Identification of differentially expressed metastatic genes and their signatures to predict the overall survival of uveal melanoma patients by bioinformatics analysis

Dan-Dan Zhao¹, Xin Zhao², Wen-Tao Li³

¹Shanxi Eye Hospital, Taiyuan 030001, Shanxi Province, China
²Datong Second People’s Hospital, Datong 037006, Shanxi Province, China
³Taiyuan University of Science and Technology, Taiyuan 030051, Shanxi Province, China

Correspondence to: Dan-Dan Zhao. Shanxi Eye Hospital, Taiyuan 030001, Shanxi Province, China. 947501467@qq.com

Received: 2019-06-06        Accepted: 2019-08-19

Abstract

● AIM: To identify metastatic genes and miRNAs and to investigate the metastatic mechanism of uveal melanoma (UVM).

● METHODS: GSE27831, GSE39717, and GSE73652 gene expression profiles were downloaded from the Gene Expression Omnibus (GEO) database, and the limma R package was used to identify differentially expressed genes (DEGs). Gene Ontology (GO) term enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed using the DAVID online tool. A comprehensive list of interacting DEGs was constructed using the Search Tool for the Retrieval of Interacting Genes (STRING) database and Cytoscape software. The Cytoscape MCODE plug-in was used to identify clustered sub-networks and modules of hub genes from the protein-protein interaction network. GEPIA online software was used to identify survival analysis of UVM patients (n=80) from the The Cancer Genome Atlas (TCGA) cohort. OncomiR online software was used to find that the miRNAs were associated with UVM prognosis from the TCGA cohort. TargetScan Human 7.2 software was then used to identify the miRNAs targeting the genes.

● RESULTS: There were 1600 up-regulated genes and 1399 down-regulated genes. The up-regulated genes were mainly involved in protein translation in the cytosol, whereas the down-regulated genes were correlated with extracellular matrix organization and cell adhesion in the extracellular space. Among the 2999 DEGs, five genes, Znf391, Mrps11, Htra3, Sulf2, and Smarcd3 were potential predictors of UVM prognosis. Otherwise, three miRNAs, hsa-miR-509-3p, hsa-miR-513a-5p, and hsa-miR-1269a were associated with UVM prognosis.

● CONCLUSION: After analyzing the metastasis-related enriched terms and signaling pathways, the up-regulated DEGs are mainly involved in protein synthesis and cell proliferation by ribosome and mitogen-activated protein kinase (MAPK) pathways. However, the down-regulated DEGs are mainly involved in processes that reduced cell-cell adhesion and promoted cell migration in the extracellular matrix through PI3K-Akt signaling pathway, focal adhesion, and extracellular matrix-receptor interactions. Bioinformatics and interaction analysis may provide new insights on the events leading up to the development and progression of UVM.

● KEYWORDS: gene ontology; bioinformatics; uveal melanoma; protein-protein interactions network; survival analysis

DOI:10.18240/ijo.2020.07.05

Citation: Zhao DD, Zhao X, Li WT. Identification of differentially expressed metastatic genes and their signatures to predict the overall survival of uveal melanoma patients by bioinformatics analysis. Int J Ophthalmol 2020;13(7):1046-1053

INTRODUCTION

Uveal melanoma (UVM) is the most common intraocular malignancy in adults[1], accounting for 5%-6% of all cases of primary melanoma and ranking second in place after skin melanoma[2]. Unfortunately, approximately half of all UVM patients die due to metastasis, and liver is where UVM cells most common metastases[3]. There are various treatments for UVM. However, the survival rate has not improved accordingly, suggesting that at the time of diagnosis, micro-metastases had already occurred[4]. These therapies, which mainly focus on the local treatment of the eye, do not extend past this organ and do not treat metastatic disease. Furthermore, there are few treatment strategies for metastatic UVM, which leads to poor disease prognosis[5]. Epithelial and cutaneous melanoma are different diseases,
revealing that unique management strategies for UVM are needed. With the development of precision medicine and genetic prognostication, genetic testing of UVM is becoming increasingly important[9]. The crux of the problem lies in the fact that the events leading up to the spread of UVM to the liver have not yet been fully defined[7]. As a result, it is difficult to develop effective interventions[9]. Most UVM patients have poor prognoses due to the advanced clinical stage at the time of diagnosis[9]. Therefore, it is important to investigate the molecular mechanism underlying the dissemination of UVM, which can also improve the accuracy of predicting the overall survival of patients with UVM.

