Research Article

Machine-Learning-Based m5C Score for the Prognosis Diagnosis of Osteosarcoma

Haijie Zhang,[1,2] Peipei Xu,[2] and Yichang Song [1]

[1]*The Eighth Affiliated Hospital of Sun Yat-Sen University, Shenzhen 518033, China*
[2]*PET/CT Center, The First Affiliated Hospital of Shenzhen University, Shenzhen 518000, China*

Correspondence should be addressed to Yichang Song; gksychang@csu.edu.cn

Background. Osteosarcoma is a common and highly metastatic malignant tumor, and m5C RNA methylation regulates various biological processes. The purpose of this study was to explore the prognostic role of m5C in osteosarcoma using machine learning. *Methods*. Osteosarcoma gene data and the corresponding clinical information were downloaded from the GEO database. Machine learning methods were used to screen m5C-related genes and construct m5C scores. In addition, the clusterProfiler package was used to predict the m5C-related functional pathways. xCell and CIBERSORT were used to calculate the immune microenvironment cells. GSVA was applied to analyze different categories of m5C genes, and the correlation between the GSVA and m5C scores was evaluated. *Results*. Twenty m5C genes were identified, and 54 related genes were screened. The m5C score was constructed based on the PCA score. With an increase in the m5C score, the expression of m5C genes and their related genes changed. Functional analysis indicated that the focal adhesion, cell-substrate adherens junction, cell adhesion molecule binding, and E2F targets might change with the m5C score. The naive B cells and CD4[+] memory T cell also changed with the m5C score. The results of the correlation analysis showed that the m5C score was significantly correlated with the reader and eraser genes. *Conclusion*. The m5C score might be a prognostic index for osteosarcoma.

1. Introduction

Osteosarcoma is a common primary bone malignancy with a high rate of incidence in children and adolescents [1, 2]. Osteosarcoma is highly aggressive, is metastatic, and has a high risk of recurrence after treatment [3]. Currently, the main treatment methods for patients with osteosarcoma include surgery, radiotherapy, chemotherapy, and combination therapies [4, 5]. Osteosarcoma is thought to arise from osteoblasts in rapidly growing bones [6]. Changes in cellular heterogeneity and immune dynamics lead to complex molecular and genetic mechanisms that make conventional treatment regimens less effective in a clinical setting [7, 8]. Studies have shown that the estimated survival rate of patients with metastatic osteosarcoma undergoing routine treatment is less than 5 years [9]. Therefore, a clear diagnosis and precisely targeted therapy are significant for patients with osteosarcoma. The new generations of

sequencing technology and data analysis methods provide an efficient and convenient technical auxiliary means for the exploration of therapeutic targets for osteosarcoma.

As a branch of computer science and statistics [10], machine learning generates predictive models mainly by learning from training data. Its application in the field of medicine has attracted significant attention, including the diagnosis, prognosis, and treatment cycle prediction of the disease. The literature has shown that machine learning methods have been applied in drug discovery [11], the management of hematologic malignancies [12], and epilepsy pathology monitoring [13].

RNA methylation is involved in the regulation of various biological processes, and its dysregulation is closely related to the occurrence of human malignant tumors [14]. In addition, 5-methylcytosine (m5C) is a type of RNA methylation modification located in the untranslated region of the mRNA transcript [15]. M5C is involved in various RNA

metabolisms, including mRNA output, RNA stabilization, and translation, through the dynamic regulation of a series of important mediator proteins (writer, eraser, and reader) [16]. Studies have shown that m5C plays an important role in the diagnosis and prognosis of cancer, including hepatocellular carcinoma [17], squamous cell carcinoma of the head and neck [18], lung adenocarcinoma [19], triple-negative breast cancer [20], and gastrointestinal cancer [21]. However, the role of m5C in osteosarcoma remains unclear.

The prognostic role of the m5C score in osteosarcoma is currently unknown. In this study, we analyzed the data in GEO. A machine learning method was used to screen and construct the m5C score. Related functional pathways were predicted. The correlation between clinical factors, m5C scores, and the prognosis was analyzed. These results provide a scientific basis for using the m5C score as a prognostic indicator for patients with osteosarcoma.

2. Methods

2.1. Datasets and Data Preprocessing. The GSE21257 and GSE39058 gene expression profile and corresponding clinical information were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/gds/), with the GSE39058 dataset as the validation set. The microarray data of the GPL10295 platform (Illumina human-6 v2.0 expression beadchip) include RNA expression and clinical characteristics (e.g., age, sex, grade, metastasis, and pathology) from 53 patients with osteosarcoma. Microarray data probes were annotated based on the gene names, and the mean value of all probes with the same gene name was obtained. All data were then normalized.

2.2. Establishment of the m5C Score. The 21 m5C genes, including the synonym of these genes, were intersected with the osteosarcoma dataset (GPL10295), leaving 20 genes (i.e., DNMT1, DNMT3A, DNMT3B, MBD1, MBD2, MBD3, MBD4, MECP2, NEIL1, NTHL1, SMUG1, TDG, UHRF1, UHRF2, UNG, ZBTB33, ZBTB4, TET1, TET2, and TET3). Next, a correlation analysis was conducted among these 20 genes, and genes with related coefficient >0.6, with $p < 0.05$, were considered as related genes. The univariate analysis was used for the related gene dimensionality reduction. After the random survival forest and PCA were performed, the risk score model was constructed as follows: m5C score $= \sum PC1_A - \sum PC1_B$, where A represents the gene with HR > 1 and B represents the gene with HR < 1. The m5C score was verified in GSE21257 and GSE39058. ROC curves were performed to diagnose the effectiveness of the model. After three main classifications (reader, eraser, and writer) of the m5C gene and the complete gene set were analyzed using GSVA, the correlation analysis was performed between them and the m5C score.

2.3. Functional Annotation Analysis. An analysis of the difference between the two groups with high- and low-risk scores in the expression matrix was conducted, and the standard was set as abs (logFC) > log2 (1.5), with p value <0.05. A gene ontology (GO) analysis was conducted for these differentially expressed genes, including cellular component (CC), molecular function (MF), and biological process (BP) analyses. The limit for a significant enrichment was set at $p < 0.05$. For the correlation analysis of all genes in the expression matrix of this gene, the R package clusterProfiler [22] was used to conduct a gene set enrichment analysis (GSEA). The criteria of statistical significance were set as |NES| >1, with NOM p value <0.05 and an FDR q value < 0.25. The clusterProfiler package was then applied to evaluate and analyze the GO biological process, KEGG pathway, and HALLMARK pathway of the genes.

2.4. Immune Cell Infiltration Analysis. In this study, xCell [23] and CIBERSORT were used to calculate immune microenvironment cells. The method integrates the benefits of a gene enrichment analysis through a deconvolution to evaluate 64 cell types involving multiple adaptive and innate immune cells, hematopoietic progenitor cells, epithelial cells, and extracellular matrix cells. These included 48 microenvironment-related tumor cells.

