MYL2 as a potential predictive biomarker for rhabdomyosarcoma

Junning Wang, MB[a,*], Shang Gao, MB[b], Keqin Dong, MS[c], Peiyuan Guo, MS[c], Meng-jie Shan, MD[d,e]

Abstract

Rhabdomyosarcoma (RMS) is a common malignant soft tissue sarcoma, which is the third most common soft tissue sarcoma after malignant fibrohistoma and liposarcoma. The discovery of potential postbiomarkers could lead to early and more effective treatment measures to reduce the mortality of RMS. The discovery of biomarker is expected to be the direction of targeted therapy, providing a new direction for the precise treatment of RMS.

Gene Expression Omnibus database was used to download the tow gene profiles, GSE28511 and GSE135517. GEO2R was applied to identify differently expressed genes (DEGs) between RMS and normal group. Database for Annotation, Visualization and Integrated Discovery and Metascape can perform the enrichment analysis for the DEGs. Protein-protein interaction network was constructed, and the hub genes was identified by the Cytoscape. Expression and overall survival analysis of hub genes were performed.

A total of 15 common DEGs were screened between RMS and normal tissues. The enrichment analysis here showed that the DEGs mainly enriched in the muscle filament sliding, myofibril, protein complex, sarcomere, myosin complex, nuclear chromosome, and tight junction. The 6 hub genes (DNA Topoisomerase II Alpha, Insulin Like Growth Factor 2, HIST1H4C, Cardiomyopathy Associated 5, Myosin Light Chain 2 [MYL2], Myosin Heavy Chain 2) were identified. Compared with the normal tissues, MYL2 were down-regulated in the RMS tissues. RMS patients with low expression level of MYL2 had poorer overall survival times than those with high expression levels ($P < .05$).

In summary, lower expression of MYL2 was 1 prediction for poor prognosis of RMS. MYL2 is hope to be the target of therapy, which leads to more effective treatment and reduces the mortality rate of RMS.

Abbreviations: CMYA5 = Cardiomyopathy Associated 5, DAVID = Database for Annotation, Visualization and Integrated Discovery, DEGs = differently expressed genes, GEO = Gene Expression Omnibus, GEPIA = Gene Expression Profiling Interactive Analysis, IGF2 = Insulin Like Growth Factor 2, MYH2 = Myosin Heavy Chain 2, MYL2= Myosin Light Chain 2, PPI = protein-protein interaction, RMS = rhabdomyosarcoma, TOP2A = DNA Topoisomerase II Alpha.

Keywords: bioinformatics, biomarker, MYL2, prognosis, rhabdomyosarcoma

1. Introduction

Rhabdomyosarcoma (RMS) is a malignant tumor arising from or derived from mesenchymal cells. RMS is a common malignant soft tissue sarcoma. It is most common in children and adolescents, and accounts for about 3% of all malignant tumors in children.[1] RMS is the third most common soft tissue sarcoma after malignant fibrohistoma and liposarcoma. RMS has a high degree of malignancy and a poor prognosis.[2] There is generally no typical radiological characteristics, and no calcification. Tumors can invade and destroy adjacent bones, especially the skull, forearm, hand, and foot. At present, the main clinical treatment means are radiotherapy and chemotherapy, and there are no clear target drugs.[3] Radiation therapy and chemotherapy have a large side effect, so it is critical to look for targets to accurately treat RMS and improve patient survival. The cause of RMS is unknown. It is a soft tissue malignancy composed of rhabdomyocytes with different degrees of differentiation. The disease may be associated with genetic factors, chromosomal abnormalities, gene fusion, and other factors. Therefore, in-depth study of the molecular mechanism of RMS is particularly important.

Bioinformatics technology is the processing and analysis of all kinds of genome data, and it is also an important content of genome research. Bioinformatics is an indispensable part of genomics research.[4] The development of bioinformatics is not only the analysis of genome data, but also the comprehensive analysis of known or new gene products. Bioinformatics is a

science that used computers as tools to store, retrieve, and analyze biological information in the study of life sciences. It is one of the important frontier fields of life science and natural science, and will also be one of the core fields of natural science in the 21st century.[5] The development of this technique has made it possible to explore and identify genetic targets related to diseases.[6]

Myosin Light Chain 2 (MYL2) is a protein-coding gene. Pathways associated with MYL2 include Guidance Cues and Growth Cone

Motility and Activation of cAMP-Dependent proteinkinase A. The Gene Ontology annotation associated with the MYL2 includes calcium ion binding and actin monomer binding.[7] However, the relationship between MYL2 and RMS is unclear.

