Exploration of Potential Biomarkers and Immune Landscape for Hepatoblastoma: Evidence from Machine Learning Algorithm

Peng Zhou,1 Shanshan Gao,2 and Bin Hu1

1Department of Pediatric, Maternal and Child Health Hospital, Zibo, China
2Department of Ultrasound, Zibo Forth People's Hospital, Zibo, China

Correspondence should be addressed to Bin Hu; 18409120@masu.edu.cn

Received 3 March 2022; Accepted 2 June 2022; Published 31 July 2022

Academic Editor: Qing Li

1. Introduction

The most prevalent pediatric liver tumor is hepatoblastoma (HB), which mostly affects children under the age of four. HB is an uncommon pediatric cancer with an annual incidence of 1.5 cases per million [1]. Up to 80% of all patients diagnosed with cancer survive thanks to complete surgical resection and treatment. Unfortunately, immunosuppression may have long-term negative effects in survivors [2]. The absence of excellent early diagnosis techniques is the primary cause of poor prognosis. Nowadays, clinicians identify the condition based on clinical signs, imaging, and methotrexate levels. Due to the many sources of AFP in patients, the sensitivity and specificity are insufficient [3]. A case reported that five individuals with normal AFP levels were diagnosed with HB [4]. As a result, novel and robust biomarkers must be discovered in order to create effective diagnostic and therapeutic procedures for HB patients.

With the development of high-throughput sequencing technology, more and more aberrantly expressed mRNAs have been identified in HB [5–7]. Moreover, the elevated N6-methyladenosine alteration in HB represents a carcinogenic pathway [8]. Several studies have also found that various genes, such as zinc finger antisense 1 [9] and TUG1 [10], are involved in proliferation, apoptosis, and glutaminolysis. Importantly, the liver has specific histology and microenvironment that controls tumor growth and therapeutic outcome: dual blood supply, vascularization of porus blood sinuses, and the presence of different mesenchymal cells [11]. The liver exhibits a specific immune response to tumor cells, which correlates with poor responsiveness to immunotherapy [12]. Therefore, assessing the altered immune microenvironment in HB could provide a more reliable therapeutic strategy for patients. Because of the various analytical methodologies, experimental methods, and sample sizes used in the investigations, the findings are debatable. As a result, more bioinformatics investigations from public databases might help to reveal the immune landscape and novel biomarkers panel in HB patients.
The study aimed to reveal reliable differentially expressed genes in HB using datasets from the Gene Expression Omnibus (GEO) database. Subsequently, we discovered hub genes using machine learning methods and utilized them to build a neural network for diagnosis. In addition, we studied the status of the immune microenvironment as well as hundreds of tumor microenvironment-related pathways in depth. We hope that the biomarkers panel in this study will lead to the development of novel diagnostic and prognostic for HB patients.

2. Methods

2.1. Datasets and Data Preprocessing. RNA-seq datasets were downloaded from the GEO database [13]: GSE131329 datasets based on the GPL6244 platform (14 noncancerous
Figure 3: Enrichment analysis. (a) The terms of KEGG enrichment analysis. (b) The terms of GO enrichment analysis.

Figure 4: Differential expression analysis. (a) The process of constructing random forest trees. (b) Ranking of gene importance. (c) The heatmap of the top 10 genes in the screening dataset.
liver tissue samples and 53 HB samples) and GSE133039 datasets based on the GPL16791 platform (32 noncancerous liver tissue samples and 31 HB samples). The GSE131329 datasets were defined as screening sets, and the GSE133039 dataset was used as an external validation set.

2.2. Screening of Biomarkers Panel. Differentially expressed genes (DEGs) were identified in the screening set using the “limma” package in R software [14]. We selected $|\log_2\text{fold change FC}| > 2$ and adj-$P$.value < 0.05 as criteria. The random forest tree [15] (the optimal number of variables in the binomial tree in the node is 6, the optimal number of trees contained in the random forest is 2000, and the top10 genes in importance analysis were selected) was used to identify 10 hub genes as final biomarkers panel.