Bioinformatics is a combination of biology, computer science and information technology to reveal the biological secrets of large and complex biological[90]. The Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) includes an extensive classification of high-throughput experimental data that is used by investigators to retrieve information on mRNA abundance from single and double channel microarray experiments, as well as genomic DNA and protein data[11]. The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) is a shared database set up by scientists around the world to unlock the secrets of all cancers, which has stored genomic data on more than 30 cancers[12]. The TCGA dataset, which contains more than two petabytes of publically available genomic data, has greatly contributed to the prevention, diagnosis and treatment of cancer. Both GEO and TCGA databases have enhanced our understanding of cancer, and efficient integrated bioinformatic methods have been developed for large-scale analysis of cross-platform high-throughput data.

In this study, we analyzed data from multiple microarrays to identify genes and miRNAs associating with metastasis and UVM prognosis. First, we downloaded three datasets from the GEO database and identified differentially expressed genes (DEGs) by bioinformatic methods. Second, we analyzed the functions of these genes, as well as the associated pathways and protein-protein interactions (PPI). Last, a prognostic gene signature was generated based on TCGA HNSC RNA-Seq data that demonstrated good performance for predicting the overall survival of patients with UVM. This robust prognostic gene signature was successfully validated in an independent cohort of patients.

**MATERIALS AND METHODS**

**Microarray Data** Three original datasets comparing gene expression profiles between metastatic and non-metastatic UVM were downloaded from the GEO database. The accession numbers were GSE27831 (18 non-metastatic samples and 11 metastatic samples), GSE39717 (39 non-metastatic samples and 2 metastatic samples) and GSE73652 (8 non-metastatic samples and 5 metastatic samples). The microarray data of GSE27831, GSE39717, and GSE73652 were based on GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array), GPL6098 (Illumina HumanRef-8 v1.0 Expression BeadChip) and GPL10558 (Illumina HumanHT-12 v4.0 Expression BeadChip), respectively.

**Screening of Differential Genes** Robust Multi-array Average (RMA) approach was used to standardize the microarray data background, and then converted into expression measures using the affy R package. The limma R package was subsequently used to identify DEGs. The cut-off criteria were P<0.05 and log₂FC (expression in mutant background) >1 for DEGs based on the Benjamini-Hochberg (BH) procedure.

**Gene Ontology Term and KEGG Pathway Enrichment Analysis** Gene Ontology (GO) term enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG, a set of high-throughput gene and protein pathways) pathway analysis were performed using the Database for Annotation, Visualization and Integration Discovery tool (DAVID, https://david.ncifcrf.gov/). GO consisted of three items, namely, the enriched biological process, cellular component and molecular function[13]. P<0.05 was set as the threshold value.

**Protein-Protein Interaction Network and Module Analysis** The STRING (http://string-db.org/) was used to comprehensively analyze the interactions between the DEGs. Cytoscape (http://www.cytoscape.org/) was employed for PPI network visualization of the DEGs. The clustered sub-networks and modules of the hub genes from the PPI network were searched by The Cytoscape MCODE plug-in. The default parameters were as follows: the combined score of >0.7 was set as the cut-off criteria for the PPI network and a degree cutoff ≥2, node score cutoff ≥0.2, K-score ≥2 and max depth of 100 for screening hub genes.

**Generation and Prediction of Prognostic Gene Signatures** We used Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/index.html) for survival analysis of the UVM patients (n=80) from the TCGA cohort. Univariate Cox proportional hazards regression equation was used to analyze the expression of genes related to overall survival. P-value <0.05 was considered statistically significant. Subsequently, we extracted the common genes from the survival-related genes and DEGs. Genes with hazard ratios (HRs) <1 were considered protective genes and HRs >1 were risk factors.