Therefore, this paper intends to use bioinformatics technology to excavate the differentially expressed genes and hub genes between RMS and normal tissue, and conduct enrichment analysis, pathway analysis, and survival analysis. RMS is a significant harm to child injuries, and the discovery of targeted
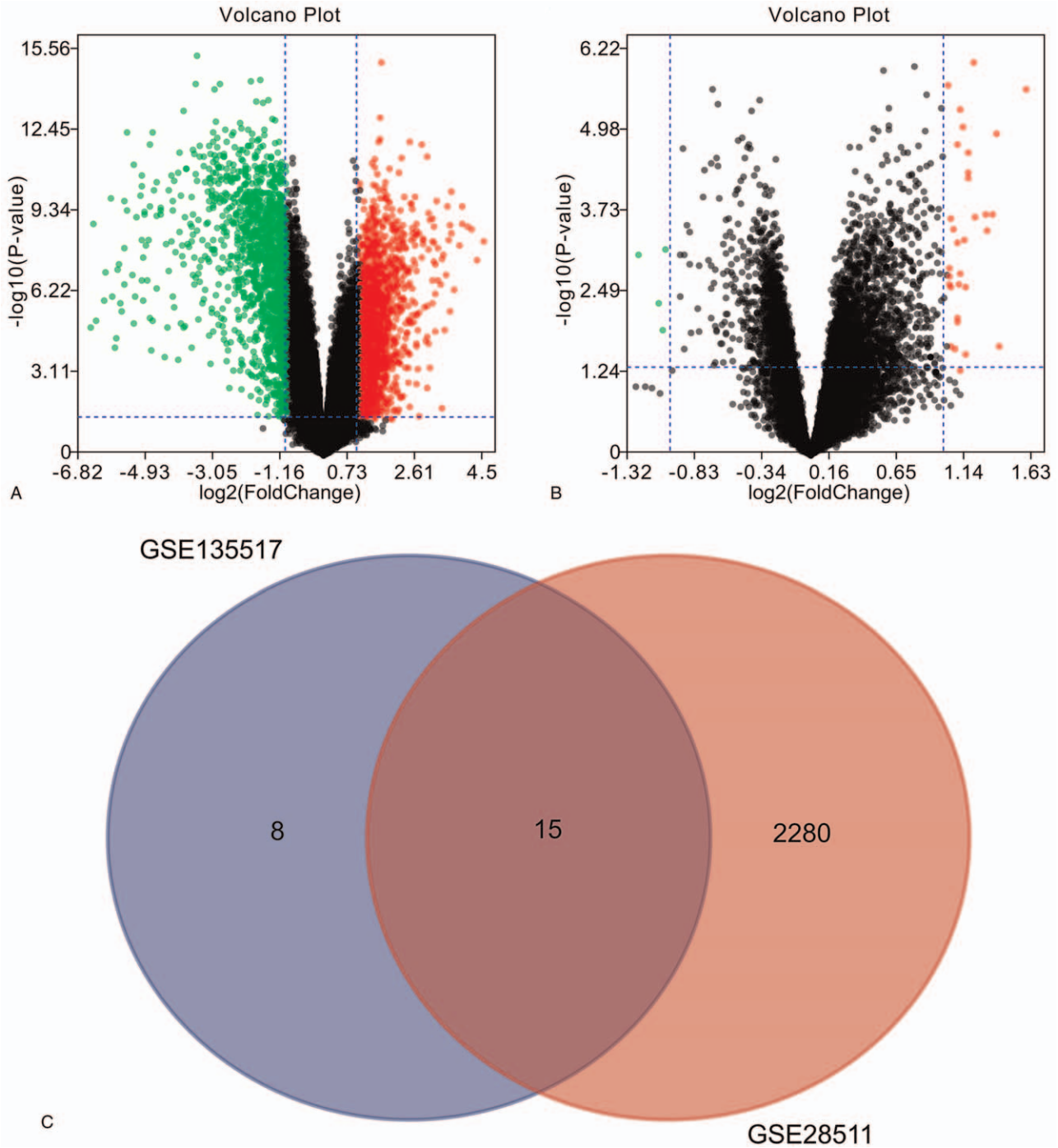

Figure 1. The DEGs. (A) The volcano map presenting DEGs in the GSE28511. (B) The volcano map presenting DEGs in the GSE135517. (C) The Venn diagram revealed 15 DEGs shared by the 2 datasets. DEG = differently expressed genes.

Table 1

Functional enrichment of DEGs in rhabdomyosarcoma.

Term	Description	Count in gene set	P value
GO:0060047	Heart contraction	2	.008486
GO:0030049	Muscle filament sliding	2	.02462
GO:0048511	Rhythmic process	2	.034821
GO:0006936	Muscle contraction	2	.067921
GO:0044267	Cellular protein metabolic process	2	.074661
GO:0030016	Myofibril	3	2.05E-04
GO:0043234	Protein complex	4	.003468
GO:0031672	A band	2	.010705
GO:0030017	Sarcomere	2	.029558
GO:0016459	Myosin complex	2	.037747
GO:0000228	Nuclear chromosome	2	.039969
hsa04530	Tight junction	2	.049648

DEGs = differentially expressed genes, GO = Gene Ontology.

therapy is essential for treatment. The exploration of biomarkers related to the prognosis of RMS is expected to be the research direction of targeted therapy and provide a new direction for the precise treatment of RMS. In addition, as a biomarker of poor prognosis, MYL2 can also guide the timely selection of effective clinical management measures to reduce the mortality of RMS.

