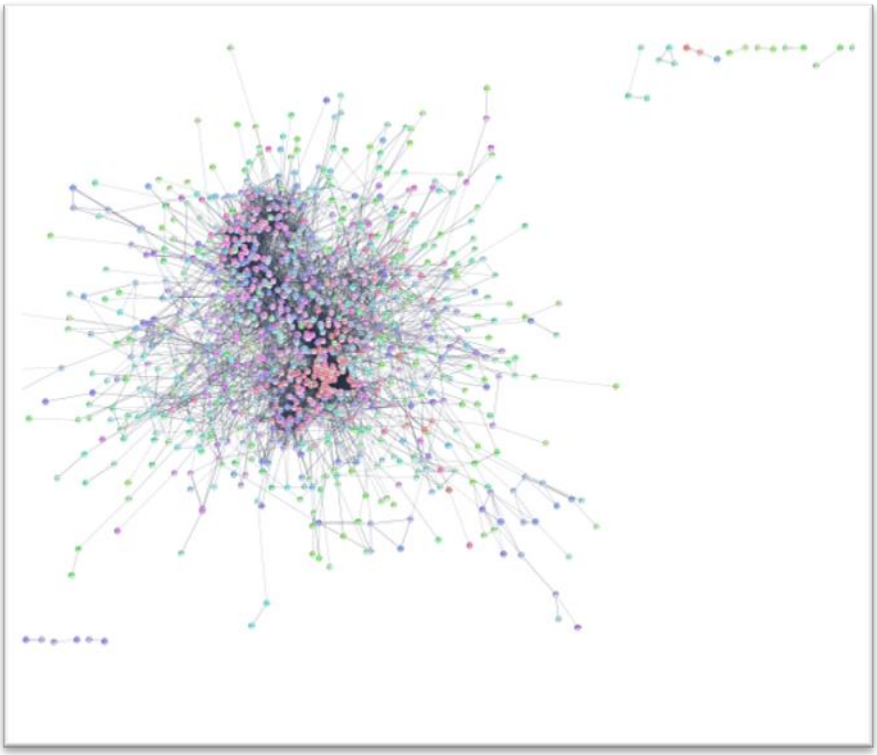

Figure 6: PPI network constructed using STRING database

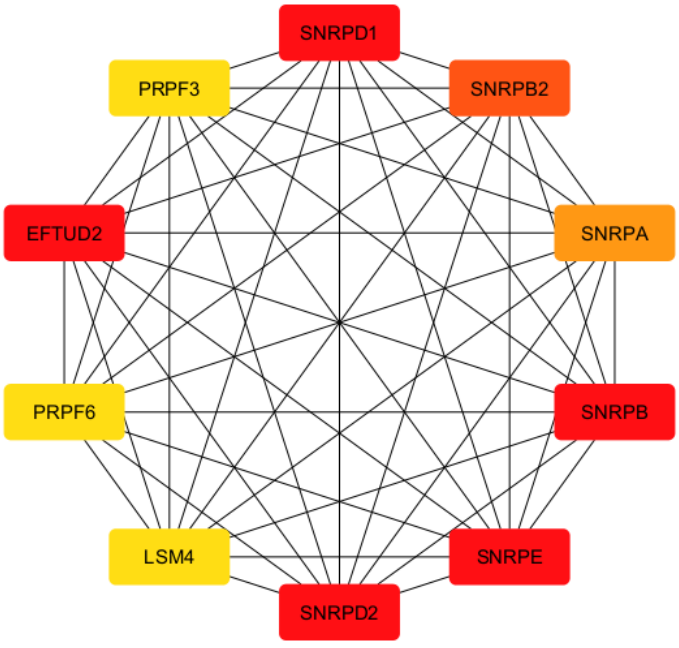

Figure 7: Interaction Network of Top Hub Genes Identified in the PPI Network- Top 10 Hub genes identified by the CytoHubba–a plugin in Cytoscape v3.10.1 using degree method. The degree scores are represented in yellow to orange colour. A darker colour indicates a higher degree score

3.5 Construction of miRNA–Hub Gene Network

To understand post-transcription regulation of the identified hub genes, miRNA-gene interaction analysis was performed. Using databases like miRNet, regulatory relationships between microRNAs and hub genes were predicted. (Figure 8), each yellow node represents a miRNA interacting with hub genes, suggesting possible regulatory miRNA involved in HCC pathogenesis. The miRNA–hub gene interaction network showed extensive post-transcriptional regulation of the identified hub genes. Several hub genes were targeted by multiple miRNAs, indicating a complex regulatory network controlling spliceosome-associated pathways.