2.5. Statistical Analysis. To analyze differentially expressed genes, we used the Benjamini–Hochberg method, which converts p values into FDR to identify important genes. All survival curves were generated and visualized using the R package survminer and a Kaplan–Meier analysis. The statistical significance of the differences was determined using the log-rank test in each dataset. A pheatmap was adopted to generate heatmaps. R (https://www.r-project.org/, version 3.6.1) was used to conduct a statistical analysis. The data were visualized using the R package ggplot2. After the Shapiro–Wilk normality test was used to check the normality of the variables, the differences between the two groups were compared using unpaired Student's t-test with the normally distributed variables, and the Wilcoxon test was conducted to compare the variables without normally distributed. The statistical significance was set at $p < 0.05$.

3. Results

3.1. Establishment of the m5C Score for Osteosarcoma. To establish the m5C score, we obtained m5C gene data from the GEO database and screened 20 m5C genes. Their survival analysis in GSE21257 was conducted independently (Figures S1 and S2). Correlation analysis among the 20 m5C genes in the osteosarcoma dataset indicated a certain amount of interrelationship among the m5C genes (Figure 1(a)). Based on 20 m5C genes, a gene set containing 3,008 genes was constructed (Figure 1(b)). A total of 1,929 genes were screened out using univariate analysis, and 54 genes were identified in the random survival forest (Figures 1(c) and 1(d)). After a univariate analysis (Figure S3) and PCA scoring of these genes, the m5C score was calculated. A survival analysis showed that patients with a high-risk score had a poor prognosis (p value less than 0.05) (Figure 1(e)). The risk score was verified in the

(a)

(b)

(c)

(d)

(e)

(f)

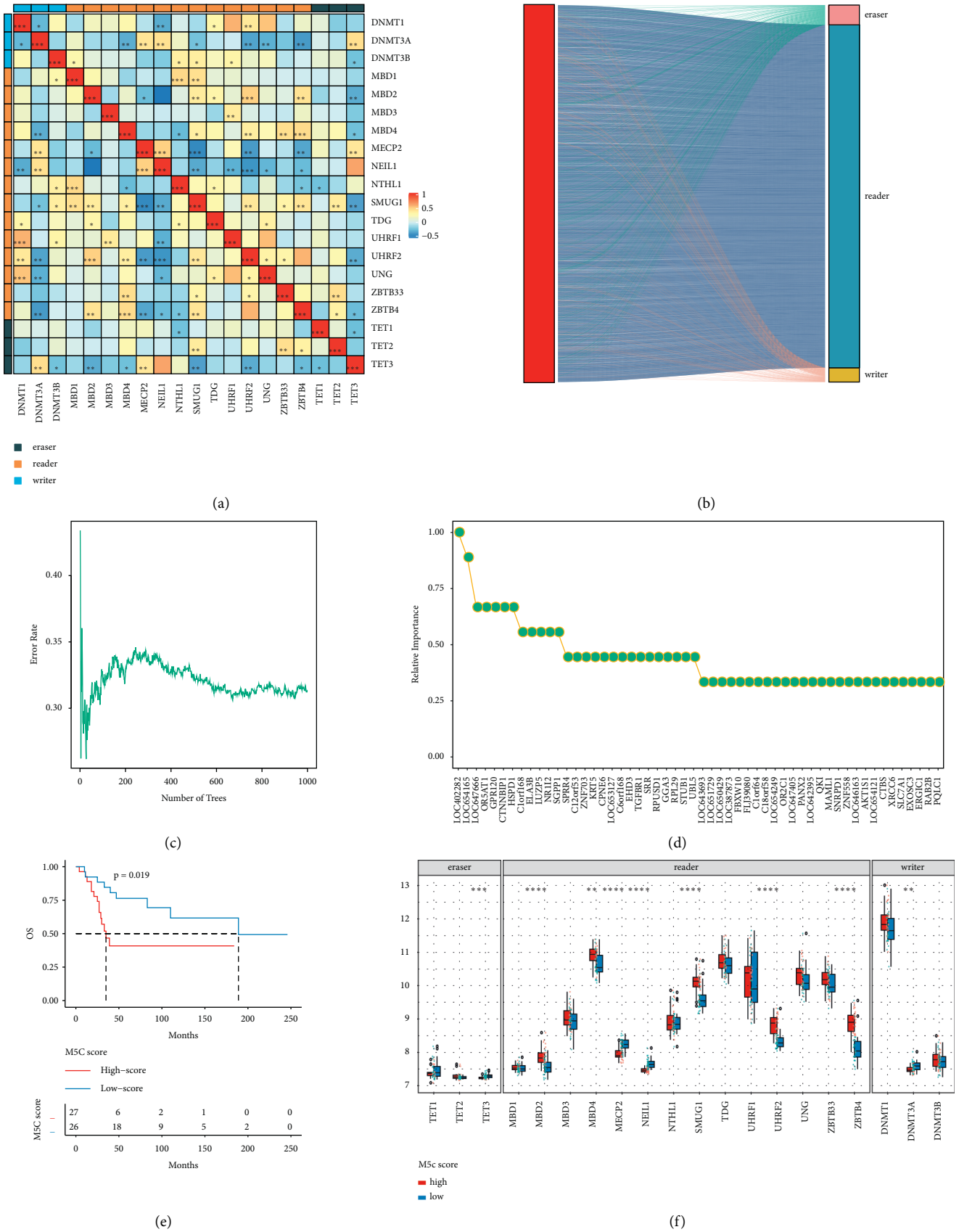FIGURE 1: Construction of the m5C risk model. (a) Correlation analysis results of m5C genes. (b) The relationship between the m5C gene and its related genes. (c) Error rate as a function of the classification tree. (d) Related importance values for the genes. (e) Overall survival in the high and low m5C score group. (f) The expression level of the m5C gene under high and low m5C scores. * indicates significance with $p < 0.05$.

GSE39058 dataset (Figure S4). ROC curves were performed to diagnose the effectiveness of the model. The AUC value of our model was higher than that of others [24] (Figure S5). The genes were divided into high and low groups based on their risk scores. The expression of m5C differed, including TET3, MBD2, UHRF2, MECP2, ZBTB4, and SMUG1 (Figure 1(f)). With an increase in the risk score, the expression of the m5C gene changed, including DNMT3A, MBD2, MECP2, NEIL1, and SMUG1 (Figure 2(a)). As the risk score increased, the expression levels of the constructed PCA score genes changed, including C1orf64, FBXW10, LOC647405, OR2C1, AKT1S1, XRCC6, EXOSC3, and SNRPD1 (Figure 2(b)).

3.2. M5C Score-Related Functional Analysis of Osteosarcoma. Next, we assessed the expression of genes in the high and low groups based on the m5C scores (Figure 3(a)). There were conspicuous differences in the gene expression between the high and low m5C score groups. A functional analysis was conducted on the differentially expressed genes, including BP, CC, and MF (Figure 3(b)). Among them, with BP, the regulation of leukocyte activation and translational initiation was significant. With CC, the focal adhesion, cell-substrate adherens junction, cell-substrate junction, and adherens junction were significant. With MF, the cell adhesion of the molecule binding, ISG15 transferase activity, and kinase regulator activity were significant. Meanwhile, we conducted a GSEA based on the m5C score, including GO enrichment (Figure 4(a)), KEGG pathway (Figure 4(b)), and HALL-MARK (Figure 4(c)) analyses. The GO analysis results showed that the ribonucleoprotein complex biogenesis, translational initiation, and application of protein localization to endoplasmic reticulum pathways were significant. A KEGG analysis showed that the ribosome, starch and sucrose metabolism, and spliceosome pathways were significant. The results of the HALLMARK analysis showed that the MYC target V1, the G2M checkpoint, and the E2F target pathways were significant.