2. Material and methods

2.1. Dataset

We downloaded tow gene profiles, GSE28511 and GSE135517, from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The GSE28511 includes 18 RMS
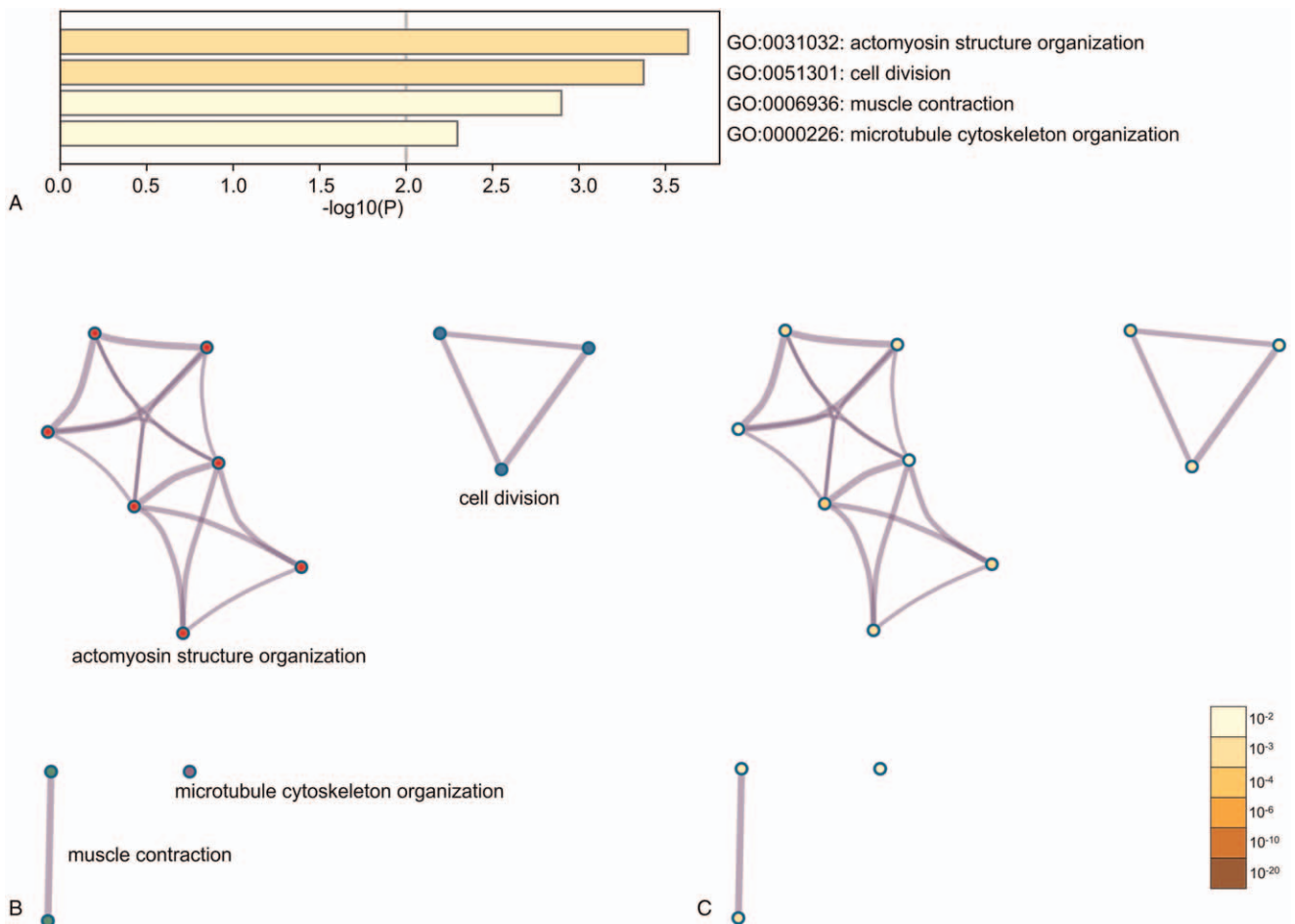
Figure 2. The enrichment analysis of DEGs by Metascape. (A) Heatmap of enriched terms across input differently expressed gene lists, colored by P values, via the Metascape. (B) Network of enriched terms colored by cluster identity, where nodes that share the same cluster identity are typically close to each other. (C) Network of enriched terms colored by P value, where terms containing more genes tend to have a more significant P value. DEG = differently expressed genes.

samples and 6 normal samples, while the GSE135517 includes 48 RMS samples and 15 normal samples.

2.2. DEGs identification

We applied GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r), an online tool based on GEOquery and limma R packages, to identify differently expressed genes (DEGs) between esophagus cancer and normal group. The cutoff criteria were that P value $\leq.05$ and a log (fold change) ≥ 1 or log (fold change) ≤ -1.

2.3. DEGs annotation

Database for Annotation, Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/home.jsp) and Metascape (http://metascape.org/gp/index.html) are 2 powerful annotation tools which can perform the biological process, cellular component, molecular function, and Kyoto Encyclopedia of Genes and Genomes analysis on genes. We annotated the function of common DEGs through DAVID and Metascape.

2.4. Protein-protein interaction network construction

The Search Tool for the Retrieval of Interacting Genes (http://string-db.org), can convert DEGs into expressed proteins and construct protein-protein interaction (PPI) network. We got PPI network of common DEGs through Search Tool for the Retrieval of Interacting Genes, and visualized it by Cytoscape (version 3.7.2).

2.5. Hub genes identification and expression

The cytoHubb was applied to screened out hub genes from the PPI network when the degree ≥ 6. Then, the clustering analysis of expression level of hub genes was performed using heatmaps based on the GSE28511 and GSE135517.

2.6. Overall survival analysis of hub genes

Effect of hub gene expression for overall survival was analyzed by the Kaplan-Meier Plotter (http://kmplot.com/analysis/index.php?p=background). The Kaplan Meier plotter is capable to assess the effect of 54k genes (mRNA, miRNA, protein) on survival in 21 cancer types. Primary purpose of the tool is a meta-analysis based discovery and validation of survival biomarkers. Furthermore, the Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/) was also used to analyze the role of hub genes on the overall survival so as to verify the results of Kaplan-Meier Plotter.

3. Results

3.1. DEGs

Two volcano plots present the DEGs in the GSE28511 (Fig. 1A) and GSE135517 (Fig. 1B). The Venn diagram revealed 15 DEGs shared by the 2 datasets (Fig. 1C).

3.2. Enrichment analysis of DEGs with DAVID

Through DAVID analysis, the results of the Gene Ontology analysis here showed that the DEGs mainly enriched in the muscle filament sliding, rhythmic process, myofibril, protein

complex, a band, sarcomere, myosin complex, nuclear chromosome, and tight junction (Table 1).

3.3. DEGs annotation with Metascape

Enrichment analysis by Metascape was displayed in Figure 2. The DEGs related with biological process, cellular component, molecular function, and Kyoto Encyclopedia of Genes and Genomes enrichment analysis were displayed in actomyosin structure organization, cell division, muscle contraction, microtubule cytoskeleton organization (Fig. 2A, B, C).