Figure 8: miRNA-Hub gene interaction

3. Discussion

An integrated bioinformatics analysis of three independent HCC datasets was used to identify DEGs associated with HCC progression. The intersection of DEGs revealed CHI3L1 (chitinase-3-like protein 1) as a common up regulated gene and 1,546 common downregulated genes were identified. The use of multiple datasets reduced dataset-specific bias and increased the reliability of the identified molecular signatures. Several studies have reported CHI3L1 to promote tumor proliferation, angiogenesis, extracellular matrix remodeling, inflammation, and immune evasion in several malignancies, including HCC and increased levels of CHI3L1 have been found to be associated with poor prognosis, increased metastatic potential, and enhanced tumor aggressiveness. (Shao et al., 2009; Libreros et al., 2013). Thus, the present study supports the role of CHI3L1 as a diagnostic biomarker and therapeutic target in HCC. However, the downregulated genes such as A2M, ABCC10, ABCC4, ABCF1, ABHD12, ABHD14B, and are found to be involved in molecular transport, lipid metabolism, cellular

detoxification, immune regulation, and intracellular signalling (Reitz et al., 2015).

Gene ontology studies showed that HCC development involves alterations in chromosomal stability and RNA-processing mechanisms. Additionally cellular component enrichment analysis demonstrated involvement of the spliceosomal complex, catalytic step 2 spliceosome, U2-type spliceosomal complex, and ribonucleoprotein complexes pointing towards dysregulation of RNA splicing machinery which has increasingly been recognized as an important contributor to tumorigenesis (Dvinge et al., 2016; Yoshida et al., 2020). Molecular function enrichment analysis results suggest that transcriptional regulation and post-transcriptional RNA processing may play important roles in the molecular mechanisms underlying HCC progression.

Most of the hub genes identified by the Protein-protein interaction network analysis encode core components of the spliceosome machinery revealing coordinated regulation of pre-mRNA splicing in HCC. Previous studies have shown that spliceosomal dysregulation contributes to abnormal transcript processing, uncontrolled proliferation, and cancer progression. Therefore, these hub genes may represent key molecular drivers and potential therapeutic targets in HCC (Behrouzifar, 2023; Wang et al., 2022; Hershberger et al., 2026).

The miRNA–hub gene interaction network showed that the hub genes were targeted by multiple miRNAs, indicating a complex regulatory network controlling spliceosome-associated pathways. This complex network revealed that dysregulation of miRNA-mediated gene silencing may result in altered expression of spliceosomal genes and subsequently influence HCC pathogenesis (Callegari et al., 2013; Oura et al., 2019) which needs to be evaluated further.

Although the present study thus provides valuable insights into molecular mechanisms involved in HCC pathogenesis, there are several limitations. The findings were derived from publicly available transcriptomic datasets and therefore require validation using independent patient cohorts and experimental approaches. Furthermore, functional studies are necessary to elucidate the precise biological roles of CHI3L1, the identified hub genes, and their associated miRNAs in HCC progression.

4. Conclusion

This present study used three publically available datasets to identified differentially expressed genes (DEGs) and key molecular pathways associated with HCC. CHI3L1 was the common upregulated gene and 1,546 common downregulated genes were identified. Use of DAVID tool revealed that these DEGs were involved in chromosome organization, RNA processing, spliceosome-related pathways, and transcriptional regulation. Protein-protein interaction network analysis using cytoscape and its pluggin cytohubba identified 10 hub genes, most of which belonged to the small nuclear ribonucleoprotein (SNRP) family, indicating their role in RNA processing and tumor development. The miRNA regulatory network analyses provided additional insights into the molecular mechanisms governing these genes. Overall, the findings of the presentstudy reveal that aberrant RNA splicing, chromosomal regulation, and miRNA-mediated control are important mechanisms contributing to hepatocellular carcinoma development. CHI3L1 and the identified hub genes may serve as promising biomarkers for diagnosis and prognosis, as well as potential therapeutic targets for HCC**.** Further experimental validation is required to confirm their clinical significance and biological functions in hepatocellular carcinoma.

5. Declaration

The authors hereby declare that the manuscript submitted for consideration is an original work and has not been published or submitted elsewhere for publication. The authors take full responsibility for the integrity, accuracy, and ethical compliance of the work presented in the manuscript, including all revisions made in response to reviewer comments.

6. AI Usage Statement

Authors declare that AI tools, if used, were solely employed to improve the clarity, grammar, and language of the manuscript (as indicated in the reviewer's comments). No data, results, or scientific content were generated or altered using AI.

7. Conflict of Interest and Ethical Compliance

All authors confirm that:

i. Any potential conflicts of interest, whether financial or non-financial, have been fully disclosed. –Not Applicable

ii. All sources of funding and financial support received for the conduct of the study have been appropriately acknowledged, including any updates made during revision. –Not Applicable

iii. Necessary ethical approvals have been obtained from the relevant institutional or regulatory bodies for studies involving human participants, animals, or sensitive data, wherever applicable, and are clearly stated in the manuscript. –Not Applicable