We used the TargetScan database (http://www.targetscan.org/vert_72/) to retrieve miRNAs targeting miRNAs, which predicted the target of miRNA by matching the miRNA seed region. Predictions with poorly conserved sites are provided, as well as mismatches in the seed region that are compensated by conserved 3’ pairing and centered sites. The intersect function in the R package was used to identify the common miRNAs.
between prognostic miRNAs identified by OncomiR and retrieved miRNAs identified by TargetScan. Venn Diagram R package was used to draw the Venn diagram.

RESULTS

Identification of Differentially Expressed Genes Among GSE27831, GSE39717, and GSE73652 Datasets Analysis of GSE27831, GSE39717, and GSE73652 was performed by the limma R package, and the data were visualized as volcano plots, with red and green dots representing significantly up-regulated (log2 FC>1, P<0.05) or down-regulated (log2 FC<-1, P<0.05) genes, respectively. Of these, 143 (124 up-regulated and 19 down-regulated), 2526 (1297 up-regulated and 1229 down-regulated) and 330 (179 up-regulated and 151 down-regulated) dots were matched to unique genes in GSE27831, GSE39717, and GSE73652 datasets, respectively (Figure 1). There were 1600 and 1399 up-regulated and down-regulated genes, respectively.

Heatmaps were generated by Morpheus software (https://software.broadinstitute.org/morpheus/) based on the expression levels of DEGs in GSE27831, GSE39717, and GSE73652 datasets (Figure 2). In the heat map, each column represented a biological sample and each row represented a gene. The color indicated the expression level of the gene between metastatic and non-metastatic tissues. The heatmaps showed that the expression levels of the DEGs between the metastatic and non-metastatic tissues were significantly different.

Gene Ontology and Pathway Enrichment Analysis For biological processes, the up-regulated genes were mainly enriched in SRP-dependent co-translational protein targeting to the membrane and translational initiation, whereas the down-regulated genes were mainly enriched in extracellular matrix organization and cell adhesion. For cellular components, the up-regulated genes were mainly found in the cytosol, whereas the down-regulated genes were mainly found in the extracellular space. For molecular functions, the up-regulated genes were mainly associated with protein binding and structural constituent of ribosome, whereas the down-regulated genes were mainly associated with extracellular matrix structural constituent (ECM-receptor interaction; Figure 3). KEGG pathway enrichment analysis showed that most of the
up-regulated genes were enriched in ribosome and mitogen-activated protein kinase (MAPK) pathways, whereas the down-regulated genes were enriched in complement and coagulation cascades, PI3K-Akt signaling pathway, focal adhesion and ECM-receptor interaction.

Construction of the PPI Network and Identification of Key Candidate Genes The identified DEGs were mapped to the PPI network using the STRING database and the Cytoscape software (Figure 4A). The top ten highest-scoring nodes were then screened as key candidate genes. The up-regulated genes were Rps27a, Rpl11, Rpl17, Rplp0, Rps16, Rps9, Rps8, Rpl8, Rpl10a, and Rps25, whereas the down-regulated genes were Agt, Gnb4, Saa1, Gpr17, F2, Plg, C5, Alb, Fga, and Fgg. The PPI network was further analyzed by using the MCODE plug-in, and then the most important modules were selected for further analysis. The modules of the up-regulated and down-regulated genes consisted of 27 nodes and 43 nodes, respectively (Figure 4B). According to the KEGG pathway analysis, the up-regulated gene module mainly involved in ribosome, whereas the down-regulated gene module mainly involved in Cytokine-cytokine receptor interaction (Figure 4C).