3.3. Immune Infiltration Analysis for the m5C Score. Next, xCell (Figure 5(a)) and CIBERSORT (Figure S6) were used to analyze the immune cell infiltration. As the m5C score increased, the expression of immune cell infiltrates changed, including the CLP, smooth muscle, hepatocytes, melanocytes, MSCs, and HSCs.

3.4. M5C Score as an Independent Prognostic Factor in Osteosarcoma Patients. To further validate the prognostic value of m5C scores, the clinical factors and m5C scores of patients with osteosarcoma were evaluated. The results of a univariate Cox regression analysis (Figure 5(b)) and a multivariate Cox regression analysis (Figure 5(c)) showed that the p values of the age, sex, and grade were all greater than 0.05. These findings suggested that these clinical factors in patients are indistinctively associated with the prognosis. The m5C score was significantly correlated with patient prognosis, with $p = 0.004$ and $p = 0.036$, respectively.

Therefore, the m5C score might be an independent prognostic factor in patients with osteosarcoma.

3.5. Relationship between the m5C Score and m5C in Osteosarcoma. To further explore the relationship between the m5C gene score and m5C score, we conducted a GSVA on the three major classifications of the m5C gene (reader, eraser, and writer) and the complete m5C gene set. A correlation analysis was applied using the m5C score and the above reader (Figure 6(a)), eraser (Figure 6(b)), and writer (Figure 6(c)) GSVA scores, as well as the GSVA score of the m5C gene set (Figure 6(d)). The GSVA score of the reader was positively correlated with the m5C score, whereas the eraser score was negatively correlated. There was no significant correlation between the GSVA score of the writer and m5C scores. The m5C gene set score was positively correlated with m5C. Thus, there might be a relationship between the m5C scores and m5C genes in osteosarcoma.

4. Discussion

Although some studies have suggested that an m5C RNA modification is associated with the genesis and progression of cancer [25], the underlying relationship between osteosarcoma and m5C remains unclear. In this study, we established a risk model based on the m5C scores. Compared with other models [24], the m5C score has higher diagnostic efficiency. Furthermore, the m5C score can be used as an independent prognostic factor in patients with osteosarcoma.

The m5C score, including C1orf64, AKT1S1, and XRCC6, was constructed in this study. Among them, C1orf64 can be used as a protective factor in patients with osteosarcoma, and its expression decreases with an increase in the m5C score. SRARP (C1orf64) is a tumor suppressor that can be used to predict the clinical outcomes of malignant tumors [26]. In addition, AKT1S1 and XRCC6 were identified as risk factors. A high expression of AKT1S1 is positively correlated with a poor prognosis in liver cancer patients [27]. In addition, XRCC6 is involved in the poor prognosis of human osteosarcoma cells [28] and prostate cancer [29]. All of these demonstrated the reliability of predicting the prognosis of patients based on their m5C score. Regardless, how the genes constructing the m5C score regulate the development of osteosarcoma still requires further research and exploration.

M5C regulates the transcriptome expression mainly through the dynamic regulation of methyltransferases (writers), binding proteins (readers), and demethylases (eraser) [30]. The eraser, reader, and writer can change significantly with the change in the m5C score, including MBD2, UHRF2, ZBTB4, and TET3. Among them, MBD2 inhibits DNA methylation and activates prometastatic genes [31, 32]. In 5-azaCdR-induced breast cancer, MBD2 depletion can inhibit cell invasion [33]. It has been reported that UHRF2 was associated with cell invasion, migration, and lymphatic metastasis of intrahepatic cholangiocarcinoma [34]. TET3 inhibits ovarian cancer by
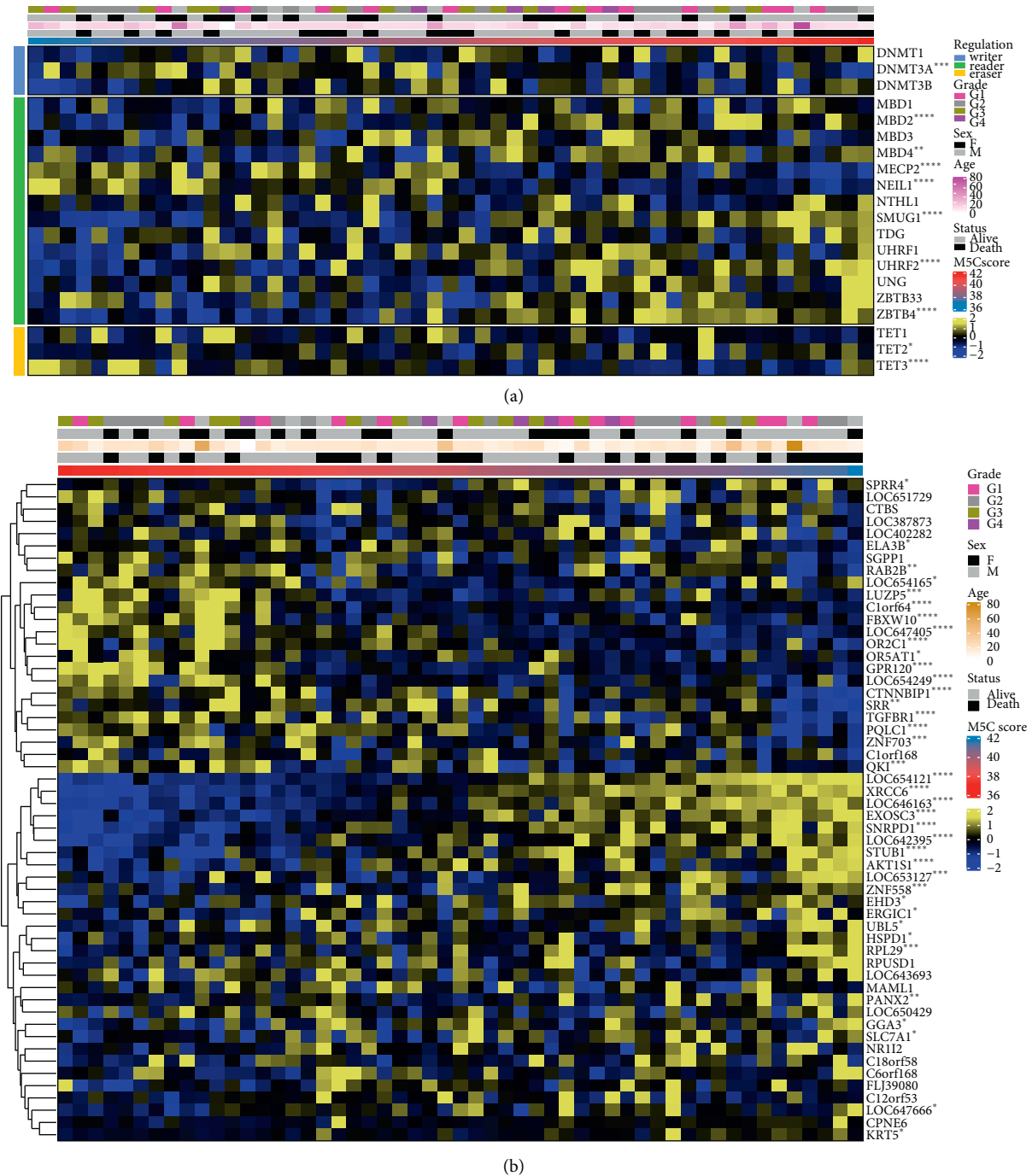
(a)

