Identification of 5 microRNA biomarkers associated with the prognosis of uveal melanoma

Yabin Sun\(^a\), Xinmin Zhang\(^b\), Zhongyi Cong\(^b\), Siying Teng\(^a\),* #

1. Introduction

Uveal melanoma (UM) is the most common malignant ophthalmic tumor in adult Caucasians,\([1]\) with an incidence rate of about 4.3 per million people.\([2,3]\) Although there are multiple treatment options, such as surgical resection, radiotherapy, chemotherapy, transpupillary thermotherapy, and bioterrorism, the 5-year survival rates of UM are only about 50%, and the survival probability of patients with metastasis is <1 year.\([4–6]\) Therefore, identifying the risk factors that affect survival is essential for the early diagnosis and prompt treatment of UM.

The progression of tumor size, location, and growth rate caused by UM cell proliferation, migration, and invasion are closely correlated with patient survival.\([1,3,5]\) An increasing number of genetic signatures, such as chromosome abnormalities,\([7–9]\) gene mutations,\([8–10]\) and abnormal expression\([11,12]\) have also been associated with poorer prognosis in UM patients.

MicroRNAs are short noncoding RNAs that can bind to the 3′ untranslated region of targeted mRNAs to induce mRNA degradation or protein translation repression.\([13,14]\) Previous studies have demonstrated that the expression profiles of microRNAs contribute not only to physiological processes, but also to the occurrence and development of diseases, including multiple tumors.\([15–18]\) However, the role of microRNAs in UM prognosis needs to be better established.

In the present study, the data for microRNA and RNA expression and the clinical traits of patients with UM were retrieved from The Cancer Genome Atlas (TCGA). By using Weighted Correlation Network Analysis (WGCNA), we divided different microRNA expression profiles into 10 modules,\([19]\) among which blue and yellow modules were associated with UM survival. Five key microRNAs, hsa-miR-513a-5p, miR-506-3p, miR-508-3p, miR-140-3p, and miR-103a-2-5p, were further identified in the blue and yellow modules using cytoHubba in Cytoscape and
least absolute shrinkage and selection operator (LASSO) Cox regression analysis. Kaplan–Meier (KM) plots showed that the 5 microRNAs were highly correlated with the survival of patients ($P < .01$). The target genes for the 5 microRNAs were identified using miRTarBase, and correlation between expression of microRNAs and corresponding target genes analyzed by spearman method shown most genes were negatively correlated corresponding microRNAs except miR-103a-2-5p. All of the target genes with $P$ values $< 0.05$ by using Cox regression analysis were positively or negatively correlated with the expression of the associated microRNAs. In addition, gene ontology (GO), pathway, and disease enrichment analyses demonstrated that these 5 microRNAs were closely correlated with multiple bioprocesses and pathways in UM. Therefore, our study indicated that these 5 microRNA signatures could be used to predict the prognosis of UM patients.

2. Method

2.1. Data collection and preprocessing

The fragments per kilobase of exon model per million reads mapped (FPKM) quantified expression data for microRNA ($n = 80$) and RNA ($n = 80$), and the clinical trait data for patients with UM were downloaded from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/). TCGA belongs to public databases. The patients involved in the database have obtained ethical approval. The expression data for each mature isoform microRNA were obtained by summing the values per million reads. The values for all samples were then merged into 1 matrix. All FPKM RNA data were also merged into 1 matrix according to the ensemble IDs, which were converted into the gene name using the R packages “org.Hs.e.g..db” and “clusterProfiler”.

Figure 1. Analysis of network topology for various soft-thresholding powers. (A) The values of the soft threshold (x-axis) and the scale-free fit index (y-axis). (B) The values of the soft threshold (x-axis) and the mean connectivity (y-axis). (C) The logarithm of the whole network connectivity (x-axis) and the logarithm of the corresponding frequency distribution (y-axis). (D) Clustering dendrogram of microRNAs, with dissimilarity based on topological overlap, together with assigned dynamic tree cut module colors. From the results, 10 co-expression modules were constructed and are shown in different colors.
2.2. Weighted microRNA co-expression network analysis of uveal melanoma