2.3. Neural Network. We selected the screening set for neural network model training, and the validation set was tested using the same process. A neural network model based on a biomarkers panel was constructed using the “neuralnet” package [16] in R software after normalizing the data to the maximum and minimum values. Subsequently, four hidden layers were set as model parameters, and the classification model of the disease (HB or normal) was constructed by the obtained gene weight information. It is worth noting that the sum of the product of the weight score and the expression level of significant genes is used as the disease classification score. In addition, we performed a 10fold validation of the model results. Finally, the “pROC” package [17, 18] in R software was used to calculate the AUC value in classification performance.

2.4. Identification of Hub MCODEs in Protein-Protein Network and Enrichment Analysis. We used metascape online tools [19] to construct hub MCODEs in DEGs and annotated biological functions in each hub MCODEs. GO enrichment analysis [20] is a commonly used bioinformatics method for comprehensive information, including biological process, cellular component, and molecular function. In addition, KEGG pathway enrichment analysis [21] is widely used to

Figure 5: Construction and validation of the artificial neural network model. (a) Construction of the artificial neural network model. (b) ROC curve in screening dataset. (c) ROC curve in the validation dataset.
explore biological mechanisms and functions. We selected q value and P value < 0.05 as criteria.

2.5. Analysis of Tumor Immune Microenvironment. We used the single-sample gene-set enrichment analysis (ssGSEA) to quantify the tumor immune microenvironment (TIME) associated pathways in each tissue sample. The “TOBR” package [22] was used to download and analyze the gene sets of “TME-associated pathways.” In addition, xCell, EPIC, MCPcounter, and timer algorithms were used to analyze the immune infiltration levels of cells in TME. The heatmap diagram was used to compare the differences between immune cells and TME-associated pathways in different tissues.

2.6. Cell Lines and Quantitative Real-Time PCR (qRT-PCR). The HB cell lines (SMMC-7721) and the normal human hepatic cell line (L02) were cultured in 10% FBS DMEM medium in an environment of 5% CO2 and 37°C [23]. Primer sequences are summarized in previous studies, including TRIB1 [24], MAT1A [25], SAA2-SAA4 [26], NAT2 [27], HABP2 [28], CYP2CB [29], APOF [30], and CFHR3 [31]. Detailed experimental procedures are described in our previous publications [32, 33].

2.7. Statistics. We implemented all statistical analyses with R version 4.1.1. Wilcoxon test was used to screen infiltrating immune cells or scores of TME-associated pathways with statistically significant. P < 0.05 was considered statistically significant.

3. Results

3.1. Differential Expression Genes in Different Samples. We performed differentially expressed genes analysis in the screening set and finally identified 277 DEGs, and the heatmap shows the top30 DEGs (Figure 1(a)). The volcano plot demonstrated 90 up-regulated genes as well as 187 down-regulated genes (Figure 1(b)).

3.2. Hub Modules in DEGs Network. Considering the interactions between DEGs, we differentiated and functionally identified the modules in 277 genes (Figure 2). In the red module, it was mainly composed of CYP and UGT family genes and was associated with DNA adducts, biological oxidations, and xenobiotic processes. In the blue module, it was mainly composed of H2BC5, AR, SDS, H2BC3, and H2AC13; the module was associated with activated PKN1 stimulating transcription of AR-regulated genes, condensation of prophase chromosomes, and HDACs deacetylate...
3.3. Enrichment Analysis. We revealed that 277 DEGs were significantly associated with HB in differentially expressed genes analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was used to annotate these DEGs (Figure 3(a)). Similar to the results enriched in the hub modules, the results of KEGG analysis also showed that DNA adducts, bile secretion, and metabolism of xenobiotics by cytochrome P450 played a key role in the occurrence of HB. The GO analysis showed that the xenobiotic metabolic process, long-chain fatty acid metabolic process, and olefinic compound metabolic process were associated with HB (Figure 3(b)).

3.4. Construction of Biomarkers Panel. The random forest tree method was used to sort the weights of all DEGs (Figures 4(a) and 4(b)). Selecting the top 10 genes for our
3.5. Construction of the Artificial Neural Network Model. We built an artificial neural network model based on the "neuralnet" package. The min-max method [0, 1] was chosen for normalizing the data segment and scale data, and then, the neural network was trained to normalize the maximum and minimum values of the data, the number of hidden layers was set to 5 before starting the computation, and the parameter for the number of neurons was set to 6. The neural network model, including a total of 4123 steps, had been carried out in the training set (Figure 5(a)). In order to evaluate the results of the neural network model more effectively, we choose the method of 10 cross-validations. After cross-validation, the area under the ROC curve is close to 1 (AUC = 0.972), indicating robustness (Figure 5(b)). It was worth noting that in the validation set, the AUC of the neural network model was 0.87 (Figure 5(c)).

