



OPEN

Novel miRNA signature for predicting the stage of hepatocellular carcinoma

Srinivasulu Yerukala Sathipati^{1,2} & Shinn-Ying Ho^{1,3,4}✉

Hepatocellular carcinoma (HCC) is one of the leading causes of cancer deaths worldwide. Recently, microRNAs (miRNAs) are reported to be altered and act as potential biomarkers in various cancers. However, miRNA biomarkers for predicting the stage of HCC are limitedly discovered. Hence, we sought to identify a novel miRNA signature associated with cancer stage in HCC. We proposed a support vector machine (SVM)-based cancer stage prediction method, SVM-HCC, which uses an inheritable bi-objective combinatorial genetic algorithm for selecting a minimal set of miRNA biomarkers while maximizing the accuracy of predicting the early and advanced stages of HCC. SVM-HCC identified a 23-miRNA signature that is associated with cancer stages in patients with HCC and achieved a 10-fold cross-validation accuracy, sensitivity, specificity, Matthews correlation coefficient, and area under the receiver operating characteristic curve (AUC) of 92.59%, 0.98, 0.74, 0.80, and 0.86, respectively; and test accuracy and test AUC of 74.28% and 0.73, respectively. We prioritized the miRNAs in the signature based on their contributions to predictive performance, and validated the prognostic power of the prioritized miRNAs using Kaplan–Meier survival curves. The results showed that seven miRNAs were significantly associated with prognosis in HCC patients. Correlation analysis of the miRNA signature and its co-expressed miRNAs revealed that hsa-let-7i and its 13 co-expressed miRNAs are significantly involved in the hepatitis B pathway. In clinical practice, a prediction model using the identified 23-miRNA signature could be valuable for early-stage detection, and could also help to develop miRNA-based therapeutic strategies for HCC.

Hepatocellular carcinoma (HCC) is the most common type of liver cancer, and the third leading cause of cancer deaths worldwide¹. Hepatitis B and C viral infection^{2,3}, cirrhosis⁴, heavy alcoholism⁵, hemochromatosis⁶ and alpha-1-antitrypsin deficiency⁷ are risk factors associated with HCC. Treatment conditions depend on cancer stage, availability of treatment resources, liver function, and clinical expertise¹. Despite advances in treatment conditions, the overall survival rate of patients with HCC has not improved⁸, largely because most cases of HCC are diagnosed at an advanced stage⁹. Early-stage detection of HCC creates opportunities to use a wider range of treatment options¹⁰. Hence, early-stage detection of HCC plays a critical role in guiding treatment decisions. Identification of biomarkers for early-stage detection will help to improve treatment strategies and ensure that more patients receive the proper treatment.

Recently, microRNAs (miRNAs) have attracted interest as biomarkers due to their critical roles in cancer development and prognosis. MiRNA dysregulation is observed in multiple types of cancers: in lung cancer, hsa-let-7a expression is associated with poor survival¹¹. MiRNAs have been used as biomarkers in HCC¹². Hsa-miR-155 is highly expressed in breast and colon cancers¹³. Differential expression of miRNAs has been observed in ovarian cancer¹⁴. In gastric cancer, a seven-miRNA signature has been used to predict overall survival and relapse-free survival¹⁵.

Several studies have reported dysregulation of miRNAs in HCC. For example, miR-222 deregulation is observed in HCC cell lines¹⁶. In humans, miR-224 targets glycine N-methyl transferase and plays an important role in HCC tumorigenesis¹⁷. A 20-miRNA signature is associated with survival in HCC patients¹⁸. Moreover, some miRNAs are thought to have therapeutic potential in HCC¹⁹. Previously, Toffannin et al., used miRNA

¹Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan. ²Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan. ³Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan. ⁴Center For Intelligent Drug Systems and Smart Bio-devices (IDS2B), National Chiao Tung University, Hsinchu, Taiwan. ✉email: syho@mail.nctu.edu.tw

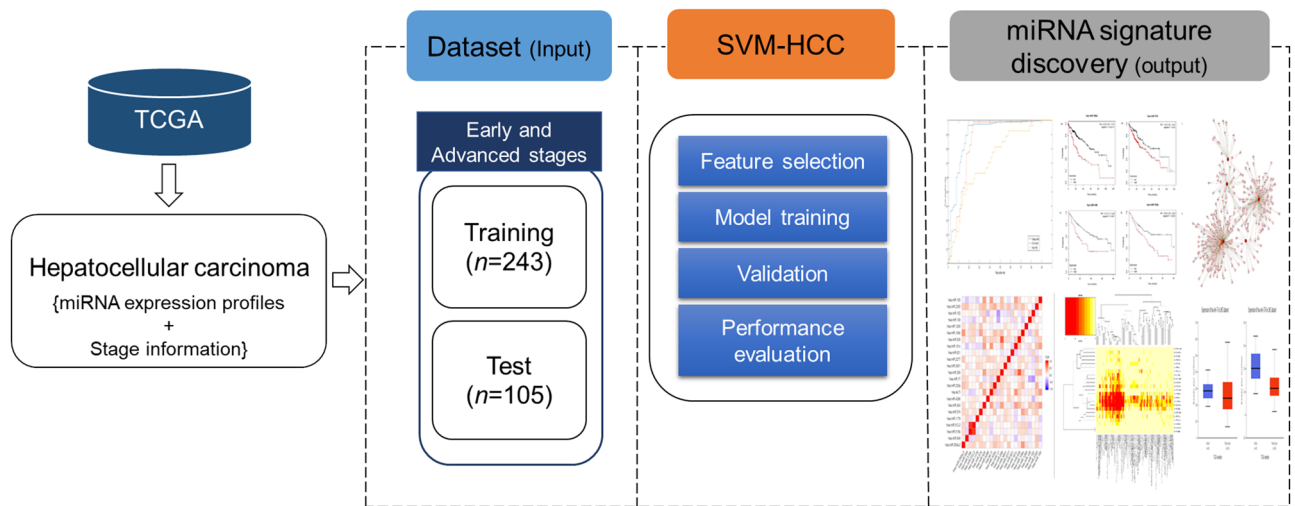


Figure 1. System flowchart representing the dataset, SVM-HCC method and miRNA signature identification.

profiling of 89 HCC patients, followed by unsupervised hierarchical clustering, to categorize HCC into three subclasses²⁰. Machine learning models have been used to predict the treatment response of trans-arterial chemoembolization in patients with HCC²¹. The random forest method and multiple urine DNA biomarkers have been used for HCC screening²². A. Nagy et al. identified 223 miRNAs as prognostic biomarkers based on previous literature and validated their prognostic power using the independent datasets; in which, 55 individual miRNAs are significantly associated with the overall survival of HCC²³. Previously developed methods and studies have mainly focused on identifying differentially expressed genes and survival variants in HCC. Early stage detection and diagnosis of cancer remains a challenge for clinicians. MiRNAs are considered as potential tumor markers due to their tissue specificity and capability to predict clinicopathological parameters²⁴. Several studies have been demonstrated that miRNAs have the potential to be new biomarkers in various cancers for early detection^{25–28}. Moreover, miRNAs can be detectable not only from tissue samples but also from a wide range of biological samples, such as urine, blood plasma, and serum. However, few studies have attempted to predict the stage of HCC using the genomic profiling. Therefore, this study aims to identify a miRNA signature consisting of a small set of miRNA biomarkers that can predict the cancer stage of patients with HCC, so that this miRNA signature can be useful for developing gene-based target therapies in HCC.