(b)

Figure 2: M5C score and gene expression. (a) The expression of m5C genes. (b) The expression of constructing PCA score genes.

blocking TGF-β1-induced EMT [35]. These results suggested that MBD2, UHRF2, and TET3 may be involved in the process of tumor metastasis. Our functional analysis also demonstrates this point. Focal adhesion was significantly increased in terms of the m5C score and high/low m5C score groups, including the focal adhesion, cell-substrate adherens junction, cell adhesion molecule binding, and E2F targets. There is a relationship between cell adhesion and cancer metabolism [36], and it has been shown in the literature that cell adhesion molecules play an important role in the

development of cancer and are clinical markers for the effectiveness of cancer therapy [37]. E2F dysfunction may contribute to cancer development [38]. Regulation of the PRMT5-E2F1 axis can promote the migration and invasion of tumor cells [39]. Thus, m5C might influence cancer development by regulating the functional downstream pathways. These results further proved that the m5C score has a certain prognostic role in patients with osteosarcoma.

However, in our analysis, the expression of ZBTB4 increased with an increase in the m5C score. In colorectal

(a)

(b)
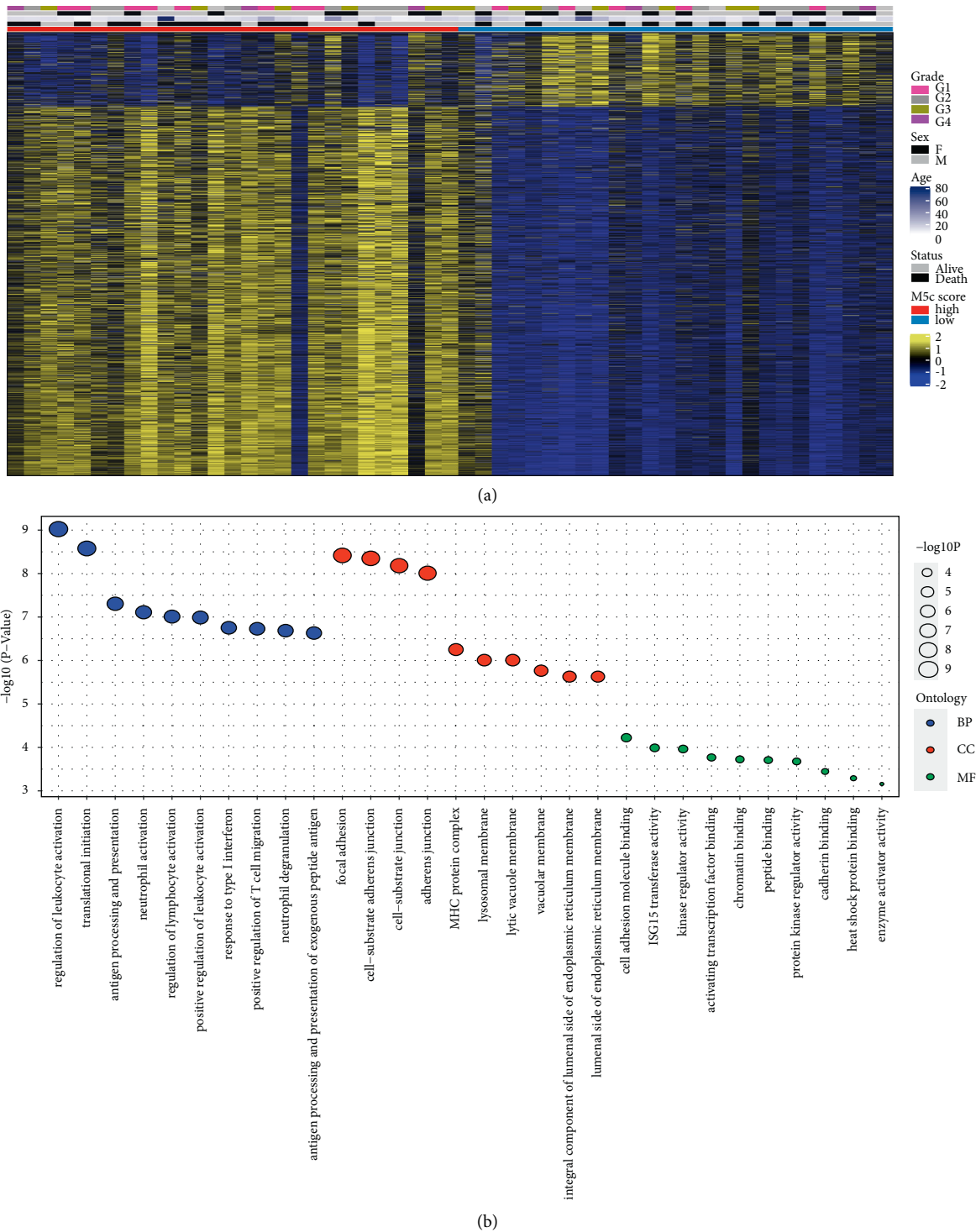
Figure 3: Differential genes and functional analysis between high and low m5C score groups. (a) The expression of differential genes. (b) Functional analysis of differential genes.

cancer, patients with a high ZBTB4 expression have a better prognosis [40]. This is inconsistent with our analysis. We speculated that this might be due to the complexity of the biological functions of cells and the complexity of the microenvironment of a tumor.

Tumor development is closely related to immune invasion [41, 42]. With the change in the m5C score, the immune cells also changed significantly, including naive B cells and CD4[+] memory T cells. Naive B cells are involved in the prognosis of hepatocellular carcinoma and colorectal

Pathway	Gene ranks	NES	pval	padj
GO_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS		1.66	1.1e−04	1.0e−01
GO_TRANSLATIONAL_INITIATION		1.80	1.1e−04	1.0e−01
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM		1.75	1.2e−04	1.0e−01
GO_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE		1.80	1.2e−04	1.0e−01
GO_CHROMATIN_BINDING		1.48	2.0e−04	1.0e−01
GO_OXIDOREDUCTASE_ACTIVITY_ACTING_ON_PEROXIDE_AS_ACCEPTOR		−2.03	3.6e−04	1.0e−01
GO_CHEMOKINE_ACTIVITY		−1.98	3.5e−04	1.0e−01
GO_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_IMMUNOGLOBULIN		−1.88	3.5e−04	1.0e−01
GO_REGULATION_OF_VIRAL_ENTRY_INTO_HOST_CELL		−1.88	2.9e−04	1.0e−01
GO_HYDROGEN_PEROXIDE_CATABOLIC_PROCESS		−1.90	2.9e−04	1.0e−01

0   5000 10000 15000 20000

(a)