3.6. The Landscape of Infiltrating Immune Cells and TME-Associated Pathways. We used the four deconvolution methods to characterize the immune cell pattern in HB. In the EPIC algorithm, the proportions of B cells and macrophage cells were significantly decreased in the tissues of HB patients. However, the proportions of CD8 T cells, CAFs, and endothelial cells were significantly increased in the tissues of HB patients (Figure 6(a)). In the MCPcounter algorithm, only the proportions of CD8 T cells were significantly decreased in the tissues of HB patients (Figure 6(b)). In the TIMER algorithm, four different types of immune cells also differed significantly between the two groups (Figure 6(c)). Finally, both the xCELL algorithm (Figure 6(d)) with more results and the enrichment results of TME-related pathways (Figure 7) showed a more active immune response in the HB tissue. The above results showed that the occurrence and development of HB were closely related to TME.

3.7. Quantitative Real-Time PCR (qRT-PCR). For validating the expression of hub genes, HB cell lines (SMMC-7721) and L02 (L02) were used to conduct a qRT-PCR assay. In the validation of TRIB1, MAT1A, SAA2-SAA4, NAT2, HABP2, CYP2CB, APOF, and CFHR3, the expression of SMMC-7721 was higher than that of L02 (Figure 8). The above results suggested that these hub genes may be potential oncogenic genes, but further experimental verification was still needed.

4. Conclusion

We developed the HB diagnostic classification model employing a random tree and 10 hub genes in this study. We also used an independent dataset to test the classification efficiency based on an artificial neural network. As a result, the model might aid with HB diagnosis. We also used GO and KEGG analyses to discover that the DEGs were mostly involved in DNA adducts, bile secretion, metabolism of xenobiotics by cytochrome P450, xenobiotic metabolic process, long-chain fatty acid metabolic process, and olefinic compound metabolic process. We studied the status of the TME as well as hundreds of TME-related pathways in depth. Finally, all algorithms and the enrichment results of TME-related pathways showed a more active immune response in the HB tissue.

The 10 genes in our panel were as follows: TNFRSF19, TOP2A, TRIB1, MAT1A, SAA2-SAA4, NAT2, HABP2, CYP2CB, APOF, and CFHR3. Unfortunately, current basic research in HB had not addressed the above-mentioned genes in depth. Therefore, the discussion focused on the function of
tissue than in normal tissue in one investigation. Reduced TNFRSF19 led to increased proliferation and invasion of HCC cells, implying that TNFRSF19 may function as a tumor suppressor [38]. The TOP2A can encode a DNA topoisomerase that is responsible for controlling and altering DNA topology during transcription [39]. In addition, TOP2A expression in different cancers is considered to be a favorable prognostic biomarker for predicting cancer progression and recurrence, and it can also serve as a risk factor for low survival [40]. In HCC, several studies have shown TOP2A up-regulation [41, 42]. TRIB1 has a conserved motif similar to the catalytic domain of serine/threonine kinases but lacks the ATP-binding or kinase catalytic domain [43]. TRIB1 acts as a scaffold or bridging protein, promotes the degradation of target proteins, and regulates several important signaling pathways [44]. Due to physical interactions, TRIB1 suppresses the tumor suppressor gene p53, the most commonly mutated gene in HCC, which plays a key role in many cancers [45]. Mammals have three different forms of MAT (MAT1, MATII, and MATIII), encoded by two different genes (MATIA and MAT2A). MAT in the liver and extrahepatic tissues is the product of two genes, MATIA and MAT2A, respectively [46]. MATIA is a strong prognostic indicator for HCC, and data from HCC patients with reduced MATIA suggest that the RETOME pathway may be involved in HCC tumorigenesis [47]. However, the remaining genes are equally poorly studied in HCC, which reminds us of the gaps in the field of HB-related hub gene research. In conclusion, we have developed a robust biomarkers panel for HB patients.