In this study, we proposed a method for predicting the early and advanced stages of HCC using miRNA expression profiles. We retrieved 348 expression profiles of 540 miRNAs (348*540) from 348 HCC patients from The Cancer Genome Atlas (TCGA) database. Our dataset includes 258 patients with early-stage disease and 90 patients with advanced HCC. We utilized a support vector machine (SVM)-based classifier²⁹, SVM-HCC, which incorporated with an inheritable bi-objective combinatorial genetic algorithm (IBCGA)³⁰ to identify a miRNA signature capable of distinguishing early-stage patients from advanced-stage HCC. Though optimization technique of the SVM-HCC was adopted from our previous study³¹, identified miRNA signature is novel in HCC stage prediction. The main purpose of this study is to identify a miRNA signature associated with cancer stage of patients with HCC. We ranked the miRNAs in the signature based on their contributions to predictive performance, and subjected the 10 top-ranked miRNAs to further analysis. Next, to investigate the prognostic power of the identified miRNA signature among the patients with HCC, Kaplan–Meier (KM) survival analysis was performed. The expression difference of the 10 top-ranked miRNAs was compared between cancer and normal samples. The biological significance of the identified miRNA signature was analyzed using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) annotations. Finally, we identified co-expressed miRNAs to the miRNA signature to provide the more information on its overall impact on HCC.

Results and discussion

The proposed method, SVM-HCC, distinguished patients with HCC into early-stage and advanced-stage groups based on their miRNA expression profiles. We used a dataset containing 540 miRNA expression profiles from 348 HCC patients, of whom 248 had early-stage and 90 had advanced-stage HCC. SVM-HCC, used a feature selection algorithm (IBCGA) to select a significant miRNA signature associated with early and advanced stages of HCC. The system flowchart of the overall process is depicted in Fig. 1.

We compared SVM-HCC with standard machine learning methods including sequential minimal optimization (SMO), multilayer perceptron (MLP), naïve Bayes, LibSVM, and random forest. For the feature selection, we used the Ranker search and correlation attribute evaluation method of Waikato Environment for Knowledge Analysis (Weka) to select 30 features to classify early-stage and advanced-stage HCC patients. SVM-HCC performed well relative to these machine learning methods in terms of training accuracy. Using the training set ($n = 348$), SVM-HCC achieved a mean training accuracy, sensitivity, specificity, and Matthews correlation coefficient (MCC) of $89.56 \pm 1.27\%$, 0.94 ± 0.01 , 0.73 ± 0.03 , and 0.71 ± 0.03 , respectively. SVM-HCC achieved the best training accuracy, sensitivity, specificity, MCC, and area under the receiver operating characteristic curve (AUC) of 92.24, 0.96, 0.81, 0.79, and 0.90, respectively. The comparison results are shown in Table 1.

Method	Training accuracy (%)	Selected miRNAs	Sensitivity	Specificity	MCC	AUC
SMO	72.70	30	0.96	0.05	0.03	0.50
MLP	66.09	30	0.79	0.27	0.07	0.60
Naïve Bayes	73.27	30	0.82	0.46	0.3	0.71
LibSVM	74.71	30	0.99	0.11	0.15	0.54
Random forest	75.28	30	0.94	0.21	0.22	0.72
SVM-HCC	92.24	37	0.96	0.81	0.79	0.90
SVM-HCC-Mean	89.56 ± 1.27	35 ± 3.9	0.94 ± 0.01	0.73 ± 0.03	0.71 ± 0.03	0.86 ± 0.02

Table 1. Comparison of predictive performance of SVM-HCC with those of other machine learning methods.

Next, to observe the difference in the prediction performance among the standard machine learning methods with the feature size, we used the greedy stepwise search and Cfs Subset Evaluator attribute evaluation method of Weka to select 19 features to classify early and advanced stages of patients with HCC. The prediction performance results shown that only a slight difference in the prediction accuracies was observed for the SMO, MLP, naïve Bayes, LibSVM, and random forest methods when compared to the Table 1. SMO, MLP, naïve Bayes, LibSVM, and random forest methods shown the accuracy differences of 1.15%, 1.43%, 0.86%, 2.87%, and 1.15%, respectively. However, there was no larger difference observed for the AUCs among these methods, shown in Supplementary Table S1.

The comparison of AUCs. Further, statistical analysis was performed to compare the prediction ability of SVM-HCC with some machine learning methods using the AUC comparison method proposed by Hanley and McNeil³². This analysis provides the statistical test comparison between the AUC of SVM-HCC and the AUCs of other machine learning methods. When compared the statistical test AUC of SVM-HCC (AUC = 0.9), SMO obtained a standard error (SE), AUC area difference (AUCd), Z value, and a *p* value of 0.03, 0.40, 10.29, and *p* < 0.001, respectively; MLP obtained SE, AUCd, Z value, and a *p* value of 0.033, 0.3, 8.09, and *p* < 0.001, respectively; naïve Bayes obtained SE, AUCd, Z value, and a *p* value of 0.029, 0.19, 5.71, and *p* < 0.001, respectively; LibSVM obtained SE, AUCd, Z value, and a *p* value of 0.034, 0.36, 9.34, and *p* < 0.001, respectively; and random forest obtained SE, AUCd, Z value, and a *p* value of 0.028, 0.18, 5.48, and *p* < 0.001, respectively. The statistical analysis shows that SVM-HCC method is significantly (*p* < 0.001) different and performed better when compared with the other standard machine learning methods, shown in Supplementary Table S2.

We performed 30 independent runs of SVM-HCC to select a robust miRNA signature using the appearance score³³, which was calculated based on the frequency of each feature over the independent runs. The most robust miRNA signature had an appearance score of 6.17. SVM-HCC identified a 23-miRNA signature associated with the early and advanced stages of HCC, and achieved a tenfold cross-validation (10-CV) accuracy, sensitivity, specificity, MCC and AUC of 92.59%, 0.98, 0.74, 0.80, and 0.86, respectively, and a test accuracy and test AUC of 74.28% and 0.73, respectively. The predictive performance of SVM-HCC was evaluated using a receiver operating characteristic (ROC) curve, and is shown in Fig. 2. Additionally, to investigate the effect of clinical characteristics on the prediction performance, we added some of the clinical characteristics of patients with HCC such as gender, risk factors, race, hepatitis serology, and vital status to the miRNA signature for the stage prediction. However, addition of these clinical features did not improve the prediction performance of SVM-HCC.

Ranking of miRNA signature. We ranked the miRNA signature identified by SVM-HCC by main effect difference (MED) analysis³⁴. The larger MED score indicates the higher contribution towards the prediction accuracy. In this analysis, miRNAs in the signature were ranked based on their MED scores. The miRNA signature contains 23 miRNAs: in order of decreasing MED scores, hsa-miR-550a, hsa-miR-549, hsa-miR-518b, hsa-miR-512, hsa-miR-1179, hsa-miR-574, hsa-miR-424, hsa-miR-4286, hsa-let-7i, hsa-miR-320a, hsa-miR-17, hsa-miR-299, hsa-miR-3651, hsa-miR-2277, hsa-miR-621, hsa-miR-181c, hsa-miR-539, hsa-miR-106b, hsa-miR-1269, hsa-miR-139, hsa-miR-152, hsa-miR-2355, and hsa-miR-150. Rankings of miRNA signatures and MED scores are provided in Table 2. According to the MED analysis, the 10 top-ranked miRNAs contributed better towards prediction performance when compared to the remaining miRNAs of the signature. Hence, we subjected the 10 top-ranked miRNAs to further analysis.