Pathway	Gene ranks	NES	pval	padj
KEGG_RIBOSOME		1.73	5.0e−04	1.3e−02
KEGG_STARCH_AND_SUCROSE_METABOLISM		1.54	1.0e−02	8.7e−02
KEGG_SPLICEOSOME		1.48	1.3e−02	1.0e−01
KEGG_CITRATE_CYCLE_TCA_CYCLE		1.45	4.8e−02	2.6e−01
KEGG_CELL_CYCLE		1.34	4.8e−02	2.6e−01
KEGG_VIRAL_MYOCARDITIS		−1.88	4.4e−04	1.3e−02
KEGG_LEISHMANIA_INFECTION		−1.93	4.3e−04	1.3e−02
KEGG_AUTOIMMUNE_THYROID_DISEASE		−1.81	3.7e−04	1.3e−02
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION		−1.91	3.6e−04	1.3e−02
KEGG_ASTHMA		−1.87	3.0e−04	1.3e−02

0   5000 10000 15000 20000

(b)

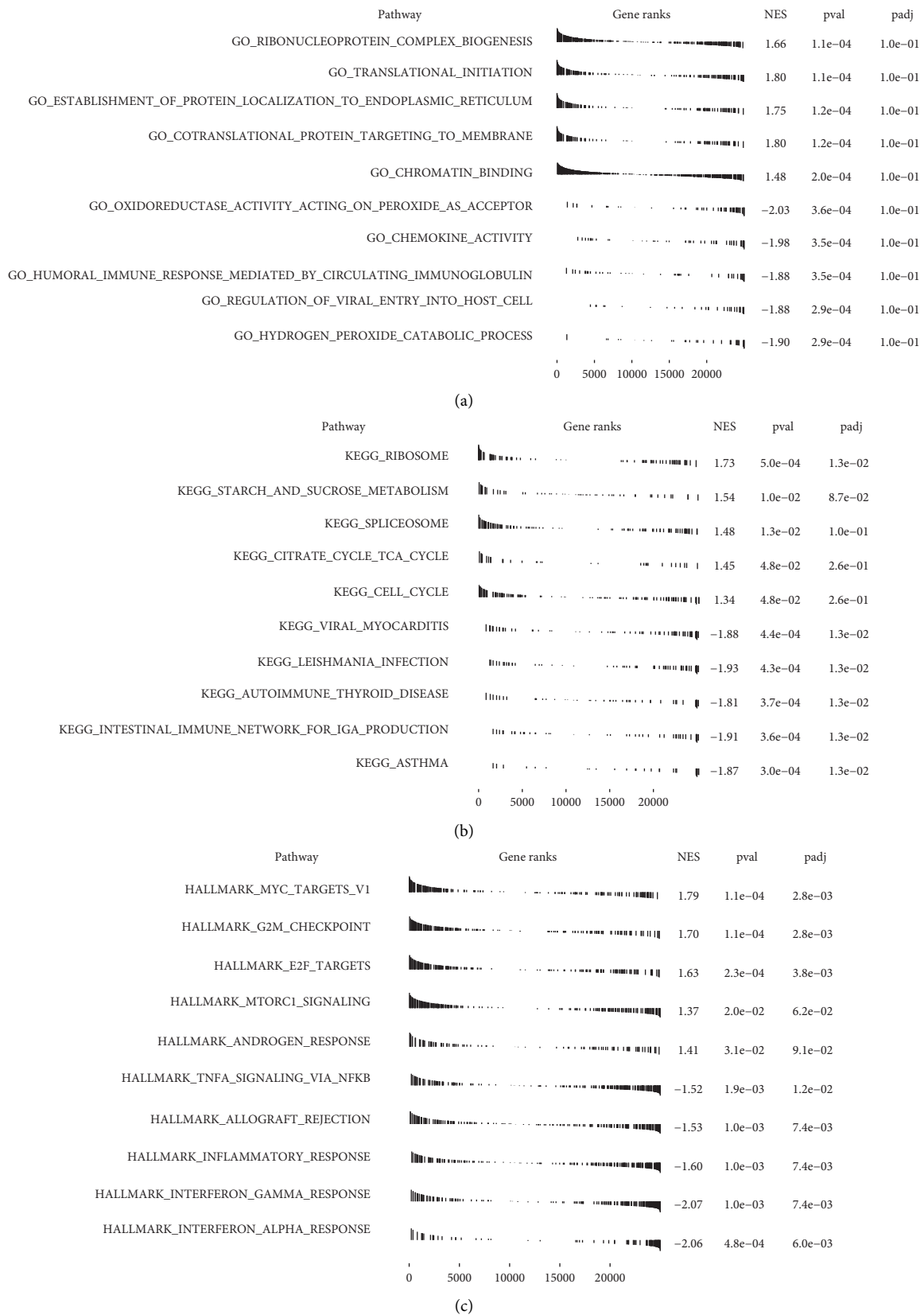| Pathway	Gene ranks	NES	pval	padj
HALLMARK_MYC_TARGETS_V1		1.79	1.1e−04	2.8e−03
HALLMARK_G2M_CHECKPOINT		1.70	1.1e−04	2.8e−03
HALLMARK_E2F_TARGETS		1.63	2.3e−04	3.8e−03
HALLMARK_MTORC1_SIGNALING		1.37	2.0e−02	6.2e−02
HALLMARK_ANDROGEN_RESPONSE		1.41	3.1e−02	9.1e−02
HALLMARK_TNFA_SIGNALING_VIA_NFKB		−1.52	1.9e−03	1.2e−02
HALLMARK_ALLOGRAFT_REJECTION		−1.53	1.0e−03	7.4e−03
HALLMARK_INFLAMMATORY_RESPONSE		−1.60	1.0e−03	7.4e−03
HALLMARK_INTERFERON_GAMMA_RESPONSE		−2.07	1.0e−03	7.4e−03
HALLMARK_INTERFERON_ALPHA_RESPONSE		−2.06	4.8e−04	6.0e−03

0   5000 10000 15000 20000

(c)