The “WGCNA” package in R Weighted was used to construct a microRNA co-expression network, including correlation calculations between microRNAs, co-expression network construction, and module detection, and the modules were related to the clinical traits of patients. Since some of the microRNA expression levels in this study did not follow a normal distribution, the correlation between microRNAs was analyzed using the Spearman correlation test. A co-expression similarity matrix was constructed using the microRNA correlation value. An appropriate soft-thresholding value was selected to transform the similarity matrix into an adjacency matrix, which was then transformed into a topological overlap matrix. Afterward, the corresponding dissimilarity was calculated, and the hierarchical clustering tree of genes was derived based on the dissimilarity of the topological overlap matrix. Modules were defined as clusters of densely interconnected microRNAs. Since the number of qualified iso-microRNA was only 645, we set 10 as the minimum module size and the dynamic cut tree method was used to determine the microRNA modules. The correlations between the engine-microRNA modules and the clinical traits were then evaluated.

2.3. Identifying hub microRNAs for modules of interest

There were multiple microRNAs in each clinical related module. The microRNA signature (mS) referred to the correlation between the microRNA expression and clinical features in the module. MicroRNA module Membership (mM) referred to the correlation between the microRNA expression and the characteristics of the module. A good correlation between mS and mM indicated better biological significance, and modules with good correlation ($P < .05$) were selected for subsequent analysis. A microRNA correlation network for the modules of interest was constructed using WGCNA and exported into an edge and node list file that could be read by Cytoscape. The app “Cytohuba” in Cytoscape was used to analyze and identify the top key node microRNAs in each module.

2.4. Identifying microRNA signatures

Data samples of microRNAs or the target genes were divided into high- and low-expression groups based on the median expression level. The “survival” package in R was used to draw KM plots, conduct log rank tests, and perform Cox proportional hazards regression. Twenty hub microRNAs with $P$ values <0.01 in the Cox proportional hazards regression were identified as microRNAs associated with overall survival (OS). The LASSO Cox regression method in the “glmnet” package in R was employed to select microRNA signatures for the prediction of OS prognosis. Then, based on a linear combination of the regression coefficients ($\beta$) of the LASSO Cox regression model multiplied by the group number (high expression was 1 and low expression was 0), a prognostic microRNA signature for UM patients was constructed. Patients were then divided into high- and low-risk groups based on the median score of the risk signature. KM analysis and the area under the receiver operating characteristic (AUROC) curve were used to evaluate the prognostic performance of the microRNA signature using the “survival” and “survivalROC” packages in R.

2.5. The interaction between microRNA and the target RNA

miRtarBase is an experimentally validated microRNA-target gene interaction database. We downloaded homo sapiens miRTarBase is an experimentally validated microRNA-target gene interaction database. [23] We downloaded homo sapiens miRTarBase data and selected the target genes for microRNA signatures.

The correlations between the expression of node microRNAs and the target genes were analyzed using the Spearman correlation test.

2.6. GO and pathway analyses of microRNAs

The 5 microRNAs related to the target genes with $P$ values <0.05 in the interaction analysis were used for pathway, GO, and disease enrichment analyses. Specifically, the R packages “org.Hs.e.g..db” and “clusterProfiler”[24] were used to perform GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses, “ReactomePA”[25] was used to analyze reactome pathways, “DOSE”[26] was used to perform disease enrichment analysis, and “clusterProfiler” and “ggplot2” were used for visualization.

3. Results

3.1. Construction of microRNA co-expression modules for UM

Among the data downloaded from the TCGA database, clinical data with duplications or lack of survival or follow-up information, microRNA data with zero count samples (more than 1/3 in the expression matrix), and obvious outlier samples in the sample clusters were excluded. Finally, 645 microRNAs and 78 samples were included in the iso-microRNA expression matrix. Seven was choose from values predicted by the R package “WGCNA”, because taking 7 as the power value, the topology scale R$^2$ was largest (0.87), the mean connectivity is still above 0 (Fig. 1A–C) and 7 was used as a soft threshold for subsequent analysis. The “average” method was used to cluster the

![Module-trait relationships](image)

Figure 2. Module-trait relationships. Row: modules; Columns: status represents the survival status of patients and survival represents the survival time. Each cell contains the corresponding correlation and $P$ value. The table is color-coded by correlation according to the color legend.
samples and the “dynamic cut tree” method was used to generate co-microRNA expression modules. As shown in Figure 1D, the microRNA expression patterns were divided into 10 modules in different colors.