Validation of top-ranked miRNA prognostics in HCC. We validated the prognostic power of the 10 top-ranked miRNAs in HCC using Kaplan–Meier (KM) survival curves generated by the KM plotter³⁵. We selected overall survival data available for 376 patients from TCGA and 166 patients from the GSE31384 dataset, while the median follow-up time of patients in TCGA and GSE31384 was 19.6 and 34 months, respectively. Four of the ten top-ranked miRNAs were significantly associated with overall survival of HCC patients in the TCGA dataset. These four miRNAs, hsa-miR-550a, hsa-miR-574, hsa-miR-424, and hsa-let-7i, had *p* values of 4.9e-07, 0.016, 0.024, and 0.032 and hazard ratios of 2.38, 1.64, 1.57, and 1.49, respectively. Three of the top-ranked miRNAs were significantly associated with overall survival of HCC patients in the GSE31384 dataset, hsa-miR-549, hsa-miR-518, and hsa-miR-512, with *p* values of 0.0021, 0.0022, and 0.0021 and hazard ratios of 2.12, 2.03, and 0.48, respectively. The KM survival curves for the seven miRNAs are shown in Fig. 3. The KM survival curves

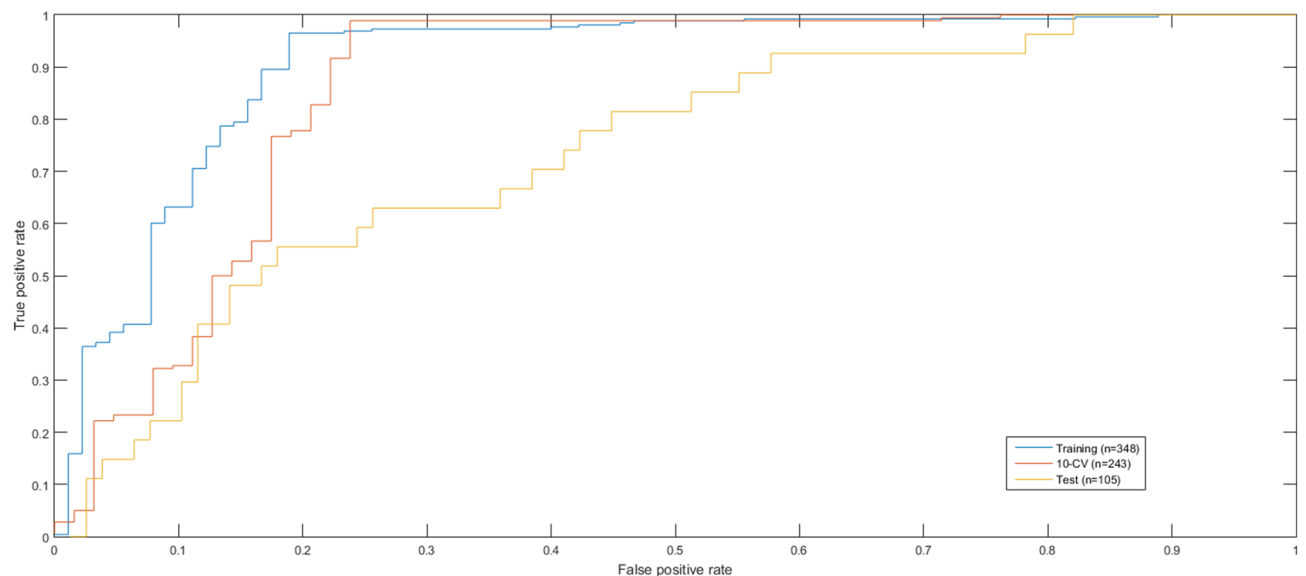


Figure 2. ROC curves for evaluating the predictive performance of SVM-HCC. The ROC curve for Training dataset (AUC=0.90 using 348-patient HCC cohort), 10-CV (AUC=0.86 using 243-patient HCC cohort), and test dataset (AUC=0.73 using 105-patient HCC cohort).

Rank	MiRNA	MED score
1	hsa-miR-550a	60.91
2	hsa-miR-549	54.72
3	hsa-miR-518b	51.15
4	hsa-miR-512	50.67
5	hsa-miR-1179	27.73
6	hsa-miR-574	27.02
7	hsa-miR-424	26.62
8	hsa-miR-4286	24.72
9	hsa-let-7i	24.16
10	hsa-miR-320a	22.97
11	hsa-miR-17	22.81
12	hsa-miR-299	22.02
13	hsa-miR-3651	17.02
14	hsa-miR-2277	13.76
15	hsa-miR-621	13.61
16	hsa-miR-181c	13.05
17	hsa-miR-539	12.97
18	hsa-miR-106b	10.83
19	hsa-miR-1269	8.53
20	hsa-miR-139	6.70
21	hsa-miR-152	6.62
22	hsa-miR-2355	3.76
23	hsa-miR-150	2.10

Table 2. Ranking of miRNAs and their corresponding scores, using MED analysis.

for the remaining three miRNAs, hsa-miR-1179, hsa-miR-4286, and hsa-miR-320a are shown in Supplementary Fig. S1.

Furthermore, we predicted the prognosis of patients with HCC using remaining 13 miRNAs of the signature across TCGA and GSE31384 datasets. Seven of these 13 miRNAs were significantly associated with the prognosis of patients with HCC across TCGA dataset. Totally, there were 11 of the 23-miRNA signature including, hsa-miR-550a ($p < 0.001$), hsa-miR-574 ($p = 0.016$), hsa-let-7i ($p = 0.03$), hsa-miR-424 ($p = 0.024$), hsa-miR-2277 ($p = 0.017$), hsa-miR-539 ($p = 0.05$), hsa-miR-106b ($p = 0.011$), hsa-miR-1269a ($p = 0.001$), hsa-miR-139

($p < 0.001$), hsa-miR-2355 ($p = 0.03$), and hsa-miR-150 ($p = 0.03$) were significantly associated with prognosis of patients with HCC across TCGA dataset.

Among the 23-miRNA signature, 10 miRNAs including, hsa-miR-549 ($p = 0.002$), hsa-miR-518b ($p = 0.002$), hsa-miR-512 ($p = 0.011$), hsa-miR-574 ($p = 0.002$), hsa-miR-299 ($p < 0.001$), hsa-miR-181c ($p = 0.003$), hsa-miR-539 ($p = 0.001$), hsa-miR-106b ($p < 0.001$), hsa-miR-152 ($p < 0.001$), and hsa-miR-150 ($p = 0.031$) were significantly associated with prognosis of patients with HCC across GSE31384 dataset.

Expression difference of top ranked miRNAs in tumor vs normal. We then compared the expression levels of the 10 top-ranked miRNAs in tumor and normal samples using UALCAN web portal³⁶, and observed a significant difference in miRNA expression levels between the two groups. Of the top 10 ranked miRNAs, eight miRNAs, hsa-miR-550a, hsa-miR-518b, hsa-miR-512, hsa-miR-574, hsa-miR-424, hsa-miR-4286, hsa-let-7i, hsa-miR-320a are significantly expressed in tumor and normal samples, a p value < 0.05 was considered a threshold to describe the statistical significant. Among, two of the ten top-ranked miRNAs (hsa-miR-549 and hsa-miR-1179), the role of hsa-miR-549 was not reported earlier in HCC, and hsa-miR-1179 was not significantly expressed between the tumor and adjacent normal tissues of the TCGA cohort. However, hsa-miR-549 contributed better towards predicting the stage of HCC (MED rank 2) and possessed a significant role in other cancers^{37–39}. Though, the expression of hsa-miR-1179 was not significant between the tumor and adjacent normal tissues of the TCGA cohort, a Quantitative Real Time-Polymerase Chain Reaction study on 40 HCC samples reported that hsa-miR-1179 was significantly expressed between HCC and matched normal tissues, and plays an important role in HCC progression and metastasis⁴⁰. The expression levels of the 10 top-ranked miRNAs in the tumor and normal samples are listed in Supplementary Table S3. Box plot representation of relative expression difference of top ranked miRNAs in tumor and normal samples is given in Supplementary Fig. S2. The individual data points of the expression analysis can be accessed from the UALCAN web portal.

Further, we attempted to distinguish the tumor and normal samples using the identified miRNA signature. We used a dataset consisting of 32 normal samples and randomly selected 32 tumor samples, and LibSVM of the WEKA to distinguish the tumor and normal samples. LibSVM achieved a leave-one-out accuracy of 100% to distinguish tumor and normal samples using the 23-miRNA signature.

Significance of top-ranked miRNAs in cancer. Nine of the ten top-ranked miRNAs are involved in HCC and various other cancers; we summarize their functions in HCC, based on reports in the experimentally validated literature, in Supplementary Table S4.