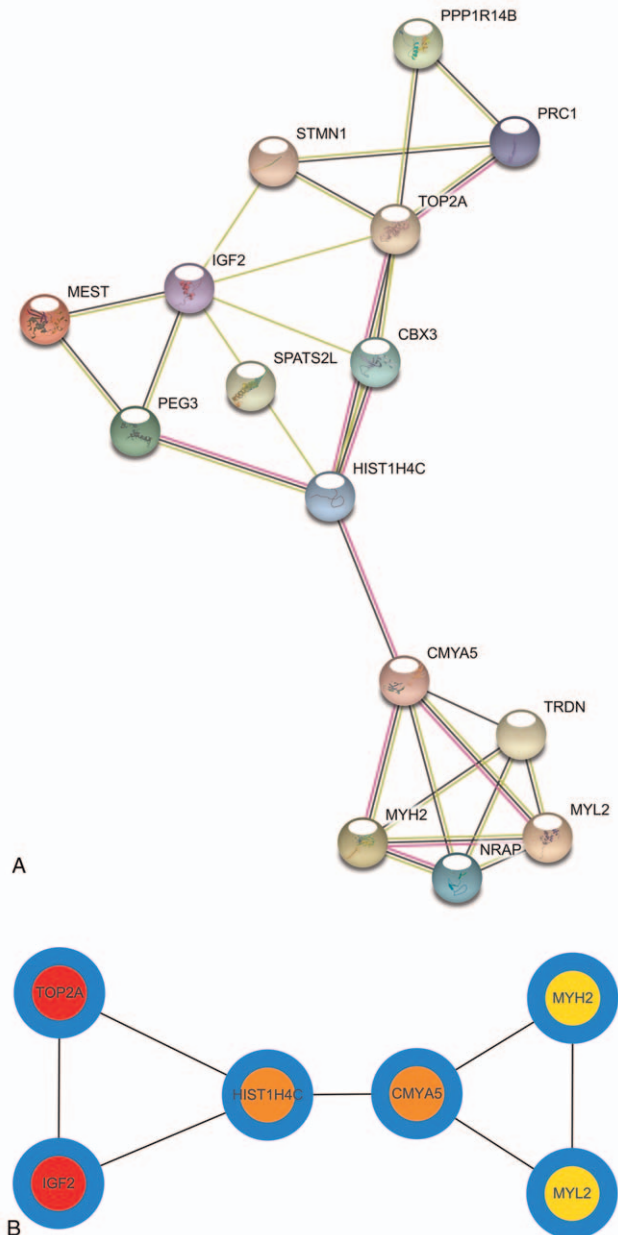**Figure 3.** Protein-protein interaction network and hub genes. (A) PPI network showed that there exists closed relationship among DEGs. (B) The 6 hub genes (TOP2A, IGF2, HIST1H4C, CMYA5, MYL2, MYH2). CMYA5 = Cardiomyopathy Associated 5, DEG = differently expressed genes, IGF2 = Insulin Like Growth Factor 2, MYH2 = Myosin Heavy Chain 2, MYL2 = Myosin Light Chain 2, PPI = protein-protein interaction, TOP2A = DNA Topoisomerase II Alpha.

3.4. Protein-protein interaction network and hub genes

PPI was displayed in Figure 3A, which showed that there exists closed relationship among DEGs. The 6 hub genes (DNA Topoisomerase II Alpha [TOP2A], Insulin Like Growth Factor 2 [IGF2], HIST1H4C, Cardiomyopathy Associated 5 [CMYA5], MYL2, Myosin Heavy Chain 2 [MYH2]) were shown in Figure 3B.

3.5. The validation of expression level of hub genes

Two heat maps showed the expressions of the hub genes in GSE28511 (Fig. 4A) and GSE135517 (Fig. 4B). Hierarchical clustering allowed for simple differentiation of RMS tissues from normal tissues via the expression levels of hub genes. Compared with the normal tissues, MYH2, CMYA5, and MYL2 were down-regulated in the RMS tissues. However, the expression of TOP2A, IGF2, and HIST1H4C in the RMS tissues was higher than the normal tissues.

3.6. Association between hub gene expression and overall survival via the Kaplan-Meier Plotter

RMS patients with high expression levels of MYH2, TOP2A, and HIST1H4C had poorer overall survival times than those with low

expression levels ($P < .05$; Fig. 5). RMS patients with low expression level of MYL2 had poorer overall survival times than those with high expression levels ($P < .05$; Fig. 5). However, there was no statistically significant effect on overall survival associated with the expression of CMYA5 and IGF2 ($P > .05$; Fig. 5).

3.7. Verification for the role of hub genes on the overall survival with the GEPIA

The significant hub genes (MYH2, MYL2, TOP2A, and HIST1H4C) in the Kaplan-Meier Plotter were entered in the GEPIA. RMS patients with low expression level of MYL2 had poorer overall survival times than those with high expression levels ($P < .05$; Fig. 6). However, there was no statistically significant effect on overall survival associated with the expression of MYH2, TOP2A, and HIST1H4C ($P > .05$; Fig. 6).