Figure 4: GSEA based on the m5C score in osteosarcoma. (a) GO enrichment analysis. (b) KEGG pathway analysis. (c) HALLMARK analysis.
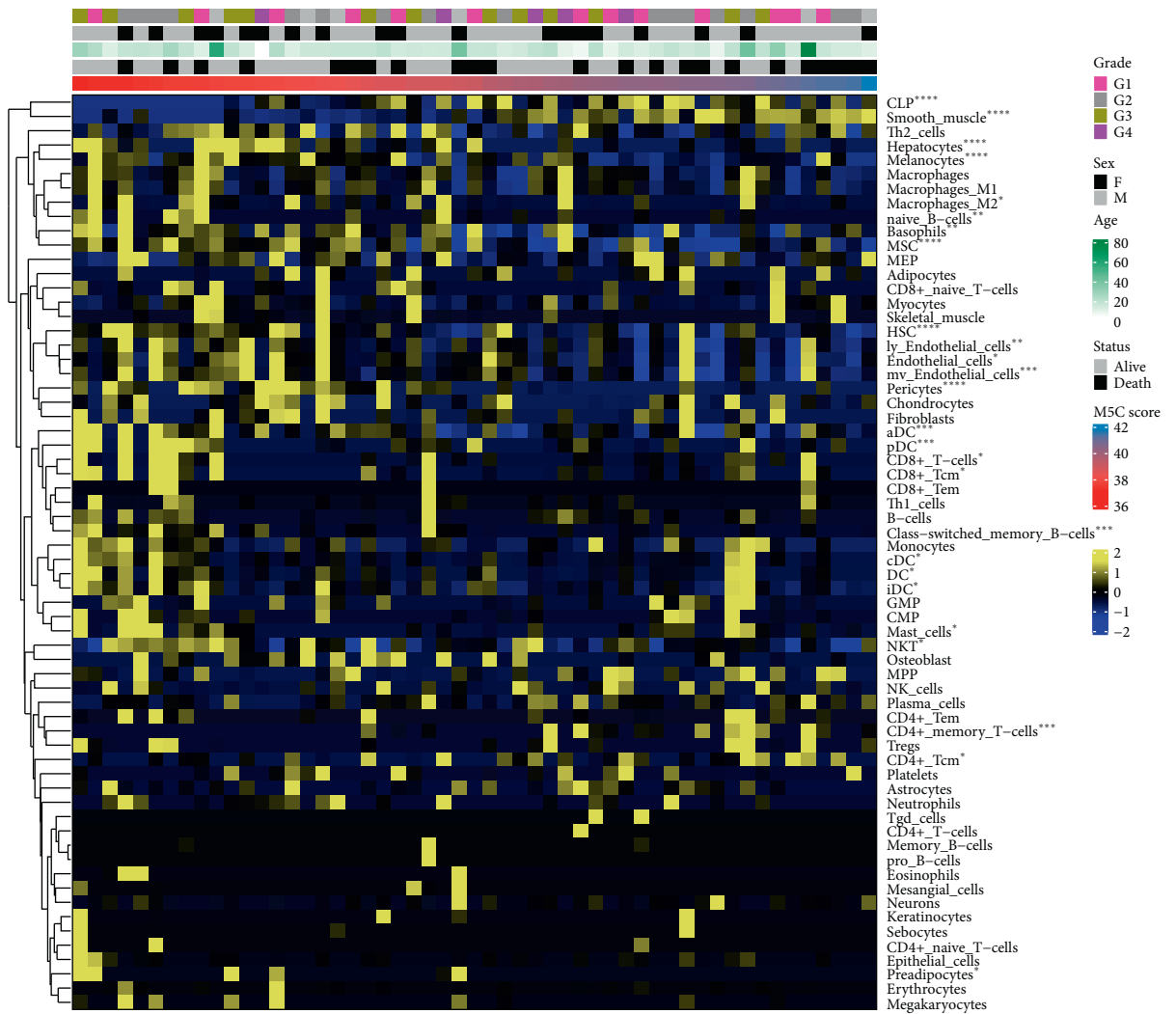
(a)
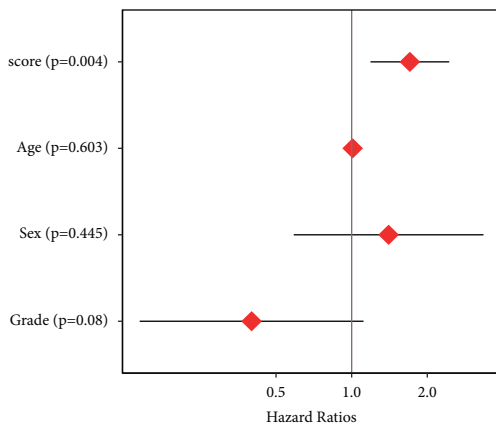
(b)
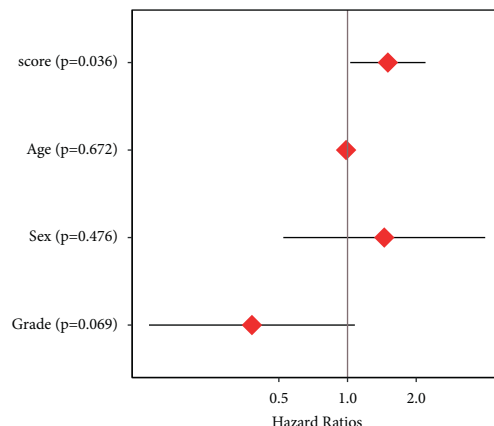
(c)

FIGURE 5: Immune cell infiltration analysis and prognosis analysis for the m5C score. (a) Immune cell infiltration analysis. (b) Univariate analysis. (c) Multivariate analysis.

cancer [43, 44]. A high abundance of B-cell infiltration could have a good prognostic effect. The study showed that CD4$^+$ follicular helper T-cell infiltration could be used to predict breast cancer survival [45]. The resistance of pancreatic cancer cells is also associated with changes in CD4$^+$ memory T cells [46]. Therefore, we speculated that m5C could affect
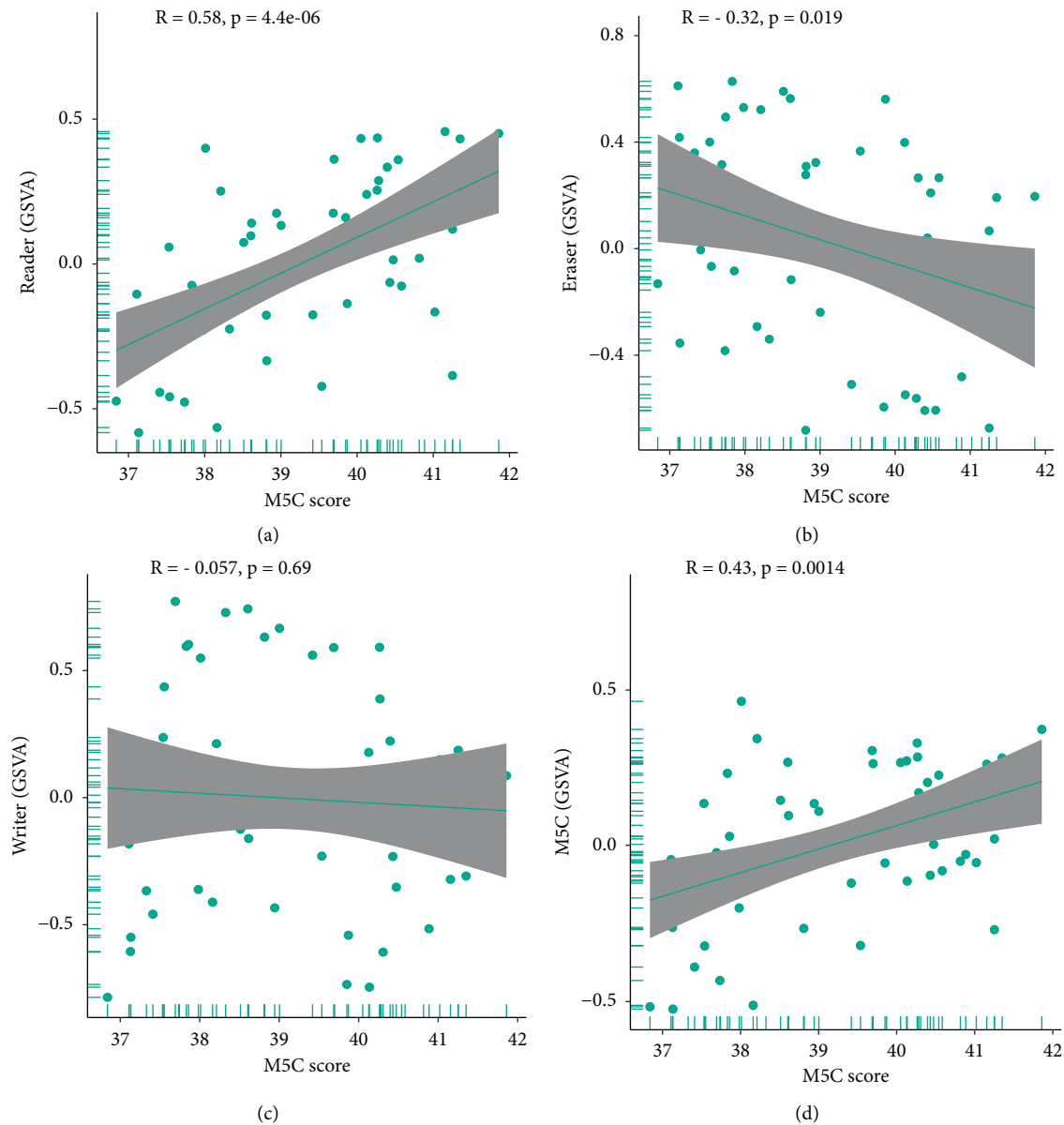
FIGURE 6: Correlation analysis between the m5C score and m5C gene. Correlation analysis of the m5C genes and the m5C scores, including (a) the reader score, (b) eraser score, (c) writer score, and (d) m5C gene set score.