Determination of the Prognostic Gene Signature To determine the prognostic signature, GEPIA was used to analyze the UVM patients (n=80) from the TCGA cohort. One hundred genes were identified as connected with the overall survival of the patients. Five out of 100 DEGs were common; 54 genes were enriched in KEGG pathways associated with cell proliferation and death, whereas 46 genes were enriched in KEGG pathways associated with PI3K-Akt signaling pathway and focal adhesion. The hazards ratio was calculated based on the Cox proportional hazards model to analyze the five genes as potential predictors of UVM prognosis. The HRs were as follows: 0.065 (Znf391), 9.5 (Mrps11), 16 (Htra3), 37 (Sulf2), and 16 (Smarcd3).

miRNAs Associated with the Prognosis of UVM To identify the miRNAs associated with the prognosis, OncomiR software was used to analyze the UVM patients from the TCGA cohort. One hundred and nineteen miRNAs were identified. TargetScan Human 7.2 software was then used to identify the miRNAs targeting the five prognostic signatures. The 100 highest-scoring miRNAs were selected, with 38 miRNAs targeting Smarcd3, which was used to reveal the common miRNAs. The common miRNAs were as follows: two miRNAs targeting Znf391, three miRNAs targeting Mrps11, five miRNAs targeting Htra3 and three miRNAs targeting Sulf2. No miRNAs targeting Smarcd3 were identified (Figure 6A, Table 1). The HR based on the Cox proportional hazards model was calculated for miRNAs with log_{FC}>1 (Figure 6B). We identified three miRNAs with significant differences in their expression levels, namely, S=-2.696 * EsriR-509-3-5p+2.249 * EsriR-513a-5p+2.653 * EsriR-1269a. The data in parentheses represent the log_{FC}, with >0 representing up-regulated genes in living patients and <0 representing up-regulated genes in deceased patients.

DISCUSSION UVM is a lethal disease affecting the uveal layer of the eye[14]. The 5-year survival rate of UVM is 75%, and the main cause of death is metastasis, in which approximately 95% patients die from hepatic metastasis. Unfortunately, metastasis is often untreatable[15]. The dissemination of UVM is different from the spread of other cancer cells, which can spread throughout the body through the lymphatic system[16]. Without the lymphatic system in the eye, cancer cells travel via the circulatory system, often to the liver where they develop into secondary lesions[17]. This may be due to the fact that the liver is rich in serum factors, which readily attract cancer cells[18]. However, metastasis is a multistep process in which cells undergo various phenotypical changes, break away from the primary tumor and travel to a distant secondary site through dense connective tissue[19-21]. And, the events leading up to metastasis are not yet known. We used three microarray datasets to identify DEGs. GO analysis revealed that up-regulated DEGs referred to SRP-dependent co-translational protein targeting to the membrane and structural constituent of ribosome in the cytosol. However, down-regulated DEGs associated with extracellular matrix organization, and extracellular matrix structural constituent, in extracellular space. For KEGG pathway, ribosome and MAPK signaling associated with up-regulated DEGs, whereas PI3K-Akt signaling pathway, focal adhesion and extracellular matrix-receptor interactions associated with down-regulated DEGs. Taken collectively, these findings indicate that up-regulated DEGs mediated protein synthesis and cell proliferation by ribosome and MAPK pathways, whereas down-regulated DEGs resulted in decreased cell adhesion and promoted cell migration in the extracellular matrix through PI3K-Akt signaling pathway, Focal adhesion and extracellular matrix-receptor interactions. Further studies of the role of these key pathways in the development and metastasis of UM may provide new clues for clinical treatment of UVM.

The PPI network provides the interactions among the DEGs. The top ten up-regulated interacting DEGs were Rps27a, Rpl11, Rpl17, Rplp0, Rps16, Rps9, Rps8, Rpl8, Rpl10a, and Rps25, with all genes encoding ribosomal proteins. Ribosomal proteins have important roles in protein biosynthesis and DNA repair as well as cell differentiation. Studies have reported ribosomal protein overexpression in various types of cancer, such as those of the liver, stomach, rectum and esophagus, and based on our results, ribosomal proteins are also involved in metastasis[22-24]. As such, further studies on the elevated expression of ribosomal proteins in UVM, may be beneficial to the development of new treatment strategies for UVM. By contrast, the top ten down-regulated interacting
**Figure 3** Gene ontology analysis of the DEGs according to their biological process, cellular component and molecular function