4. Discussion

As a common malignant soft tissue tumor in children, RMS has high degree of malignancy, low survival rate, and unclear pathogenesis.[1] In-depth exploration of the molecular mecha-
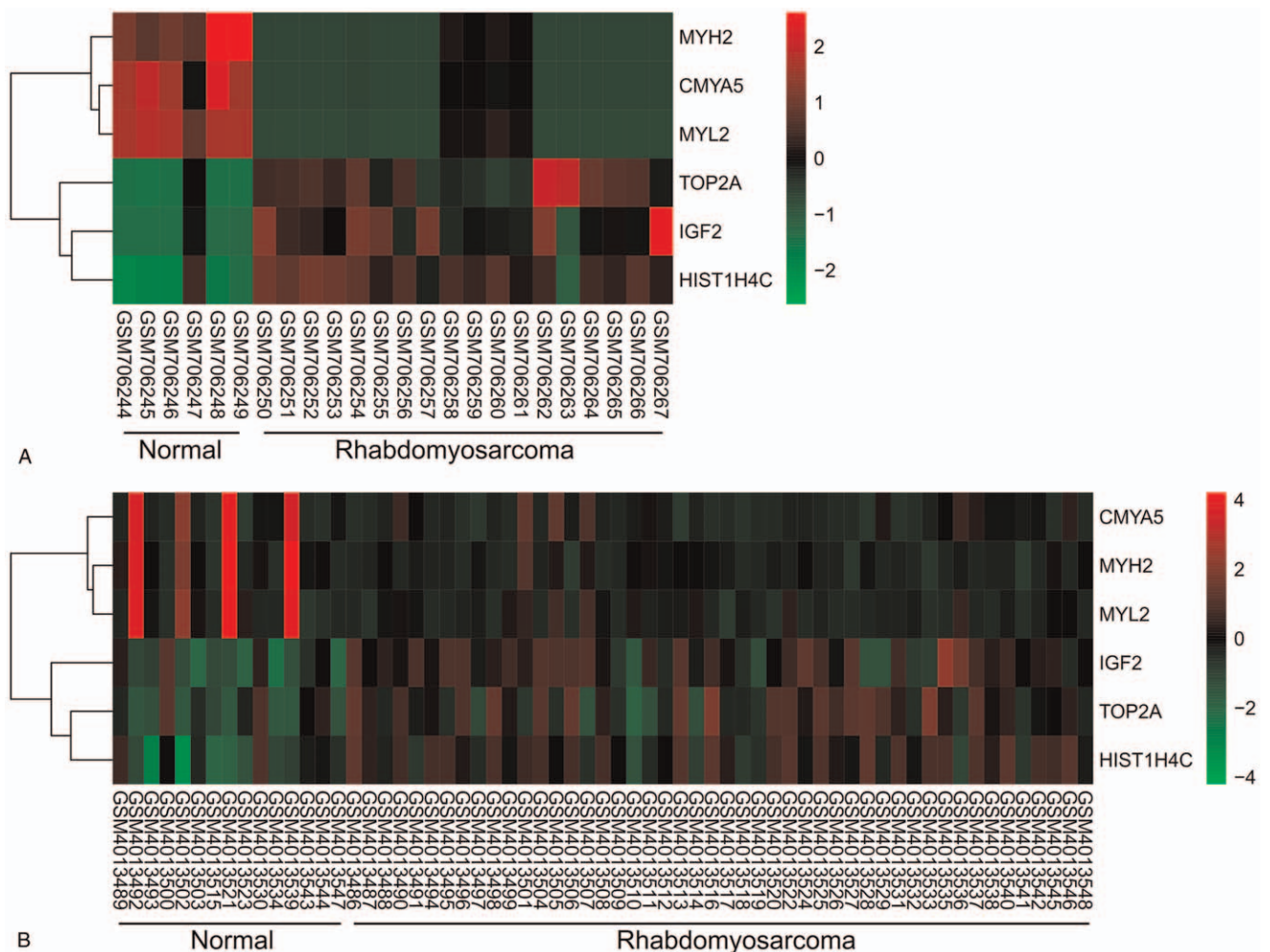
Figure 4. Two heat maps showed the expressions of the hub genes. (A) GSE28511; (B) GSE135517. CMYA5 = Cardiomyopathy Associated 5, IGF2 = Insulin Like Growth Factor 2, MYH2 = Myosin Heavy Chain 2, MYL2 = Myosin Light Chain 2, TOP2A = DNA Topoisomerase II Alpha.

Figure 5. Association between hub gene expression and overall survival via the Kaplan-Meier Plotter. CMYA5 = Cardiomyopathy Associated 5, IGF2 = Insulin Like Growth Factor 2, MYH2 = Myosin Heavy Chain 2, MYL2 = Myosin Light Chain 2, TOP2A = DNA Topoisomerase II Alpha.