Hsa-miR-549, the second-ranked miRNA, is differentially expressed in cancer cells relative to normal cells. For example, hsa-miR-549 is highly expressed in colon cancer³⁷, colorectal cancer³⁸, and breast cancers³⁹, with log-fold changes of 0.51, 1.75, and 0.66, respectively, relative to normal cells. However, the role of hsa-miR-549 in HCC has not been reported previously. Our results suggest that hsa-miR-549 is significantly associated with overall survival in HCC patients, and that it actively participates in other major cancers. Hence, it is a worthy subject of further investigation.

Other than the 10 top-ranked miRNAs, several of the remaining miRNAs in the signature are actively involved in HCC and other cancers. For example, expression of hsa-miR-11 is associated with poor prognosis in HCC patients⁴¹, whereas hsa-miR-299 acts as a tumor suppressor in HCC⁴². Hsa-miR-3651, hsa-miR-621, hsa-miR-181c, hsa-miR-539, hsa-miR-106b, hsa-miR-1269, hsa-miR-139, hsa-miR-152, and hsa-miR-150 are all significantly involved in HCC^{43–50}.

We constructed a miRNA target interaction network using Cytoscape⁵¹ to investigate regulatory interactions compiled in the miRTarBase database. The top-ranked miRNAs annotated with miRBase accession numbers and predicted miRNA interactions using miRTarBase was 2,274. The predicted miRNA target interaction network is shown in Supplementary Fig. S3.

KEGG pathway and gene ontology enrichment analysis. Next, we investigated the biological significance of the top-ranked miRNAs using KEGG pathway and GO annotation analysis. First, we used the DIANA-miRPath web tool⁵² to examine their functional annotations. Fisher's exact test was used for the enrichment analysis. The 10 top-ranked miRNAs are involved in several pathways, the most significant of which are fatty acid metabolism, fatty acid biosynthesis, fatty acid elongation, endocytosis, fatty acid degradation, pathways in cancer, lysine degradation, viral carcinogenesis, glioma, and the Hippo signaling pathway. The top-ranked miRNAs, along with the numbers of predicted target genes in each pathway, are listed in Table 3. The heatmap of the 10 top-ranked miRNAs enriched in KEGG pathways is shown in Fig. 4(A) and the number of target genes involved in pathways is shown in Fig. 4(B). The 23-miRNA signature enriched in KEGG pathways is shown in Supplementary Fig. S4.

Second, we analyzed the involvement of the top-ranked miRNAs in biological pathways, molecular functions, and cellular components using GO annotations. We found that these miRNAs are significantly involved in biological pathways including the mitotic cell cycle, blood coagulation, cellular protein metabolic process, membrane organization, epidermal growth factor receptor signaling pathway, and cell death, with p values $< 1.11E-16$. They are also involved in molecular functions including protein binding transcription factor activity, nucleic acid binding transcription factor activity, ion binding, and RNA binding. Finally, they are involved in cellular components including cytosol, protein complex, neoplasms, and organelles. Details of these associations are given in Supplementary Table S5, and the enrichment of the 23-miRNA signature in GO annotations is shown in Supplementary Fig. S5.

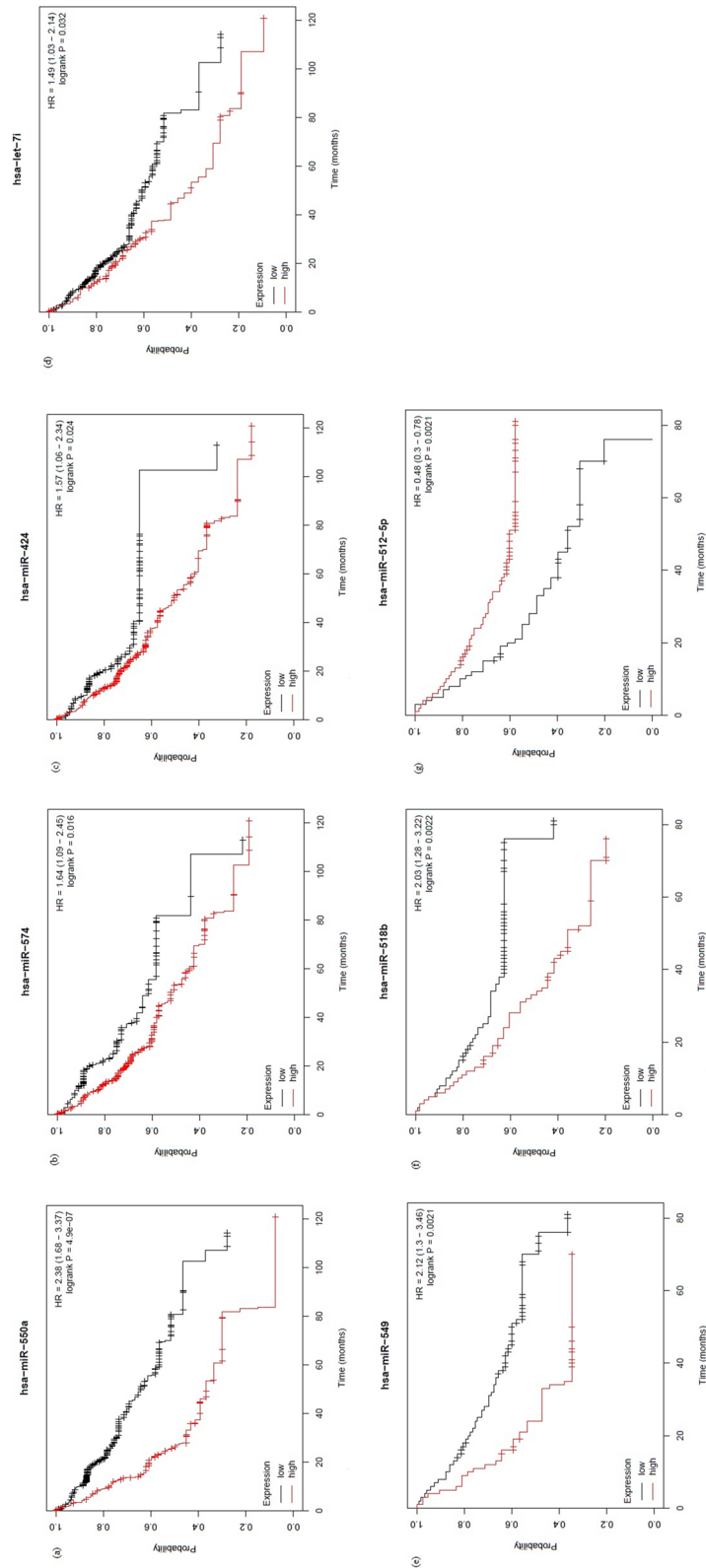


Figure 3. Kaplan-Meier plots of (a) hsa-miR-550a, (b) hsa-miR-574, (c) hsa-miR-424, and (d) hsa-let-7i using TCGA dataset, and (e) hsa-miR-549, (f) hsa-miR-518, and (g) hsa-miR-512 using GSE31384 dataset for the high-expression and low-expression groups of the HCC cohort.

KEGG pathway	Genes	MiRNAs	Adjusted <i>p</i> value
Fatty acid metabolism	10	3	<0.001
Fatty acid biosynthesis	4	2	<0.001
Fatty acid elongation	1	1	<0.001
Endocytosis	45	3	<0.001
Fatty acid degradation	1	1	<0.001
Pathways in cancer	141	4	<0.001
Lysine degradation	19	3	<0.001
Viral carcinogenesis	55	2	<0.001
Glioma	29	4	<0.001
Hippo signaling pathway	55	3	<0.001
Chronic myeloid leukemia	33	4	<0.001
Hepatitis B	55	3	<0.001
Proteoglycans in cancer	73	2	<0.001
TGF-beta signaling pathway	29	3	<0.001
Adherens junction	33	3	<0.001
Renal cell carcinoma	24	2	<0.05
Prostate cancer	41	3	<0.05
Estrogen signaling pathway	27	3	0.05
Thyroid cancer	12	3	0.07
Cell cycle	51	2	0.06
Transcriptional misregulation in cancer	46	3	0.18
Colorectal cancer	27	3	0.12

Table 3. KEGG pathway analysis of the 10 top-ranked miRNAs.