5. Conclusions

We studied the status of the immune microenvironment as well as hundreds of tumor microenvironment-related pathways in-depth and revealed more active immune responses in the HB compared to normal tissues. Moreover, we hope that the biomarkers panel will lead to diagnostic development for HB patients.

Data Availability

Data are available at the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Ethical Approval

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Peng Zhou and Bin Hu conceived and designed the study. Shanshan Gao drafted the article. All authors had final approval of the submitted versions.

References

[1] C. N. Grant, D. Rhee, E. T. Tracy et al., "Pediatric solid tumors and associated cancer predisposition syndromes: workup, management, and surveillance. A summary from the APSA Cancer Committee," Journal of Pediatric Surgery, vol. 57, no. 3, pp. 430–442, 2022.
[2] J. Hager and C. M. Sergi, "Hepatoblastoma," in Liver Cancer, C. M. Sergi, Ed., Exon Publications, Brisbane, Australia, 2021.
[3] H. Chen, Q. Guan, H. Guo, L. Miao, and Z. Zhuo, "The genetic changes of hepatoblastoma," Frontiers in Oncology, vol. 11, Article ID 690641, 2021.
[4] S. J. Shi, D. L. Wang, W. Hu, F. Peng, and Q. Kang, "Ex vivo liver resection and autotransplantation with cardiopulmonary bypass for hepatoblastoma in children: a case report," Pediatric Transplantation, vol. 22, no. 7, Article ID e13268, 2018.
[5] G. Nagae, S. Yamamoto, M. Fujita et al., "Genetic and epigenetic basis of hepatoblastoma diversity," Nature Communications, vol. 12, no. 1, p. 5423, 2021.
[6] H. Lee, T. El Jabbour, S. Ainechi et al., "General paucity of genomic alteration and low tumor mutation burden in refractory and metastatic hepatoblastoma: comprehensive genomic profiling study," Human Pathology, vol. 70, pp. 84–91, 2017.
[7] R. Sun, S. Li, K. Zhao, M. Diao, and L. Li, "Identification of ten core hub genes as potential biomarkers and treatment target for hepatoblastoma," Frontiers in Oncology, vol. 11, Article ID 591507, 2021.
[8] G. Li, L. Deng, N. Huang et al., "m6A mRNA methylation regulates LKB1 to promote autophagy of hepatoblastoma cells through upregulated phosphorylation of AMPK," Genes, vol. 12, no. 11, p. 1747, 2021.
[9] X. Cui, Z. Wang, L. Liu et al., "The long non-coding RNA ZFAS1 sponges miR-193a-3p to modulate hepatoblastoma growth by targeting RALY via HGF/c-Met pathway," Frontiers in Cell and Developmental Biology, vol. 7, p. 271, 2019.
[10] M. X. Yuan, C. Y. Ji, H. Q. Gao, X. Y. Sheng, W. X. Xie, and Q. Yin, "IncRNA TUG1 regulates angiogenesis via the mR-204-5p/JAK2/STAT3 axis in hepatoblastoma," Molecular Medicine Reports, vol. 24, no. 2, p. 553, 2021.
[11] W. J. Ho, L. Danilova, S. J. Lim et al., "Viral status, immune microenvironment and immunological response to checkpoint inhibitors in hepatocellular carcinoma," Journal for Immunotherapy of Cancer, vol. 8, no. 1, Article ID e00394, 2020.
References

[12] J. Fu, W. Z. Li, N. A. McGrath et al., "Immune checkpoint inhibitor associated hepatotoxicity in primary liver cancer versus other cancers: a systematic review and meta-analysis," *Frontiers in Oncology*, vol. 11, Article ID 650292, 2021.

[13] S. Feng, T. Xia, Y. Ge et al., "Computed tomography imaging-based radiogenomics analysis reveals hypoxia patterns and immunological characteristics in ovarian cancer," *Frontiers in Immunology*, vol. 13, Article ID 868067, 2022.

[14] M. E. Ritchie, B. Hippsion, D. Wu et al., “Limma powers differential expression analyses for RNA-sequencing and microarray studies,” *Nucleic Acids Research*, vol. 43, no. 7, p. e47, 2015.

[15] M. Savargiv, B. Masoumi, and M. R. Keyvanpour, "A new random forest algorithm based on learning automata," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 5572781, 19 pages, 2021.