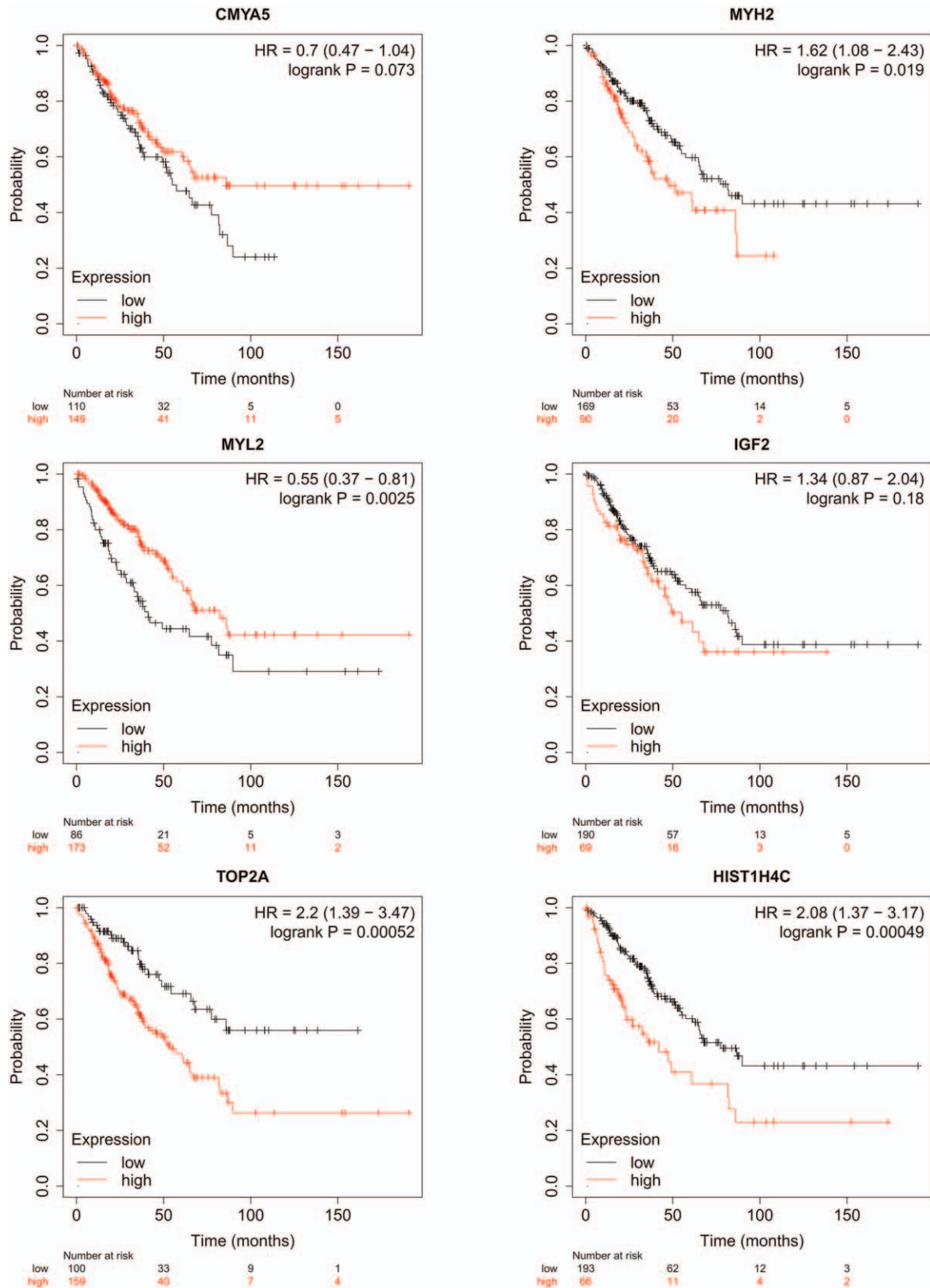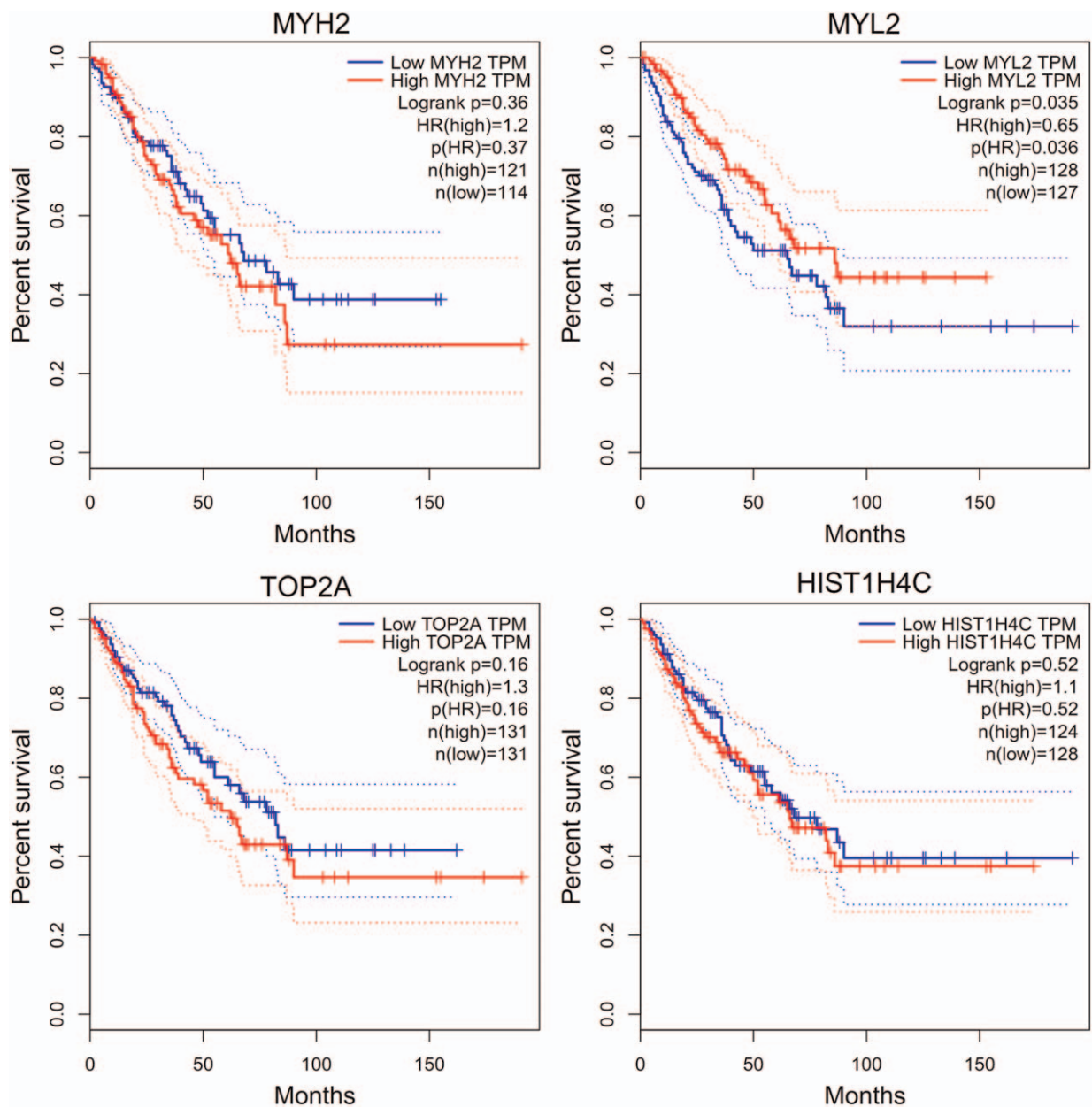**Figure 6.** Verification for the role of hub genes on the overall survival with the GEPIA. GEPIA = Gene Expression Profiling Interactive Analysis, MYH2 = Myosin Heavy Chain 2, MYL2 = Myosin Light Chain 2, TOP2A = DNA Topoisomerase II Alpha.