the prognosis of osteosarcoma by regulating the infiltration of immune cells.

Combined with the above analysis, we believe that the m5C score can be used as a prognostic index for osteosarcoma. The specific mechanism by which m5C regulates the osteosarcoma prognosis requires further exploration and specific experimental verification.

5. Conclusion

In this study, patients with high m5C scores had a poor prognosis. The m5C score could influence cellular changes in the functional regulation and immune microenvironment related to osteosarcoma metastases. In conclusion, the m5C score could be used as an independent prognostic factor for the diagnosis of osteosarcoma.

Data Availability

The microarray data and clinical data we used in this research were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21257; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39058; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL10295).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Supplementary Materials

Figures S1 and S2: the survival analysis of DNMT1, DNMT3A, DNMT3B, MBD1, MBD2, MBD3, MBD4, MECP2, NEIL1, NTHL1, SMUG1, TDG, TET1, TET2, TET3, UHRF1, UHRF2, UNG, ZBTB4, and ZBTB33. Figure S3: the univariate analysis results of 54 genes. Univariate analysis showed that these 54 m5C-related genes might affect the survival time of the patients. Figure S4: the survival analysis of the m5C score in GSE39058. Figure S5: the ROC curves of our model (A) and others (B). Figure S6: the immune cell infiltration analysis by the CIBERSORT algorithm. (*Supplementary Materials*)

References

[1] B. Pingping, Z. Yuhong, L. Weiqi et al., "Incidence and mortality of sarcomas in shanghai, China, during 2002-2014," *Frontiers in Oncology*, vol. 9, p. 662, 2019.

[2] L. Mirabello, R. J. Troisi, and S. A. Savage, "Osteosarcoma incidence and survival rates from 1973 to 2004," *Cancer*, vol. 115, no. 7, pp. 1531–1543, 2009.

[3] F. Cersosimo, S. Lonardi, G. Bernardini et al., "Tumor-associated macrophages in osteosarcoma: from mechanisms to therapy," *International Journal of Molecular Sciences*, vol. 21, no. 15, 2020.

[4] I. F. Ciernik, A. Niemierko, D. C. Harmon et al., "Proton-based radiotherapy for unresectable or incompletely resected osteosarcoma," *Cancer*, vol. 117, no. 19, pp. 4522–4530, 2011.

[5] G. Rosen, B. Caparros, A. G. Huvos et al., "Preoperative chemotherapy for osteogenic sarcoma:Selection of postoperative adjuvant chemotherapy based on the response of the primary tumor to preoperative chemotherapy," *Cancer*, vol. 49, no. 6, pp. 1221–1230, 1982.

[6] A. J. Mutsaers and C. R. Walkley, "Cells of origin in osteosarcoma: mesenchymal stem cells or osteoblast committed cells?" *Bone*, vol. 62, pp. 56–63, 2014.

[7] Y. Suehara, D. Alex, A. Bowman et al., "Clinical genomic sequencing of pediatric and adult osteosarcoma reveals distinct molecular subsets with potentially targetable Alterations," *Clinical Cancer Research*, vol. 25, no. 21, pp. 6346–6356, 2019.

[8] D. Wang, X. Niu, Z. Wang et al., "Multiregion sequencing reveals the genetic heterogeneity and evolutionary history of osteosarcoma and matched pulmonary metastases," *Cancer Research*, vol. 79, no. 1, pp. 7–20, 2019.

[9] B. Zhang, Y. Zhang, R. Li, J. Li, X. Lu, and Y. Zhang, "The efficacy and safety comparison of first-line chemotherapeutic agents (high-dose methotrexate, doxorubicin, cisplatin, and ifosfamide) for osteosarcoma: a network meta-analysis," *Journal of Orthopaedic Surgery and Research*, vol. 15, no. 1, p. 51, 2020.

[10] Q. Bi, K. E. Goodman, J. Kaminsky, and J. Lessler, "What is machine learning? A primer for the epidemiologist," *American Journal of Epidemiology*, vol. 188, no. 12, pp. 2222–2239, 2019.

[11] L. Patel, T. Shukla, X. Huang, D. W. Ussery, and S. Wang, "Machine learning methods in drug discovery," *Molecules*, vol. 25, no. 22, 2020.

[12] N. Radakovich, M. Nagy, and A. Nazha, "Machine learning in haematological malignancies," *The Lancet Haematology*, vol. 7, no. 7, pp. e541–e550, 2020.

[13] B. Abbasi and D. M. Goldenholz, "Machine learning applications in epilepsy," *Epilepsia*, vol. 60, no. 10, pp. 2037–2047, 2019.

[14] X. Han, M. Wang, Y. L. Zhao, Y. Yang, and Y. G. Yang, "RNA methylations in human cancers," Semin Cancer Biol, 2020.

[15] T. Huang, W. Chen, J. Liu, N. Gu, and R. Zhang, "Genome-wide identification of mRNA 5-methylcytosine in mammals," *Nature Structural & Molecular Biology*, vol. 26, no. 5, pp. 380–388, 2019.

[16] Y. S. Chen, W. L. Yang, Y. L. Zhao, and Y. G. Yang, *Dynamic Transcriptomic M(5) C and its Regulatory Role in RNA Processing*, , p. e1639, Wiley Interdiscip Rev RNA, 2021.

[17] Y. He, X. Yu, J. Li, Q. Zhang, Q. Zheng, and W. Guo, "Role of m(5)C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma," *Am J Transl Res*, vol. 12, no. 3, pp. 912–922, 2020.