[16] M. W. Beck, "NeuralNetTools: visualization and analysis tools for neural networks," *Journal of Statistical Software*, vol. 85, no. 11, pp. 1–20, 2018.

[17] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.

[18] S. Feng, H. Yin, K. Zhang et al., "Integrated clinical characterististics and omics analysis identifies a ferroptosis and iron-metabolism-related lncRNA signature for predicting prognosis and therapeutic responses in ovarian cancer," *Journal of Ovarian Research*, vol. 15, no. 1, 2022.

[19] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, p. 1523, 2019.

[20] G. Ontology Consortium, "Gene Ontology consortium: going forward," *Nucleic Acids Research*, vol. 43, pp. D1049–D1056, 2015.

[21] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[22] Y. Zhu, S. Feng, Z. Song, Z. Wang, and G. Chen, "Identification of immunological characteristics and immune subtypes based on single-sample gene set enrichment analysis algorithm in lower-grade glioma," *Frontiers in Genetics*, vol. 13, Article ID 894865, 2022.

[23] G. Liu, Q. Zhu, H. Wang, J. Zhou, and B. Jiang, "Long non-coding RNA Linc00205 promotes hepatoblastoma progression through regulating microRNA-154-3p/Rho-associated coiled-coil Kinase 1 axis via mitogen-activated protein kinase signaling," *Aging* (Albany NY), vol. 14, no. 4, pp. 1782–1796, 2022.

[24] L. J. Zhang, F. Wang, P. Y. Qi, W. Y. Zhou, and B. Wang, "miR-513b-5p inhibits the proliferation and promotes apoptosis of retinoblastoma cells by targeting TRIB1," *Open Medicine*, vol. 16, no. 1, pp. 1364–1371, 2021.

[25] H. Yang, M. E. Cho, T. W. Li et al., "MicroRNAs regulate methionine adenosyltransferase 1A expression in hepatocellular carcinoma," *Journal of Clinical Investigation*, vol. 123, no. 1, pp. 285–298, 2013.

[26] J. M. Zamora-Fuentes, E. Hernández-Lemus, and J. Espinal-Enriquez, "Gene expression and co-expression networks are strongly altered through stages in clear cell renal carcinoma," *Frontiers in Genetics*, vol. 11, Article ID 578679, 2020.

[27] X. Li, H. Zhang, L. Xu et al., "miR-15a-3p protects against isoniazid-induced liver injury via suppressing N-acetyltransferase 2 expression," *Frontiers in Molecular Biosciences*, vol. 8, Article ID 752072, 2021.
facilitates HCC progression by regulating Snail expression," 
*Bioengineered*, vol. 12, no. 2, pp. 12967–12979, 2021.

[43] Z. Hegedus, A. Czibula, and E. Kiss-Toth, “Tribbles: novel 
regulators of cell function; evolutionary aspects,” *Cellular and 
Molecular Life Sciences*, vol. 63, no. 14, pp. 1632–1641, 2006.

[44] R. Cunard, “Mammalian tribbles homologs at the crossroads 
of endoplasmic reticulum stress and Mammalian target of 
rapamycin pathways,” *Scientifica*, vol. 2013, Article ID 750871, 
17 pages, 2013.

[45] C. Miyajima, Y. Inoue, and H. Hayashi, “Pseudokinase 
tribbles 1 (TRB1) negatively regulates tumor-suppressor ac-
tivity of p53 through p53 deacetylation,” *Biological and 
Pharmaceutical Bulletin*, vol. 38, no. 4, pp. 618–624, 2015.

[46] P. Y. Chu, H. J. Wu, S. M. Wang, P. M. Chen, F. Y. Tang, and 
E. P. I. Chiang, “MAT2A localization and its independently 
prognostic relevance in breast cancer patients,” *International 
Journal of Molecular Sciences*, vol. 22, no. 10, p. 5382, 2021.

[47] P. M. Chen, C. H. Tsai, C. C. Huang et al., “Downregulation of 
methionine cycle genes MAT1A and GNMT enriches protein-
associated translation process and worsens hepatocellular 
carcinoma prognosis,” *International Journal of Molecular 
Sciences*, vol. 23, no. 1, p. 481, 2022.