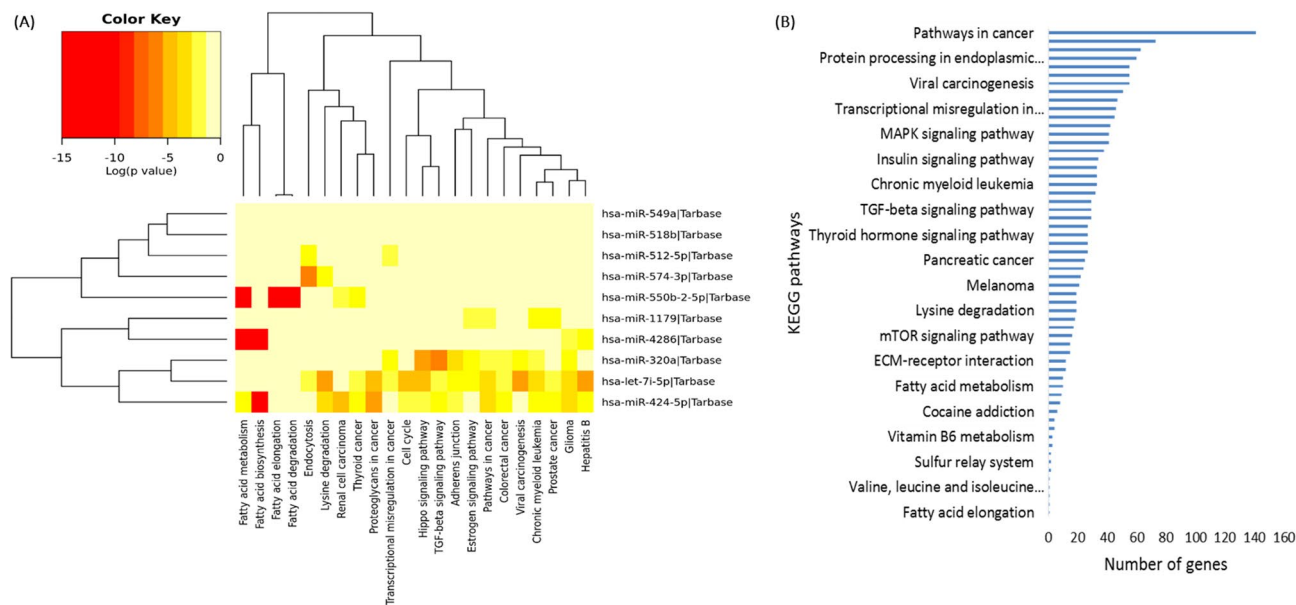


Figure 4. KEGG pathway analysis of the 10 top-ranked miRNAs. **(A)** Heatmap showing enrichment of the 10 top-ranked miRNAs in KEGG pathways. **(B)** The 10 top-ranked miRNAs involved in KEGG pathways are shown in the vertical dimension, and the number of target genes involved in pathways is shown in the horizontal dimension. The size of the pathway is proportional to the number of genes involved in that particular pathway.

MiRNAs co-expressed with the top-ranked miRNAs. Identifying more relevant miRNAs may lead to the identification of robust miRNAs that are essential for cancer. Moreover, correlated miRNAs may represent similar biological processes. In its optimization process, SVM-HCC selects a minimal set of biomarkers; hence it selected 23 biomarker miRNAs as a signature associated with HCC stage. However, it might not select some important biomarker miRNAs that are also associated with cancer stage in patients with HCC; also, the priority of miRNA selection may change with the size and number of miRNA profiles used. Hence, to select robust

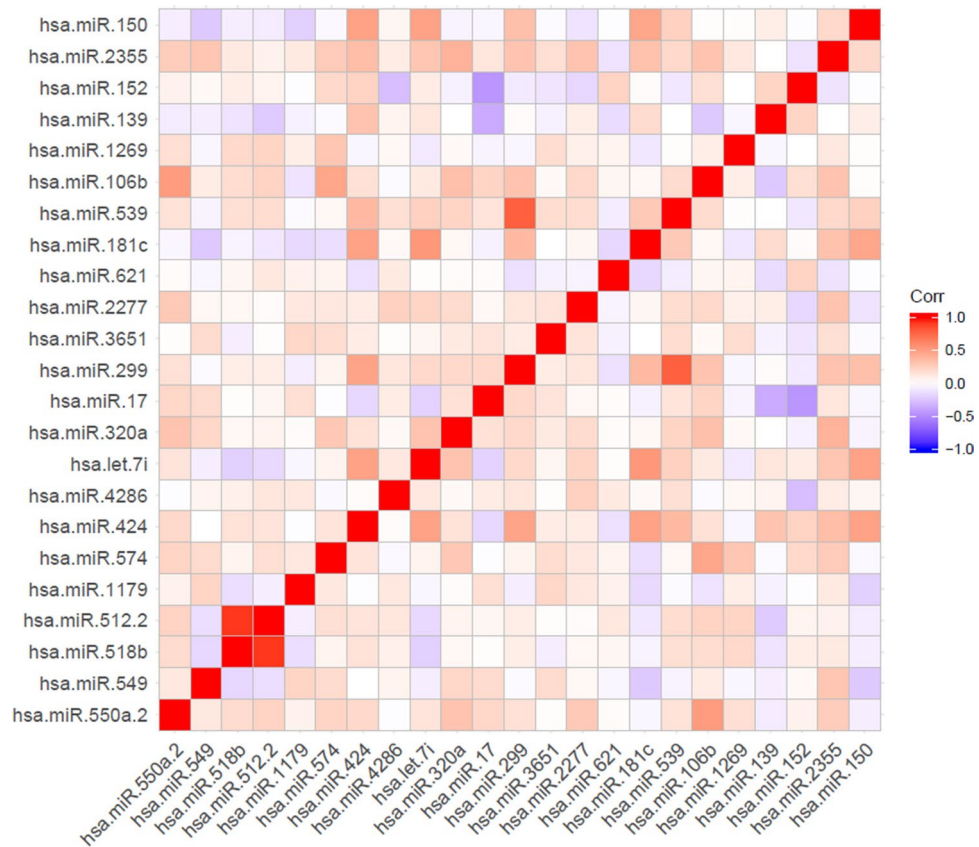


Figure 5. Heatmap showing correlation coefficients among the members of the 23-miRNA signature.

miRNAs outside the 23-miRNA signature, we sought to identify miRNAs that were co-expressed with those 23 miRNAs.

We computed the correlations among each miRNA in the signature using the Pearson correlation coefficient. The miRNAs with the highest correlation coefficient in the signature (0.92) were hsa-miR-512 and hsa-miR-518. These two miRNAs were also significantly associated with overall survival in HCC patients, with p values of 0.0022 and 0.0021. Because these miRNAs had high correlation coefficients, we considered them for further analysis. Thus, we sought to analyze the top-ranked individual miRNAs in the signature. The correlation heatmap of the miRNA signature is shown in Fig. 5.

Additionally, we sought to identify the miRNAs that were highly correlated with the miRNA signature in the 540 expression profiles constituting our dataset. To this end, we measured the correlations of the 23-miRNA signature with the 540 miRNA expression profiles. We considered miRNAs with higher correlations to be co-expressed with the signature. These co-expressed miRNAs and their correlation coefficients are listed in Supplementary Table S6. We considered $R \geq 0.5$ to be statistically significant. Five miRNAs in the top-ranked miRNA signatures had co-expressed miRNAs with correlations ≥ 0.5 . Hsa-miR-518 and hsa-miR-512 had nine co-expressed miRNAs in common. Three miRNAs, hsa-miR-424, hsa-let-7i, and hsa-miR-320a, had 15, 16, and 1 co-expressed miRNA, respectively.