[18] M. Xue, Q. Shi, L. Zheng, Q. Li, L. Yang, and Y. Zhang, "Gene signatures of m5C regulators may predict prognoses of patients with head and neck squamous cell carcinoma," *American Journal of Tourism Research*, vol. 12, no. 10, pp. 6841–6852, 2020.

[19] L. Sun, W.-K. Liu, X.-W. Du et al., "Large-scale transcriptome analysis identified RNA methylation regulators as novel prognostic signatures for lung adenocarcinoma," *Annals of Translational Medicine*, vol. 8, no. 12, p. 751, 2020.

[20] Z. Huang, J. Pan, H. Wang et al., "Prognostic significance and tumor immune microenvironment heterogenicity of m5C RNA methylation regulators in triple-negative breast cancer," *Frontiers in Cell and Developmental Biology*, vol. 9, p. 657547, 2021.

[21] S. Xiang, Y. Ma, J. Shen et al., "m5C RNA methylation primarily affects the ErbB and PI3K-akt signaling pathways in gastrointestinal cancer," *Frontiers in Molecular Biosciences*, vol. 7, p. 599340, 2020.

[22] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.

[23] D. Aran, Z. Hu, and A. J. Butte, "xCell: digitally portraying the tissue cellular heterogeneity landscape," *Genome Biology*, vol. 18, no. 1, p. 220, 2017.

[24] K. W. Xiao, Z. B. Liu, Z. H. Zeng et al., "Construction and validation of a macrophage-associated risk model for predicting the prognosis of osteosarcoma," *Journal of oncology*, vol. 2021, p. 9967954, 2021.

[25] P. Nombela, B. Miguel-López, and S. Blanco, "The role of m(6)A, m(5)C and Ψ RNA modifications in cancer: novel therapeutic opportunities," *Molecular Cancer*, vol. 20, no. 1, 2021.

[26] A. Naderi, "SRARP and HSPB7 are epigenetically regulated gene pairs that function as tumor suppressors and predict clinical outcome in malignancies," *Molecular Oncology*, vol. 12, no. 5, pp. 724–755, 2018.

[27] Z. Qi, T. Zhang, L. Song et al., "PRAS40 hyperexpression promotes hepatocarcinogenesis," *EBioMedicine*, vol. 51, p. 102604, 2020.

[28] B. Zhu, D. Cheng, S. Li, S. Zhou, and Q. Yang, "High expression of XRCC6 promotes human osteosarcoma cell proliferation through the β-catenin/Wnt signaling pathway and is associated with poor prognosis," *International Journal of Molecular Sciences*, vol. 17, no. 7, 2016.

[29] L. A. Henríquez-Hernández, A. Valenciano, P. Foro-Arnalot et al., "Association between single-nucleotide polymorphisms in DNA double-strand break repair genes and prostate cancer

aggressiveness in the Spanish population," *Prostate Cancer and Prostatic Diseases*, vol. 19, no. 1, pp. 28–34, 2016.

[30] L. Trixl and A. Lusser, "The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark," *Wiley interdisciplinary reviews. RNA*, Wiley Interdiscip Rev RNA, vol. 10, no. 1, , 2019.

[31] N. Shukeir, P. Pakneshan, G. Chen, M. Szyf, and S. A. Rabbani, "Alteration of the methylation status of tumor-promoting genes decreases prostate cancer cell invasiveness and tumorigenesis in vitro and in vivo," *Cancer Research*, vol. 66, no. 18, pp. 9202–9210, 2006.

[32] H.-H. Ng, Y. Zhang, B. Hendrich et al., "MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex," *Nature Genetics*, vol. 23, no. 1, pp. 58–61, 1999.

[33] D. Cheishvili, F. Chik, C. C. Li et al., "Synergistic effects of combined DNA methyltransferase inhibition and MBD2 depletion on breast cancer cells; MBD2 depletion blocks 5-aza-2′-deoxycytidine-triggered invasiveness," *Carcinogenesis*, vol. 35, no. 11, pp. 2436–2446, 2014.

[34] R. Peng, X. Huang, C. Zhang, X. Yang, Y. Xu, and D. Bai, "Overexpression of UHRF2 in intrahepatic cholangiocarcinoma and its clinical significance," *OncoTargets and Therapy*, vol. Volume 10, pp. 5863–5872, 2017.

[35] Z. Ye, J. Li, X. Han et al., "TET3 inhibits TGF-β1-induced epithelial-mesenchymal transition by demethylating miR-30d precursor gene in ovarian cancer cells," *Journal of Experimental & Clinical Cancer Research*, vol. 35, no. 1, p. 72, 2016.

[36] B. Sousa, J. Pereira, and J. Paredes, "The crosstalk between cell adhesion and cancer metabolism," *International Journal of Molecular Sciences*, vol. 20, no. 8, 2019.

[37] N. Beauchemin and A. Arabzadeh, "Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis," *Cancer and Metastasis Reviews*, vol. 32, no. 3-4, pp. 643–671, 2013.

[38] L. N. Kent and G. Leone, "The broken cycle: E2F dysfunction in cancer," *Nature Reviews Cancer*, vol. 19, no. 6, pp. 326–338, 2019.

[39] W. Barczak, L. Jin, S. M. Carr et al., "PRMT5 promotes cancer cell migration and invasion through the E2F pathway," *Cell Death & Disease*, vol. 11, no. 7, p. 572, 2020.

[40] T. Xiang, K. He, S. Wang, W. Chen, and H. Li, "Expression of zinc finger and BTB domain-containing 4 in colorectal cancer and its clinical significance," *Cancer Management and Research*, vol. Volume 12, pp. 9621–9626, 2020.

[41] H. Zhang, Z. Dai, W. Wu et al., "Regulatory mechanisms of immune checkpoints PD-L1 and CTLA-4 in cancer," *Journal of Experimental & Clinical Cancer Research*, vol. 40, no. 1, p. 184, 2021.

[42] N. Zhang, Z. Dai, W. Wu et al., "The predictive value of monocytes in immune microenvironment and prognosis of glioma patients based on machine learning," *Frontiers in Immunology*, vol. 12, p. 656541, 2021.

[43] D. Wu, Y. Ding, T. Wang et al., "Significance of tumor-infiltrating immune cells in the prognosis of colon cancer-infiltrating immune cells in the prognosis of colon cancer," *OncoTargets and Therapy*, vol. 13, pp. 4581–4589, 2020.

[44] Z. Zhang, L. Ma, S. Goswami et al., "Landscape of infiltrating B cells and their clinical significance in human hepatocellular carcinoma," *OncoImmunology*, vol. 8, no. 4, 2019.

[45] C. Gu-Trantien, S. Loi, S. Garaud et al., "CD4+ follicular helper T cell infiltration predicts breast cancer survival," *Journal of Clinical Investigation*, vol. 123, no. 7, pp. 2873–2892, 2013.

[46] J. Gu, J. Zhang, W. Huang et al., "Activating miRNA-mRNA network in gemcitabine-resistant pancreatic cancer cell associates with alteration of memory CD4+ T cells," *Annals of Translational Medicine*, vol. 8, no. 6, p. 279, 2020.