nism of RMS is very important for the research of targeted drugs. In this study, bioinformatics techniques were used to analyze RMS and normal tissues to screen differentially expressed genes. The main result of this study was that the expression of MYL2 gene was lower in RMS than the normal tissues. Patients with RMS have a poor prognosis when MYL2 is poorly expressed.

MYL2 is a member of the myosin light chain family. As a regulatory light chain, MYL2 is a small molecule protein composed of 167 amino acids located on chromosome 17.[8] Different types of muscle fibers contain different types of myosin light chain isomers. In vertebrate rhabdomyosin, the genes encoding the light chain subunits of rhabdomyosin include MYLPF, MYL7, and MYL2. Early studies found that MYL2

gene is mainly concentrated in muscle tissues, and the mutation of this gene can lead to muscle R cell proliferation. Earlier studies found that MYL2 gene is mainly concentrated in muscle tissue, and the mutation of this gene can lead to muscle cell proliferation.[9] In patients with muscle-related diseases, the expression of MYL2 gene is significantly lower than that of normal people, while the expression of MYL2 gene is significantly lower in patients with severe diseases than that in patients with mild diseases.[10] Recent studies have shown that phosphorylation of MYL2 in adult muscle cells causes muscle dilation,[11] which was related with numerous muscle-related diseases.[12] This suggests that MYL2 plays an important role in the normal maintenance of muscle cells. One study has shown

that MYL2 gene is expressed in mature mouse muscle tissue but not in the differentiation of C2C12 myoblasts cultured in vitro.[7] When the gene was knocked out of the muscle cells of mice, their embryonic or primary growth was severely abnormal.[13] In fibroblasts, the integrity of muscle fiber type mainly be maintained by MYL2 genes. MYL2 plays a role of regulating the activity of muscle fibers, transformation in the muscle fiber type, and affects the growth of muscle.[14] Therefore, it is speculated that MYL2 gene plays an important role in muscle growth and development, and may be a candidate gene for regulating muscle development.

One study has shown that MYL2 promoter is a muscle-specific promoter, and multiple binding sites that can promote the proliferation and differentiation of muscle cells are predicted in the upstream of the initial codon.[15] And MYL2 gene is speculated to be involved in the regulation of the growth and differentiation of skeletal muscle cells. The researchers cloned the MYL2 gene, a promoter specifically expressed in muscle.[16] The muscle specificity of the promoter was determined by dual-luciferase assay, and it was found that the MYLP2 promoter contained the MEF2 binding site near the starting site, which could activate the expression of the chloramphenicol acetyl-transferase reporter gene, and played a very important role in zebrafish muscle cells.[17] In addition, 1 study has shown that goat MYL2 protein is specifically expressed in goat skeletal muscle fibers, mainly in type I muscle fibers (slow contraction muscle fibers).[18] MYL2 gene is closely related to the growth and development of muscle and plays a crucial role in the early growth and differentiation of muscle.[7]

MYL2 gene, as a member of the myosin light chain family, is a regulatory light chain. Early study found that the fundamental light chain plays an important role in the maintenance of the heavy chain configuration, while the regulatory light chain plays a regulatory role in the activity of muscle fibers.[19] Therefore, it is believed that the myosin light chain plays an important role in the type and growth of muscle fibers. MYL2 is a regulatory light chain that regulates muscle fiber activity.[19] Then it affects the activity of ATPase on the heavy chain of myosin and transforms the types of muscle fiber, and regulates the growth of muscle.[20,21] MYL2 may play an important role in mammalian muscle growth and development. MYL2 gene was found to be expressed in skeletal muscle of mature mice, but not in the differentiation of myoblasts in vitro MYL20.[13] When the MYL2 gene was knocked out of the heart muscle and skeletal muscle, the mice developed severe abnormalities in embryonic or primary growth.[14,22] MYL2 plays an important role in the maintenance of muscle fiber integrity, cell structure, cell adhesion, cell migration, and cell division.[23] Obtaining the length cDNA sequence of this gene is of great significance for further study of gene function.

The above literature review is consistent with our results. MYL2 gene expression is reduced in muscle-related diseases. Similarly, MYL2 gene is low expressed in RMS, and the normal development and growth of muscles are restricted at this time, and the proliferation and differentiation of muscle cells are also affected accordingly, eventually leading to the occurrence of tumors.

RMS is a significant harm to child injuries, and the discovery of targeted therapy is essential for treatment. The exploration of biomarkers related to the prognosis of RMS is expected to be the research direction of targeted therapy and provide a new direction for the precise treatment of RMS. In addition, as a

biomarker of poor prognosis, MYL2 can also guide the timely selection of effective clinical management measures to reduce the mortality of RMS.

5. Limitation

Despite the rigorous bioinformatics analysis in this paper, there are still some deficiencies. In this study, MYL2 excavated by bioinformatics have not been verified at the cell and animal levels. Overexpression or knockout of MYL2 were conducted to further verify its function.[24] Therefore, in the future research, we should carry out in-depth exploration in this aspect.