Furthermore, we examined the biological significance of the top-ranked miRNAs and their co-expressed miRNAs to determine whether they were involved in any common pathways. KEGG pathway analysis of hsa-miR-518 and hsa-miR-512 and their co-expressed miRNAs included glycosphingolipid biosynthesis-lacto and neolacto series (hsa00601), folate biosynthesis (hsa00790), one-carbon pool by folate (hsa00670), mucin-type O-Glycan biosynthesis (hsa00512), and central carbon metabolism in cancer (hsa05230). Details of the involvement of hsa-miR-518, hsa-miR-512, and their co-expressed miRNAs in KEGG pathways are provided in Supplementary Table S7. Hsa-miR-424 and its co-expressed miRNAs are significantly involved in several cancer pathways, including proteoglycans in cancer (hsa05205), the Hippo signaling pathway (hsa04390), viral carcinogenesis (hsa05203), pathways in cancer (hsa05200), and glioma (hsa05214). Details of the involvement of hsa-miR-424 and its co-expressed miRNAs in KEGG pathways are provided in Supplementary Table S8. Hsa-let-7i and its co-expressed miRNAs are involved in several cancer pathways, including proteoglycans in cancer, viral carcinogenesis, pathways in cancer, chronic myeloid leukemia, thyroid cancer, bladder cancer, colorectal cancer, glioma, and prostate cancer. Interestingly, we found that hsa-let-7i and its co-expressed miRNAs, hsa-miR-145-5p, hsa-miR-10a-3p, hsa-let-7b-5p, hsa-miR-155-5p, hsa-miR-142-5p, hsa-miR-125a-3p, hsa-miR-199a-5p, hsa-miR-214-3p, hsa-miR-424-3p, hsa-miR-708-3p, hsa-miR-542-5p, hsa-miR-342-5p, and hsa-miR-450a-5p, are significantly (p value of $1.18E-10$) involved in the hepatitis B pathway (hsa05161), and target 77 genes. Chronic hepatitis B infection has

been linked to HCC⁵³. Details of the involvement of hsa-let-7i and its co-expressed miRNAs in KEGG pathways are provided in Supplementary Table S9. Experimentally validated gene interactions for hsa-let-7i and its co-expressed miRNAs in the hepatitis B pathway are shown in Supplementary Table S10.

Hsa-miR-320a had a co-expressed miRNA, hsa-miR-1301. These two miRNAs are significantly involved in cancer pathways including transcriptional misregulation in cancer, glioma, viral carcinogenesis, pathways in cancer, colorectal cancer, and pancreatic cancer. Their involvement in biological pathways is shown in detail in Supplementary Table S11. Together, these analyses revealed that not only the 23-miRNA signature, but also its co-expressed miRNAs, are involved in important pathways, and are therefore worthy of further exploration in the context of HCC. These findings could facilitate the development of miRNA-based therapeutic strategies for HCC.

MiRNAs correlated to the hepatitis infection. We measured the correlation between identified miRNA signature and clinicopathological features of HCC using Spearman correlation coefficient. Three miRNAs, hsa-let-7i, hsa-miR-320a, and hsa-miR-2355 of the signature were significantly correlated with the hepatitis infection, shown in Supplementary Table S12. Additionally, correlation was measured for hsa-let-7i and its 13 co-expressed miRNAs, which were involved in hepatitis B pathway. Two of these miRNAs, hsa-miR-145 and hsa-miR-125a were significantly correlated with the hepatitis infection, shown in Supplementary Table S13.

Conclusions

Detecting liver cancer at an early stage is difficult because its symptoms often appear only at the later stages. Currently, the diversity of miRNAs, and their differential expression in multiple types of cancer, make them worthy of investigation in the context of cancer research. Recently, miRNAs have been explored as biomarkers of various cancers. Identifying miRNA signatures associated with early-stage HCC could provide useful insight into miRNA-mediated diagnosis of this disease. Developing computational methods for early-stage detection based on miRNA expression could elucidate the variants involved in cancer progression. Besides, potential feature selection methods can easily deal with high-dimensional samples such as gene expression profiles.

In this study, we introduced a SVM-based prediction method, SVM-HCC, which incorporates an optimal feature selection algorithm (IBCGA) to identify miRNA signatures capable of distinguishing early-stage and advanced-stage patients with HCC. SVM-HCC identified a 23-miRNA signature associated with early-stage and advanced-stage HCC, and achieved a 10-CV mean accuracy, sensitivity, specificity, and MCC of 92.44 ± 0.99 , 0.96 ± 0.01 , 0.78 ± 0.03 , and 0.79 ± 0.02 , respectively. We prioritized the 23-miRNA signature based on MED scores; miRNAs with higher MED score contributed more to prediction accuracy. The highest-ranked miRNAs were subjected to further analysis.

We validated the prognostic power of the top-ranked miRNAs in HCC using KM survival curves. The results revealed that 7 of the 10 top-ranked miRNAs, hsa-miR-550a, hsa-miR-574, hsa-miR-424, hsa-let-7i, hsa-miR-549, hsa-miR-518, and hsa-miR-512, were significantly associated with overall survival in patients with HCC. In addition, the top-ranked miRNAs are all significantly involved in HCC, with the exception of hsa-miR-549. This miRNA plays an important role in other cancers, but its role in HCC had not been reported previously. However, our results suggest that hsa-miR-549 is significantly associated with overall survival in patients with HCC, and is therefore worthy of further investigation. KEGG pathway and GO enrichment analyses revealed the functional mechanisms of top-ranked miRNAs in several cancer and non-cancer pathways. Although, the identified miRNA signature is potential to predict the stage of HCC, additional information on co-expressed miRNAs to the miRNA signature was provided to explore the possible miRNAs beyond this 23-miRNA signature that might provide specific information/knowledge on its overall impact on HCC. Interestingly, we found that hsa-let-7i and its 13 co-expressed miRNAs were significantly involved in the hepatitis B pathway.

Together, our findings help to explore the role of miRNAs in HCC, and could facilitate early-stage detection and prevention.

Materials and methods

Dataset. From the TCGA database, we retrieved a dataset containing miRNA expression profiles from 348 patients with liver HCC; these profiles were obtained using the Illumina HiSeq 2000 platform. After filtering, the final dataset contained 540 miRNA expression profiles from 348 patients. The HCC stage system in the TCGA dataset was based on the size of the primary breast tumor (T), the spread of cancer to lymph nodes (N) and distant metastasis (M) according to the American Joint Committee on Cancer. For classification purposes, the dataset was divided into early-stage (stages 1 and 2) and advanced-stage (stages 3 and 4) groups. There were 258 patients in the early-stage group and 90 patients in the advanced-stage group. Clinical characteristics of the patients with HCC used in this current study is displayed in Supplementary Fig. S6.

Establishing the SVM-HCC. We proposed a method, SVM-HCC, to identify a miRNA signature capable of distinguishing early-stage and advanced-stage HCC based on miRNA expression profiles. SVM-HCC is based on an SVM²⁹ incorporating the feature selection algorithm IBCGA. SVMs are powerful statistical learning algorithms that use non-linear transformation to map data from input space to higher-dimensional space to identify better predictive models. SVMs have become popular in the biomedical sciences, especially in cancer research, due to their potential predictive performance⁵⁴.

We used miRNA expression profiles of patients with HCC as inputs. SVMs work implicitly by only computing the corresponding kernels in the feature space between two data points, x_i and x_j . The SVM kernel function is defined as

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) \quad (1)$$

where $\varphi(x)$ is the mapping function. SVM-HCC was developed using the LIBSVM package⁵⁵, where the radial basis function (RBF) is used as the kernel function for the implementation of the SVM. RBF is defined as follows:

$$K(x_i, x_j) = \exp(-\gamma x_i - x_j)^2 \quad (2)$$

In this study, the SVM parameters C and γ were optimized based on 10-CV. While establishing the SVM-HCC, an optimal feature selection algorithm, IBCGA, was incorporated into the SVM. IBCGA is an intelligent evolutionary algorithm⁵⁶ that uses an orthogonal array crossover to solve large parameter optimization problems. In the optimization process, IBCGA selects a minimum number of features, in this case miRNAs, while improving its predictive performance. We have successfully applied IBCGA to various types of cancer predictions^{31,33,57,58}. To distinguish early-stage from advanced-stage HCC, the parameter settings of SVM and IBCGA were encoded into binary “genes.” In this study, genetic algorithm (GA) terms were used to represent the genes and “chromosomes.” We used 540 miRNA ($m = 540$) expression profiles from 348 HCC patients ($n = 348$) as input. IBCGA parameters were $r_{start} = 10$, $r_{end} = 50$, $N_{pop} = 50$, and $G_{max} = 60$, as used in³¹. The steps involved in IBCGA are as follows.