3.2. MicroRNA co-expression modules corresponding to clinical traits

Each module had a representative engine-microRNA, which was used to analyze their correlation with the clinical characteristics. The results showed that the OS of patients was closely related to the blue, red, and yellow modules with correlation coefficients of 0.36, −0.34, and −0.33, and P values of 0.001, 0.002, and 0.0046, respectively (Fig. 2). The stronger the correlation between mS and MM was, the more significant was the biological impact on the associated clinical trait. Scatterplots revealed that the microRNAs in the blue (P = 1.7e-10) and yellow modules (P = .0033), but not those in the red module (P = .14) had a strong association between mS and MM (Fig. 3). Therefore, the blue and yellow modules were selected for further studies.

3.3. Identification of microRNAs as OS signatures

MicroRNA interaction networks in the blue and yellow modules predicted by WGCNA were exported into edge and node list files for reading and visualization with Cytoscape (Fig. 4). Using the Matthews correlation coefficient (MCC) algorithm in the cytoHubba App in Cytoscape, the top 10 node microRNAs for each of the 2 modules were shown as triangles in Figure 4. Sixteen of the 20 hub microRNAs strongly correlated with UM OS (P < .01). LASSO Cox regression analysis and the log-rank test identified 5 microRNAs (larger triangles in Fig. 4), specifically hsa-miR-513a-5p, miR-506-3p, miR-508-3p, miR-140-3p,
and miR-103a-2-5p, as OS signatures. The KM plot indicated that the high- and low-risk groups had significantly different OS ($P < .001$, Fig. 5A). The AUROCs for 1, 3, and 5 years were 0.93, 0.92, and 0.904, respectively (Fig. 5B). Cox multivariate regression indicated that combining these 5 microRNAs into 1 signature yielded an independent prognostic factor ($P < .001$, Fig. 5C), suggesting the risk model for these 5 microRNAs could predict the prognosis of UM patients.

3.4. Interaction between microRNA signatures and target genes

MicroRNA regulates target genes expression by translational repression\cite{27} or activation,\cite{28} mRNA degradation\cite{29,30} and up regulating gene expression.\cite{30} Except for miR-103a-2-5p, the expression levels of the microRNAs were negatively correlated with most of their target genes selected from miRTarBase ($P < .05$, Fig. 6A,B). The correlations between the expression pattern of each target gene and UM OS were analyzed using Cox proportional hazards regression, and the genes with $P$ values <0.05 were selected. The interactions between the microRNAs and these selected genes were visualized with Cytoscape. Surprisingly, all of the selected target genes were significantly negatively or positively correlated with the corresponding microRNA ($P < .01$, Fig. 6C), and all of hazard ratio (HR) of negatively correlated target genes were <1, that indicated that these microRNAs maybe regulated negatively correlated target genes by mRNA degradation, while might regulated the

Figure 5. (A) KM plot for high- and low-risk groups of UM patients (B) Receiver operating characteristic curves for the prognostic microRNA signature of UM patients. The receiver operating characteristic curves for (1, 3) and 5 years are illustrated in gold, purple, and brown. (C) Nomogram of multivariate Cox regression analysis of risk score, age, stage, and cell-type.
Figure 6. (A) Violin and box plot of correlation between each microRNA and its target genes. (B) Violin and box plot of the $p$ values for the correlation between each microRNA and its target genes. (C) Visualization of the network connections between microRNAs and their target RNAs. Geometric figures in pink indicate the $P$ values for the correlation were $<$0.01, those in lavender indicate the $P$ values for the correlation were $<$0.05, and those in baby blue indicate the $P$ values for the correlation were equal to or $\geq$0.05. Ellipses and rectangles indicate negative and positive correlations, respectively. Black text indicates the $P$ values for Cox regression were $<$0.01, and white text indicates the $p$ values for Cox regression were $<$0.05. Black border indicates hazard ratio was more than 1, and white border indicates hazard ratio was $<$1.
positively correlated target genes by translational repression, activation or other unknown mechanisms.

