Genes & Diseases (2023) 10, 1194–1196

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

RAPID COMMUNICATION

Integrative multi-platform meta-analysis of hepatocellular carcinoma gene expression profiles for identifying prognostic and diagnostic biomarkers

Hepatocellular carcinoma (HCC), as one of the most common types of primary liver cancer (PLV), accounts for approximately 75%–90% of all PLV cases identified worldwide.[1] Currently, the most effective treatment for HCC patients constitutes liver transplantation. However, due to high rates of recurrence a poor prognosis is predicted.[2] Therefore, accurate HCC biomarkers are urgently needed to develop innovative therapeutics. Most of earlier investigations to identify biomarkers have been severely limited by sample size, as those studies used only data sets generated on the same chip platform. Considering the decisive role of sample size, we aimed to identify and validate diagnosis and prognosis biomarkers associated with HCC based on the expression data of seven transcriptome microarray gene expression (MAGE) datasets from two different platforms, namely Illumina and Affymetrix. To discover differentially expressed genes (DEGs), a meta-analysis based on an empirical Bayesian approach (ComBat) was first applied to the gene expression profiles of 939 tissue samples (Table S1). Irigoyen et al[3] technique was used to preprocess and integrate the data. Data import, data filtering, and normalization for each experiment were carried out separately (Text S1.1). Then, using the bioinformatics technique of weighted correlation network analysis (WGCNA), the highly associated prioritized DEGs (Text S1.2, S1.3) were clustered. To identify sub-clusters and confirm candidate genes, co-expression networks and topological analysis were applied (Text S1.4). Finally, for these genes, the least absolute shrinkage and selection operator (LASSO) regression model was used and validated to provide a diagnostic model (Text S1.5). To construct a prognostic model, the Cox proportional hazard regression analysis was applied and validated (Text S1.6). All statistical tests were two-sided (Text S1.7) and the results with an adjusted p-value ≤ 0.05 were considered statistically significant (Text S1.8). The flowchart diagram is presented in Figure 1A. The comparative boxplots and PCA plots in Fig. S1A–D, confirm the efficacy of the ComBat for batch effect removal. From the meta-analysis, 292 genes were identified as DEGs, satisfying the criteria of the absolute value of log2-fold change (logFC) > 1 and false discovery rate (FDR) < 0.05 (Table S2). Moreover, to assess bias and reproducibility across microarray experiments, we modified a comparison of individual analyses using data from two platforms and a meta-analysis (Table S3, S4). The result of integrative meta-analysis after prioritization showed 239 common DEGs (Text S2.1 and Table S5). We found possible relationships between the expression profiles of 239 mutual DEGs using Pearson's correlation coefficients (PCCs). The hierarchical cluster tree and topological overlapping matrix were used to screen out cluster modules. DEGs were divided into four parts, of which blue, red, and turquoise modules were considered the most significant parts (Fig. 1B, a). Through the application of the molecular complex detection (MCODE) plug-in, the sub-clusters of differential co-expression module (DCEM) were found and visualized (Fig. 1B, b).

The 20 genes from sub-clusters were served as key genes for diagnosis and prognosis analysis.

Biological process and pathway enrichment analysis showed that the modules primarily focused on cellular processes, metabolic processes and biological regulation in the p53 pathway (Text S2.2). To more deeply validate results, the top five hub genes (*CDK4*, *CCNB1*, *IGF1*, *SOCS2*,

Peer review under responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2022.07.018