**Figure 4** The PPI network of the DEGs in GSE27831, GSE39717 and GSE73652 datasets

A: The PPI network of the DEGs in GSE27831, GSE39717, and GSE73652 datasets; B: The sub-networks were generated by the Cytoscape MCODE plug-in; C: KEGG pathway analysis of the DEGs in the sub-networks to identify the canonical pathways. A1, B1, C1: Up-regulated genes; A2, B2, C2: Down-regulated genes.
DEGs were Agt, Gnb4, Saa1, Gpr17, F2, Plg, C5, Alb, Fga, and Fgg. These genes which mainly related to the process of hemostasia and cruor, maybe due to the damage of liver. Five prognostic gene signatures, namely, Znf391, Mrps11, Htra3, Sulf2, and Smarcd3 were identified from the TCGA dataset. Znf391 and Mrps11 were up-regulated, whereas Htra3,
Sulf2, and Smarcd3 were down-regulated. Znf391 may be involved in the transcriptional regulation of other genes [29], and Mrps11 encodes a nuclear protein that is involved in mitochondrial protein synthesis, as reported in CESC, DLBC, and OV overexpression studies [28]. Htra3 protein was first found to play an important role in female pregnancy. In recent years, studies have shown that the expression of Htra3 in a variety of solid tumors is significantly different from that in its corresponding normal tissues, suggesting that Htra3 may be involved in the development of tumors. However, exactly how it is involved in this process has not been clarified [27]. Sulf2 is responsible for modifying the sulfated sugar (heparan sulfate), which is present on the cell surface and plays a key role in controlling cell functions, including promoting cell growth and preventing cell death [28]. Smarcd3 is a member of the SWI/SNF family of proteins that regulate the transcriptional activation of several genes by altering the chromatin structure [29]. According to Univariate Cox hazards regression, HRs ≥1 indicated that genes were risk factors for the disease, whereas HRs ≤1 indicated that genes were protective factors. Znf391, whose expression was up-regulated, was a protective factor; however, Mrps11, Htra3, Sulf2, and Smarcd3 were risk factors. The expression of Htra3, Sulf2, and Smarcd3 was down-regulated. Mrps11 was the most robust gene for the poor prognosis of UVM, and it has practical value in predicting whether dissemination will occur. Further studies are needed to address the limitations of this study and to gain new insights on significance of the prognostic gene signatures.

We identified 13 common miRNAs from the 119 miRNAs identified by OncomiR and the miRNAs targeting the five prognostic signatures. Among the 13 miRNAs, hsa-miR-509-3p (3.92), hsa-miR-513a-5p (3.2) and hsa-miR-1269a (2.1) showed significant differences in their expression levels. All the three miRNAs targeted the risk factors, namely, HTRA3 and SULF2. Therefore, the three miRNAs can be regarded as protective factors. miR-509 has been reported to regulate Htra3, Sulf2, and Smarcd3 and to target Pik1, thereby controlling cell proliferation, cell migration, vimentin-based cytoskeletal dynamics and focal adhesion assembly [30].

In conclusion, we identified genes and miRNAs involved in metastasis that may have clinical application in the early diagnosis and the prediction of recurrence and overall survival of patients with UVM, which may improve personalized therapies in the future.

ACKNOWLEDGEMENTS

Foundations: Supported by the Natural Science Foundation for Young Scholars of Shanxi (No.201801D221256); the Science Foundation for Young Scholars of Shanxi Eye Hospital (No.Q201803).

Conflicts of Interest: Zhao DD, None; Zhao X, None; Li WT, None.

REFERENCES


27 Bowden MA, Di Nezza-Cossens LA, Jobling T, Salamonsen LA, Nie GY. Serine proteases HTRA1 and HTRA3 are down-regulated with increasing grades of human endometrial cancer. Gynecol Oncol 2006;103(1):253-260.