9. References

1. Llovet, J. M., Kelley, R. K., Villanueva, A., et al. (2021). Hepatocellular carcinoma. Nature Reviews Disease Primers, 7(1), 6.
2. Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide. CA: A Cancer Journal for Clinicians, 71(3), 209–249.
3. Taher, M., Asri, M. A. I. M., Rahman, N. D. A., Zamzuri, N. E., Mohammad Farizal, N. S., Khotib, J., Susanti, D., Hasdan, N. A., Haris, M. S., & Rakhmawati, R. (2026). Serum biomarkers as personalised medicine for diagnostic and therapeutic approaches of hepatocellular carcinoma. Journal of the Egyptian National Cancer Institute, 38(1), 7. https://doi.org/10.1186/s43046-026-00342-1
4. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics, 10(1), 57–63.
5. Weinstein, J. N., Collisson, E. A., Mills, G. B., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics, 45(10), 1113–1120.
6. Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. Contemporary Oncology, 19(1A), A68–A77.
7. Agarwal, R., Narayan, J., Bhattacharyya, A., Saraswat, M., & Tomar, A. K. (2017). Gene expression profiling, pathway analysis and subtype classification reveal molecular heterogeneity in hepatocellular carcinoma and suggest subtype specific therapeutic targets. Cancer Genetics, 216–217, 37–51.
8. Yamazoe, T., Kakazu, E., Matsuda, M., et al. (2026). Genomic and transcriptomic profiling of hepatocellular carcinoma in patients with Fontan-associated liver disease. Hepatology. doi: 10.1097/HEP.0000000000001693
9. Li, P., Guo, A., Zhao, M., & Chen, G. (2026). Comparative transcriptomics and computational drug discovery identify ASPM as a key oncogenic driver and therapeutic target in hepatocellular carcinoma. Frontiers in Bioinformatics, 6, Article 1795889.doi: 10.3389/fbinf.2026.1795889
10. Allison, D. B., Cui, X., Page, G. P., & Sabripour, M. (2006). Microarray data analysis: From disarray to consolidation and consensus. Nature Reviews Genetics, 7(1), 55–65.
11. Wurmbach, E., Chen, Y. B., Khitrov, G., et al. (2007). Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. Hepatology, 45(4), 938–947.
12. Zhang, C., Peng, L., Zhang, Y., Liu, Z., Li, W., Chen, S., & Li, G. (2017). The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. Medical Oncology, 34(6), 101.
13. Oliveros, JC. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html
14. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57. doi: 10.1038/nprot.2008.211. PMID: 19131956
15. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022 Jul 5;50(W1): W216-W221. doi: 10.1093/nar/gkac194. PMID: 35325185; PMCID: PMC9252805.
16. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Annika, G. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research, 51(D1), D638–D646.

17. Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., ... Bader, G. D. (2007). Integration of biological networks and gene expression data using Cytoscape. Nature Protocols, 2(10), 2366–2382.

18. Shao, R., Hamel, K., Petersen, L., et al. (2009). YKL-40, a secreted glycoprotein, promotes tumor angiogenesis. Oncogene, 28(50), 4456–4468. doi: https://doi.org/10.1038/onc.2009.292

19. Libreros, S., Garcia-Areas, R., & Iragavarapu-Charyulu, V. (2013). CHI3L1 plays a role in cancer progression. International Journal of Cancer, 133(6), 1286–1295. doi: https://doi.org/10.1002/ijc.28138

20. Reitz, K., Yao, P., & Emmons, G. T. (2015). ABHD family proteins and lipid metabolism. Biochimica et Biophysica Acta, 1851(6), 626–635. doi: https://doi.org/10.1016/j.bbalip.2014.09.01

21. Dvinge, H., Kim, E., Abdel-Wahab, O., & Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. Nature Reviews Cancer, 16(7), 413–430. doi: https://doi.org/10.1038/nrc.2016.51

22. Yoshida, K., & Ogawa, S. (2020). Splicing factor mutations and cancer. Wiley Interdisciplinary Reviews: RNA, 11(6), e1573. doi: https://doi.org/10.1002/wrna.1573

23. Behrouzifar, S. (2023). Identifying downregulated hub genes and key pathways in HBV-related hepatocellular carcinoma using systems biology approach. doi: https://doi.org/10.48550/arXiv.2306.16173

24. Wang, H., Xu, F., Lu, L., et al. (2022). The diagnostic and prognostic significance of SNRPD1 aberrantly high expression in hepatocellular carcinoma. Journal of Cancer, 13(1), 184–201. doi: https://doi.org/10.7150/jca.65225

25. Hershberger, C. E., Daniels, N. J., Patterson, W. M., & Rotroff, D. M. (2025). The alternative splicing landscape of hepatocellular carcinoma and its potential for HCC detection. Hepatology Communications, 10(1), e0838. doi: 10.1097/HC9.0000000000000838

26. Callegari, E., Elamin, B. K., Sabbioni, S., Gramantieri, L., & Negrini, M. (2013). Role of microRNAs in hepatocellular carcinoma. Journal of Cellular Physiology, 228(4), 625–634. doi: https://doi.org/10.1002/jcp.24224

27. Oura, K., Morishita, A., Masaki, T. (2019). Molecular and functional roles of microRNAs in hepatocellular carcinoma. International Journal of Molecular Sciences, 20(24), 6329. doi: https://doi.org/10.3390/ijms20246329