3.5. Gene ontology, pathway, and disease enrichment analyses of the hub microRNAs

Although the data in miRTarBase have been experimentally verified, we used the Spearman correlation test to further screen the data for genes that were more likely to be target genes of microRNAs in UM. The genes with $P$ values $<0.05$ for interactions with microRNAs were selected for GO, KEGG pathway, reactome pathway, and Disease Ontology (DO) enrichment analysis to predict the biological processes, pathways, and diseases associated with the microRNAs involved. As shown in Figure 7A, GO analysis revealed that the 5 microRNAs were all involved in more than 1 biological process closely related to cancer, such as cell proliferation, cell cycle, cell migration, cell adhesion, cell differentiation and development, protein targeting, modification, and localization. KEGG and reactome pathway enrichment analyses also revealed that the 5 microRNAs were enriched in cancer-associated pathways, including the cell cycle, senescence and apoptosis, specific cancers, mitogen-activated protein kinase (MAPK), estimated glomerular filtration rate (EGFR), or Hippo (Fig. 7B and C). Furthermore, DO analysis also demonstrated that all 5 microRNAs were mainly enriched in tumor-related diseases (Fig. 7D).

4. Discussion

WGCNA can cluster genes with similar expression patterns into different modules and explore their relationship with specific traits or phenotypes of interest. It is widely used in the study of carcinogenesis as well as tumor development, treatment, prognosis, and other related processes. TCGA is a freely accessible and harmonized cancer genomics database. Previous studies have investigated the relationship between gene transcriptomes and the carcinogenesis as well as tumor development, treatment, prognosis, the negatively correlated target genes for these 5 microRNAs and reactome pathway analyses, and disease enrichment analyses, the negatively correlated target genes for these 5 microRNAs were found to be involved in important biological processes, pathways, and diseases associated with cancer.

In summary, we identified 5 microRNA signatures that could be used as potential prognostic biomarkers and treatment targets.

Figure 7. Enrichment analysis of gene ontology, pathways, and diseases of the hub microRNAs. (A) Top ranking results of GO enrichment analysis. (B) Top ranking results of KEGG pathway enrichment analysis. (C) Top ranking results of reactome pathway enrichment analysis. (D) Top ranking results of disease enrichment analysis.
for UM patients. However, since UM databases that can be used for verification are limited, experimental results are needed to validate our bioinformatics analysis of these microRNAs.

References
[1] Kaliki S, Shields CL, Shields JA. Uveal melanoma: estimating prognosis. Indian J Ophthalmol. 2015;63:93–102.
[2] Chattopadhyay C, Kim DW, Gombos DS, et al. Uveal melanoma: from diagnosis to treatment and the science in between. Cancer. 2016;122:2299–312.
[3] Tarlan B, Kiratli H. Uveal melanoma: current trends in diagnosis and management. Turk J Ophthalmol. 2016;46:123–37.
[4] Falzone L, Romano GL, Salemi R, et al. Prognostic significance of deregulated microRNAs in uveal melanomas. Mol Med Rep. 2019;12:2599–610.
[5] Sun T, Shao X, Zhang X, et al. Identification of a nine-miRNA signature for the prognosis of Uveal Melanoma. Exp Eye Res. 2019;180:242–49.
[6] Falzone L, Romano GL, Salemi R, et al. Prognostic significance of microRNA expression in uveal melanoma. Indian J Ophthalmol. 2015;63:1046–53.
[7] Dogrusoz M, Jager MJ. Genetic prognostic signature in uveal melanoma. Acta Ophthalmol. 2018;96:331–47.
[8] Wang Y, Lei X, Gao C, et al. MiR-506-3p suppresses the proliferation of uveal melanoma cells by targeting the SKA2 axis. Mol Immunol. 2020;12:20366–79.
[9] Su T, Shao X, Zhang X, et al. Value of circulating miRNA-1 in the diagnosis associated with metastasis and survival in patients with uveal melanoma. Oncotarget. 2020;11:14345–47.
[10] Sun C, Chiarugi D, Valleriani A. Single-molecule modeling of mRNA degradation and miRNA expression associated with metastasis and survival in patients with uveal melanoma. Onco Targets Ther. 2020;13:6051–62.
[11] Dongsoo M, Jager MJ. Genetic prognostic signature in uveal melanoma. Acta Ophthalmol. 2018;96:331–47.
[12] Li Y, Luo JT, Liu YM, et al. MiR-508-3p inhibits proliferation, induces cell cycle arrest and apoptosis in retinoblastoma by directly targeting the SKA2 axis. Mol Oncol. 2020.