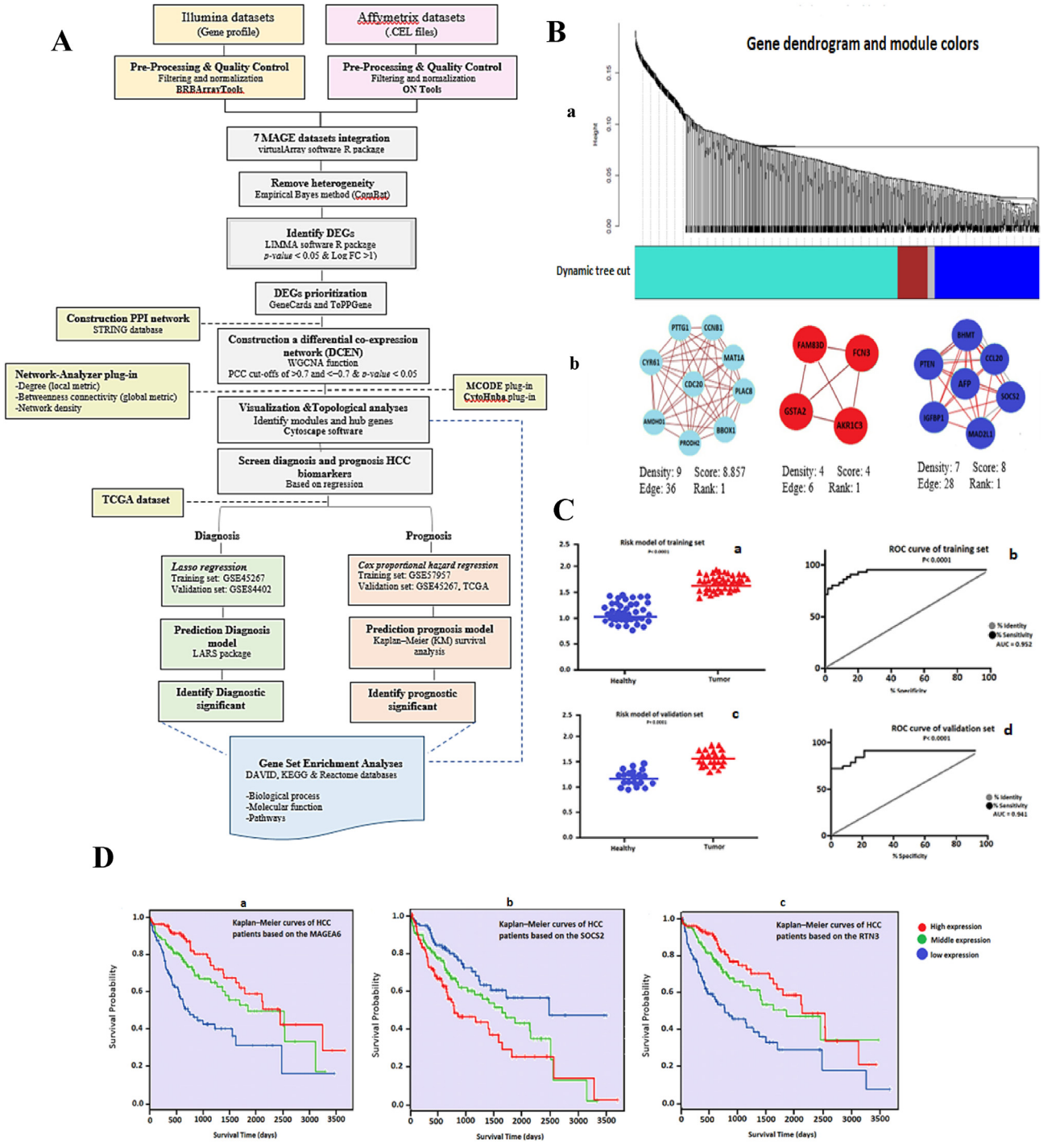

Figure 1 Methodological approach and results. **(A)** The flowchart diagram outlined the procedure step-by-step of the whole integrated meta-analysis to screen specific HCC biomarkers by integrating multi-platform MAGE datasets. **(B)** The WGCNA analysis to identify differentially expressed co-expression modules: **(a)** The result of hierarchical clustering is represented via a Gene Cluster Dendrogram. In the dendrogram, each line represents a gene. The colored column below represents the module conducted by the static tree cutting method. The turquoise, blue and red colors show DCEM; **(b)** Visualization of 20 key genes from sub-clusters DCEM (turquoise, blue and red module) and their sub-network topology. **(C)** Construction of risk model for diagnosis of hepatocellular carcinoma and receiver operating characteristic (ROC) curve: **(a)** Distribution of risk scores of healthy and tumor group of the training set; **(b)** ROC curve of a risk score for the training set (Area under the ROC Curve or AUC = 0.952); **(c)** the distribution of risk scores of healthy and tumor group of the validation set; **(d)** ROC curve of risk score of validation set (AUC = 0.941). **(D)** Kaplan—Meier curves of HCC patients of three-gene prognostic models. Patients were divided into 3 groups, and the lower third of the patients was defined as having low mRNA expression, while the upper third had high mRNA expression: **(a)** Kaplan—Meier curves of HCC patients based on MAGEA6; **(b)** Kaplan—Meier curves of HCC patients based on SOCS2; **(c)** Kaplan—Meier curves of HCC patients based on RTN3.

and *AFP*) according to the highest score in topological features were presented (Table S6).