6. Conclusions

In summary, lower expression of MYL2 was 1 prediction for the poor prognosis of patients with RMS. The exploration of MYL2 related to the prognosis of RMS is expected to be the research direction of targeted therapy and provide a new direction for the precise treatment of RMS. In addition, as a biomarker of poor prognosis, MYL2 may also guide the timely selection of effective clinical management measures to reduce the mortality of RMS. MYL2 might provide novel research idea for the diagnosis and treatment of RMS.

Author contributions

Conceptualization: Meng-jie Shan.
Formal analysis: Junning Wang, Shang Gao
Investigation: Junning Wang, Shang Gao
Methodology: Junning Wang, Meng-jie Shan.
Project administration: Peiyuan Guo.
Software: Keqin Dong.
Supervision: Shang Gao.
Validation: Keqin Dong, Peiyuan Guo.
Visualization: Peiyuan Guo.
Writing – original draft: Junning Wang.
Writing – review & editing: Shang Gao, Meng-jie Shan.

References

[1] Parham DM, Barr FG. Classification of rhabdomyosarcoma and its molecular basis. Adv Anat Pathol 2013;20:387–97.
[2] Sun X, Guo W, Shen JK, Mankin HJ, Hornicek FJ, Duan Z. Rhabdomyosarcoma: advances in molecular and cellular biology. Sarcoma 2015;2015:232010.
[3] Raney B, Huh W, Hawkins D, et al. Outcome of patients with localized orbital sarcoma who relapsed following treatment on Intergroup Rhabdomyosarcoma Study Group (IRSG) Protocols-III and -IV, 1984–1997: a report from the Children's Oncology Group. Pediatr Blood Cancer 2013;60:371–6.
[4] Leipzig J. A review of bioinformatic pipeline frameworks. Brief Bioinform 2017;18:530–6.
[5] Parad JE, Liao W. Bioinformatic applications in psoriasis: genetics, transcriptomics, and microbiomics. Semin Cutan Med Surg 2019;38:E3–3E11.
[6] Behzadi P, Ranjbar R. DNA microarray technology and bioinformatic web services. Acta Microbiol Immunol Hung 2019;66:19–30.
[7] Sheikh F, Lyon RC, Chen J. Functions of myosin light chain-2 (MYL2) in cardiac muscle and disease. Gene 2015;569:14–20.
[8] Kang PB, Griggs RC. Advances in muscular dystrophies. JAMA Neurol 2015;72:741–2.
[9] Weterman MA, Barth PG, van Spaendonck-Zwarts KY, et al. Recessive MYL2 mutations cause infantile type I muscle fibre disease and cardiomyopathy. Brain 2013;136:282–93.
[10] Szczesna D, Ghosh D, Li Q, et al. Familial hypertrophic cardiomyopathy mutations in the regulatory light chains of myosin affect their structure, Ca2+ binding, and phosphorylation. J Biol Chem 2001;276:7086–92.

[11] Li Y, Wu G, Tang Q, et al. Slow cardiac myosin regulatory light chain 2 (MYL2) was down-expressed in chronic heart failure patients. Clin Cardiol 2011;34:30–4.

[12] Huang W, Liang J, Yuan CC, et al. Novel familial dilated cardiomyopathy mutation in MYL2 affects the structure and function of myosin regulatory light chain. FEBS J 2015;282:2379–93.

[13] Sutherland CJ, Esser KA, Elsom VL, Gordon ML, Hardeman EC. Identification of a program of contractile protein gene expression initiated upon skeletal muscle differentiation. Dev Dyn 1993;196:25–36.

[14] Chen J, Kubalak SW, Minamisawa S, et al. Selective requirement of myosin light chain 2v in embryonic heart function. J Biol Chem 1998;273:1252–6.

[15] Shimizu A, Shimizu N. Dual promoter expression system with insulator ensures a stringent tissue-specific regulation of two reporter genes in the transgenic fish. Transgenic Res 2013;22:435–44.

[16] Bottje W, Kong BW, Reverter A, Waardenberg AJ, Lassiter K, Hudson NJ. Progesterone signalling in broiler skeletal muscle is associated with divergent feed efficiency. BMC Syst Biol 2017;11:29.

[17] Garcia de la Serrana D, Wreggelsworth K, Johnston IA. Duplication of a single myhz1.1 gene facilitated the ability of goldfish (Carassius auratus) to alter fast muscle contractile properties with seasonal temperature change. Front Physiol 2018;9:1724.

[18] Liu M, Wei Y, Li X, et al. Quantitative phosphoproteomic analysis of caprine muscle with high and low meat quality. Meat Sci 2018;141:103–11.

[19] Zhang SZ, Xu Y, Xie HQ, Li XQ, Wei YQ, Yang ZM. The possible role of myosin light chain in myoblast proliferation. Biol Res 2009;42:121–32.

[20] Geeves MA, Fedorov R, Manstein DJ. Molecular mechanism of actomyosin-based motility. Cell Mol Life Sci 2005;62:1462–77.

[21] Vicente-Manzanares M, Ma X, Adelstein RS, Horwitz AR. Non-muscle myosin II takes centre stage in cell adhesion and migration. Nat Rev Mol Cell Biol 2009;10:778–90.

[22] Wang Y, Szczesna-Cordary D, Craig R, et al. Fast skeletal muscle regulatory light chain is required for fast and slow skeletal muscle development. FASEB J 2007;21:2205–14.

[23] Park I, Han C, Jin S, et al. Myosin regulatory light chains are required to maintain the stability of myosin II and cellular integrity. Biochem J 2011;434:171–80.

[24] Meng LB, Shan MJ, Qiu Y, et al. TPM2 as a potential predictive biomarker for atherosclerosis. Aging (Albany NY) 2019;11:6960–82.