Step 1: (Evaluation) Evaluate the fitness value of all individuals using the fitness function, which is the prediction accuracy in terms of 10-CV.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Step 2: (Selection) Use a tournament selection method that selects the winner from two randomly selected individuals to generate a mating pool.

Step 3: (Crossover) Select two parents from the mating pool to perform an orthogonal array crossover operation.

Step 4: (Mutation) Apply a conventional mutation operator to randomly selected individuals in the new population. To prevent the highest fitness value from deteriorating, mutation is not applied to the best individuals.

Step 5: (Termination test) If the stopping condition for obtaining the solution is satisfied, then output the best individual as the solution. Otherwise, go to Step 2.

Step 6: (Inheritance) If $r < r_{end}$, randomly change one bit in the binary GA genes for each individual from 0 to 1; increase the number r by one, and go to Step 2. Otherwise, stop the algorithm.

Weka classifier. We used Weka⁵⁹, a powerful data mining tool that uses well-known machine learning algorithms. We compared the predictive performance of SVM-HCC with those of some machine learning methods such as SMO, MLP, naïve Bayes, LIBSVM, and random forest. We performed 10-CV to evaluate the performance of the machine learning models.

Evaluation metrics. We evaluated the predictive performance of the classifier using the following evaluation metrics: sensitivity (SN), specificity (SP), Matthews correlation coefficient (MCC), accuracy (ACC), and area under the ROC curve (AUC).

$$SN = \frac{TP}{TP + FN} \quad (4)$$

$$SP = \frac{TN}{TN + FP} \quad (5)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Data availability

All data analyzed during this study are publicly available at TCGA data portal (<https://portal.gdc.cancer.gov/>).

Received: 19 March 2020; Accepted: 13 August 2020

Published online: 02 September 2020

References

1. El-Serag, H. B. Hepatocellular carcinoma. *N Engl J Med* **365**, 1118–1127. <https://doi.org/10.1056/NEJMra1001683> (2011).
2. de Martel, C. *et al.* Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol* **13**, 607–615. [https://doi.org/10.1016/s1470-2045\(12\)70137-7](https://doi.org/10.1016/s1470-2045(12)70137-7) (2012).
3. Beasley, R. P. Hepatitis B virus. The major etiology of hepatocellular carcinoma. *Cancer* **61**, 1942–1956 (1988).

4. Forner, A., Llovet, J. M. & Bruix, J. Hepatocellular carcinoma. *Lancet* **379**, 1245–1255. [https://doi.org/10.1016/s0140-6736\(11\)61347-0](https://doi.org/10.1016/s0140-6736(11)61347-0) (2012).
5. Lin, C. W. *et al.* Heavy alcohol consumption increases the incidence of hepatocellular carcinoma in hepatitis B virus-related cirrhosis. *J Hepatol* **58**, 730–735. <https://doi.org/10.1016/j.jhep.2012.11.045> (2013).
6. Crownover, B. K. & Covey, C. J. Hereditary hemochromatosis. *Am Fam Physician* **87**, 183–190 (2013).
7. Stoller, J. K., Lacobawan, F. L. & Aboussouan, L. S. in *GeneReviews*(R) (eds M. P. Adam *et al.*) (University of Washington, Seattle University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., 1993).
8. Blum, H. E. Treatment of hepatocellular carcinoma. *Best Pract Res Clin Gastroenterol* **19**, 129–145. <https://doi.org/10.1016/j.bpg.2004.11.008> (2005).
9. Chen, C. H. *et al.* Long-term trends and geographic variations in the survival of patients with hepatocellular carcinoma: analysis of 11,312 patients in Taiwan. *J Gastroenterol Hepatol* **21**, 1561–1566. <https://doi.org/10.1111/j.1440-1746.2006.04425.x> (2006).
10. Marrero, J. A. Current treatment approaches in HCC. *Clin Adv Hematol Oncol* **11**, 15–18 (2013).
11. Yanaihara, N. *et al.* Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189–198. <https://doi.org/10.1016/j.ccr.2006.01.025> (2006).
12. Mohamed, A. A. *et al.* MicroRNAs and clinical implications in hepatocellular carcinoma. *World J Hepatol* **9**, 1001–1007. <https://doi.org/10.4254/wjh.v9.i23.1001> (2017).
13. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* **103**, 2257–2261. <https://doi.org/10.1073/pnas.0510565103> (2006).
14. Iorio, M. V. *et al.* MicroRNA signatures in human ovarian cancer. *Can. Res.* **67**, 8699. <https://doi.org/10.1158/0008-5472.CAN-07-1936> (2007).
15. Li, X. *et al.* Survival prediction of gastric cancer by a seven-microRNA signature. *Gut* **59**, 579. <https://doi.org/10.1136/gut.2008.175497> (2010).
16. Wong, Q. W. L. *et al.* MicroRNA-223 is commonly repressed in hepatocellular carcinoma and potentiates expression of Stathmin1. *Gastroenterology* **135**, 257–269. <https://doi.org/10.1053/j.gastro.2008.04.003> (2008).
17. Hung, J.-H. *et al.* MicroRNA-224 down-regulates Glycine N-methyltransferase gene expression in Hepatocellular Carcinoma. *Sci Rep* **8**, 12284. <https://doi.org/10.1038/s41598-018-30682-5> (2018).
18. Wei, R. *et al.* Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma. *Clin. Cancer Res.* **19**, 4780. <https://doi.org/10.1158/1078-0432.CCR-12-2728> (2013).
19. Borel, F., Konstantinova, P. & Jansen, P. L. M. Diagnostic and therapeutic potential of miRNA signatures in patients with hepatocellular carcinoma. *J. Hepatol.* **56**, 1371–1383. <https://doi.org/10.1016/j.jhep.2011.11.026> (2012).
20. Toffanin, S. *et al.* MicroRNA-based classification of hepatocellular carcinoma and oncogenic role of miR-517a. *Gastroenterology* **140**, 1618–1628.e1616. <https://doi.org/10.1053/j.gastro.2011.02.009> (2011).
21. Abajian, A. *et al.* Predicting treatment response to intra-arterial therapies for hepatocellular carcinoma with the use of supervised machine learning—an artificial intelligence concept. *J Vasc Interv Radiol* **29**, 850–857.e851. <https://doi.org/10.1016/j.jvir.2018.01.769> (2018).
22. Wang, J. *et al.* Development and evaluation of novel statistical methods in urine biomarker-based hepatocellular carcinoma screening. *Sci Rep* **8**, 3799. <https://doi.org/10.1038/s41598-018-21922-9> (2018).
23. Nagy, Á., Lánckzy, A., Menyhárt, O. & Györfy, B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep* **8**, 1–9 (2018).
24. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838. <https://doi.org/10.1038/nature03702> (2005).
25. Chen, X. *et al.* Identification of ten serum microRNAs from a genome-wide serum microRNA expression profile as novel noninvasive biomarkers for nonsmall cell lung cancer diagnosis. *Int. J. Cancer* **130**, 1620–1628. <https://doi.org/10.1002/ijc.26177> (2012).
26. Zhu, C. *et al.* A five-microRNA panel in plasma was identified as potential biomarker for early detection of gastric cancer. *Br. J. Cancer* **110**, 2291–2299. <https://doi.org/10.1038/bjc.2014.119> (2014).
27. Wang, L. G. & Gu, J. Serum microRNA-29a is a promising novel marker for early detection of colorectal liver metastasis. *Cancer Epidemiol.* **36**, e61–67. <https://doi.org/10.1016/j.canep.2011.05.002> (2012).
28. Kahraman, M. *et al.* MicroRNA in diagnosis and therapy monitoring of early-stage triple-negative breast cancer. *Sci Rep* **8**, 11584. <https://doi.org/10.1038/s41598-018-29917-2> (2018).
29. Vapnik, V. N. An overview of statistical learning theory. *IEEE Trans. Neural Netw* **10**, 988–999. <https://doi.org/10.1109/72.788640> (1999).
30. Shinn-Ying, H., Jian-Hung, C. & Meng-Hsun, H. Inheritable genetic algorithm for biobjective 0/1 combinatorial optimization problems and its applications. *IEEE Trans Syst. Man Cybern. Part B Cybern.* **34**, 609–620. <https://doi.org/10.1109/TSMCB.2003.817090> (2004).
31. Yerukala Sathipati, S. & Ho, S.-Y. Identifying a miRNA signature for predicting the stage of breast cancer. *Sci. Rep.* **8**, 16138. <https://doi.org/10.1038/s41598-018-34604-3> (2018).
32. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36. <https://doi.org/10.1148/radiology.143.1.7063747> (1982).
33. Yerukala Sathipati, S. & Ho, S.-Y. Identifying the miRNA signature associated with survival time in patients with lung adenocarcinoma using miRNA expression profiles. *Sci. Rep.* **7**, 7507. <https://doi.org/10.1038/s41598-017-07739-y> (2017).
34. Tung, C. W. & Ho, S. Y. Computational identification of ubiquitylation sites from protein sequences. *BMC Bioinform.* **9**, 310. <https://doi.org/10.1186/1471-2105-9-310> (2008).
35. Nagy, Á., Lánckzy, A., Menyhárt, O. & Györfy, B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* **8**, 9227. <https://doi.org/10.1038/s41598-018-27521-y> (2018).
36. Chandrashekar, D. S. *et al.* UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia (New York, NY)* **19**, 649–658. <https://doi.org/10.1016/j.neo.2017.05.002> (2017).
37. Sarver, A. L. *et al.* Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states. *BMC Cancer* **9**, 401. <https://doi.org/10.1186/1471-2407-9-401> (2009).
38. Balaguer, F. *et al.* Colorectal cancers with microsatellite instability display unique miRNA profiles. *Clin. Cancer Res.* **17**, 6239–6249. <https://doi.org/10.1158/1078-0432.ccr-11-1424> (2011).
39. Lee, C. H. *et al.* MicroRNA-regulated protein-protein interaction networks and their functions in breast cancer. *Int. J. Mol. Sci.* **14**, 11560–11606. <https://doi.org/10.3390/ijms140611560> (2013).
40. Gao, H. B., Gao, F. Z. & Chen, X. F. MiRNA-1179 suppresses the metastasis of hepatocellular carcinoma by interacting with ZEB2. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 5149–5157. https://doi.org/10.26355/eurrev_201906_18179 (2019).
41. Zheng, J., Dong, P., Gao, S., Wang, N. & Yu, F. High expression of serum miR-17-5p associated with poor prognosis in patients with hepatocellular carcinoma. *Hepatogastroenterology* **60**, 549–552. <https://doi.org/10.5754/hge12754> (2013).
42. Dang, S. *et al.* MiR-299-3p functions as a tumor suppressor via targeting Sirtuin 5 in hepatocellular carcinoma. *Biomed. Pharmacother.* **106**, 966–975. <https://doi.org/10.1016/j.biopha.2018.06.042> (2018).
43. Zhu, H. R. *et al.* Microarray expression profiling of microRNAs reveals potential biomarkers for hepatocellular carcinoma. *Tohoku J. Exp. Med.* **245**, 89–98. <https://doi.org/10.1620/tjem.245.89> (2018).