We set AUC $>$ 0.90 as the cut-off value to get the simplified diagnostic risk model with a minimum quantity of gene count. A three-gene diagnostic risk model was built: $-0.3034*Exp_{(CYP2E1)} + 0.3526*Exp_{(AKR1C3)} + 0.4645*Exp_{(AFP)}$. (Text S2.3, Fig. 1C: a-d). Recently reported application of AFP as a biomarker for early diagnosis of HCC has been limited because of the high false-positive rate and its low sensitivity (55%). Furthermore, its overexpression is detected in other liver abnormalities and other tumors. Consequently, the specificity of this biomarker is not too high (87%).[4] CYP2E1 is an enzyme important in ethanol metabolism, and our results show its increased activity in HCC. AKR1C3 may play a role in controlling cell growth and/or differentiation. Zhu et al[5] indicated that AKR1C3 might participate in MAPK/ERK and androgen receptor signaling pathways. The formula for the prognostic risk scores used in this study was as follows: $-0.1652*Exp_{(SOCS2)}$ - $0.1284*Exp_{(MAGEA6)} - 0.1992*Exp_{(RDH16)} + 0.1897*Exp_{(RTN3)}$. To study the relationships between gene biomarkers and survival outcome, we used Kaplan—Meier survival analysis. The lower thirds of the sorted mRNA expression values were defined as low expression, and the upper thirds were set as the cutoff point for high expression. Thus, all patients were divided into three groups based on the expression value. The differences among the survival curves of the low, middle, and high *MAGEA6* (Fig. 1D, a), *SOCS2* (Fig. 1D, b), and *RTN3* (Fig. 1D, c) expression groups were statistically significant. We demonstrated that *SOCS2* and *RDH16* expressions were significantly downregulated in HCC, and *MAGEA6* and *RTN3* were significantly upregulated as compared with normal liver tissues. High expression of *RDH16* in HCC cells might suppress cell growth, clonogenicity, and cell motility, which can be associated with increased level of retinoic acid, which has been widely evidenced to inhibit tumor development and progression. RTN3 facilitates p53 Ser392 phosphorylation via Chk2 (RTN3 adsorbs Chk2 to the endoplasmic reticulum and enhances its activation) and can promote subsequent nuclear translocation of p53. Thus, RTN3 can restrain HCC growth and induce apoptosis by activating p53, hence it could be a potential prognosis biomarker for HCC.

In summary, robust DEGs were identified using our integrative two-platform meta-analysis method, which may play an important role in the HCC biological process. In addition, we supply two gene-based models involved in the diagnosis and prognosis, which may provide new insights into clinical treatment decision-making. However, it appears necessary to emphasize that this approach is still being carefully reworked for future research and that additional experimental trials are required to label these biomarkers.

Funding

We appreciate the support from Kerman University of Medical Sciences, Kerman, Iran (No. 98000927). This study was approved by the Ethics Committee of Kerman University of Medical Sciences (Iran) (No. IR.KMU.REC.1399.116).

Conflict of interests

The authors declare no potential conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2022.07.018.

References

1. Mak D, Kramvis A. Epidemiology and aetiology of hepatocellular carcinoma in Sub-Saharan Africa. *Hepatoma Res*. 2021;7:39.
2. Li H, Wei N, Ma Y, et al. Integrative module analysis of HCC gene expression landscapes. *Exp Ther Med*. 2020;19(3):1779—1788.
3. Irigoyen A, Jimenez-Luna C, Benavides M, et al. Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers. *PLoS One*. 2018;13(4):e0194844.
4. Chia TS, Wong KF, Luk JM. Molecular diagnosis of hepatocellular carcinoma: trends in biomarkers combination to enhance early cancer detection. *Hepatoma Res*. 2019;5:9.
5. Zhu P, Feng R, Lu X, et al. Diagnostic and prognostic values of AKR1C3 and AKR1D1 in hepatocellular carcinoma. *Aging (Albany NY)*. 2021;13(3):4138—4156.

Maryam Gholizadeh [a], Morteza Hadizadeh [b], Seyed Reza Mazlooman [c], Saeid Eslami [a], Sadegh Raoufi [d], Marziye Farsimadan [e], Maryam Rashidifar [f], Marek Drozdzik [g,*], Mehrnaz Mehrabani [b,**]

[a] *Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad 91388-13944, Iran*
[b] *Physiology Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman 7616913555, Iran*
[c] *Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran 1477893780, Iran*
[d] *Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman 7616913555, Iran*
[e] *Department of Biology, Faculty of Sciences, University of Guilan, Rasht 13111-41937, Iran*
[f] *Department of Plant Biotechnology, Faculty of Life Science and Biotechnology, Shahid Beheshti University, Tehran 198396-3113, Iran*
[g] *Department of Experimental and Clinical Pharmacology, Pomeranian Medical University, Szczecin 70-111, Poland*

*Corresponding author.

**Corresponding author. Physiology Research Center, Institute of Basic and Clinical Physiology Sciences, Kerman University of Medical Sciences, Kerman 7616913555, Iran.
E-mail addresses: marek.drozdzik@pum.edu.pl (M. Drozdzik), mehrnaz.mehrabani@yahoo.com (M. Mehrabani)

9 April 2022
Available online 13 August 2022