44. Zhang, Y. *et al.* Downregulated miR-621 promotes cell proliferation via targeting CAPRN1 in hepatocellular carcinoma. *Am. J. Cancer Res.* **8**, 2116–2129 (2018).
45. Ding, M. *et al.* Integrated analysis of miRNA, gene, and pathway regulatory networks in hepatic cancer stem cells. *Journal of translational medicine* **13**, 259–259. <https://doi.org/10.1186/s12967-015-0609-7> (2015).
46. Liu, Y. *et al.* miR-539 inhibits FSCN1 expression and suppresses hepatocellular carcinoma migration and invasion. *Oncol. Rep.* **37**, 2593–2602. <https://doi.org/10.3892/or.2017.5549> (2017).
47. Jiang, L., Li, X., Cheng, Q. & Zhang, B. H. Plasma microRNA might as a potential biomarker for hepatocellular carcinoma and chronic liver disease screening. *Tumour Biol.* **36**, 7167–7174. <https://doi.org/10.1007/s13277-015-3446-7> (2015).
48. Yang, X. W. *et al.* MicroRNA-1269 promotes proliferation in human hepatocellular carcinoma via downregulation of FOXO1. *BMC Cancer* **14**, 909. <https://doi.org/10.1186/1471-2407-14-909> (2014).
49. Mo, Y. *et al.* Long non-coding RNA XIST promotes cell growth by regulating miR-139-5p/PDK1/AKT axis in hepatocellular carcinoma. *Tumour Biol.* **39**, 1010428317690999. <https://doi.org/10.1177/1010428317690999> (2017).
50. Zhou, J. *et al.* MicroRNA-152 inhibits tumor cell growth by directly targeting RTKN in hepatocellular carcinoma. *Oncol. Rep.* **37**, 1227–1234. <https://doi.org/10.3892/or.2016.5290> (2017).
51. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
52. Vlachos, I. S. *et al.* DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucl. Acids Res.* **43**, W460–W466 (2015).
53. Di Bisceglie, A. M. Hepatitis B and hepatocellular carcinoma. *Hepatology (Baltimore, MD)* **49**, S56–S60. <https://doi.org/10.1002/hep.22962> (2009).
54. Chu, F. & Wang, L. Applications of support vector machines to cancer classification with microarray data. *Int. J. Neural Syst.* **15**, 475–484. <https://doi.org/10.1142/s0129065705000396> (2005).
55. Chang, C.-C. & Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27. <https://doi.org/10.1145/1961189.1961199> (2011).
56. Shinn-Ying, H., Li-Sun, S. & Jian-Hung, C. Intelligent evolutionary algorithms for large parameter optimization problems. *IEEE Trans. Evol. Comput.* **8**, 522–541. <https://doi.org/10.1109/TEVC.2004.835176> (2004).
57. Yerukala Sathipati, S., Huang, H.-L. & Ho, S.-Y. Estimating survival time of patients with glioblastoma multiforme and characterization of the identified microRNA signatures. *BMC Genom.* **17**, 1022–1022. <https://doi.org/10.1186/s12864-016-3321-y> (2016).
58. Yerukala Sathipati, S., Sahu, D., Huang, H.-C., Lin, Y. & Ho, S.-Y. Identification and characterization of the lncRNA signature associated with overall survival in patients with neuroblastoma. *Sci. Rep.* **9**, 5125. <https://doi.org/10.1038/s41598-019-41553-y> (2019).
59. Frank, E., Holmes, G., Witten, I. H., Trigg, L. & Hall, M. Data mining in bioinformatics using Weka. *Bioinformatics* **20**, 2479–2481. <https://doi.org/10.1093/bioinformatics/bth261> (2004).

Author contributions

Srinivasulu Yerukala Sathipati (SYS) and Shinn-Ying Ho (SYH) designed the system and carried out the detailed study. SYS participated in the design of the system and implemented programs. All authors participated in manuscript preparation and approved the final manuscript.

Funding

This work was supported by Ministry of Science and Technology ROC under the contract numbers MOST 109-2221-E-009-129-, 109-2740-B-400-002-, 108-2221-E-009-127-, 108-3011-F-075-001-, and 108-2218-E-029-004-, and was financially supported by the “Center For Intelligent Drug Systems and Smart Bio-devices (IDS²B)” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-71324-z>.

Correspondence and requests for materials should be addressed to S.-Y.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020