



OPEN

Novel prognostic biomarkers, METTL14 and YTHDF2, associated with RNA methylation in Ewing's sarcoma

Jie Jiang^{1,3}, Qie Fan^{2,3}, Haishun Qu², Chong Liu¹, Tuo Liang¹, Liyi Chen¹, Shengsheng Huang¹, Xuhua Sun¹, Jiarui Chen¹, Tianyou Chen¹, Hao Li¹, Yuanlin Yao¹ & Xinli Zhan¹✉

Ewing's sarcoma has a poor prognosis and high metastasis rate; thus, it is critical to explore prognostic biomarkers of m6A-related genes. Two datasets were downloaded from the Gene Expression Omnibus database, m6A-related genes were extracted, and prognostic models were constructed using the least absolute shrinkage and selection operator and multivariate COX regression analyses. Immune cell composition and drug sensitivity analyses were performed, and our analysis was validated using laboratory methods of immunohistochemical specific staining and qRT-PCR. Ewing's sarcoma prognostic model demonstrated that the survival rate of cases in the high-risk group was much lower than that of the low-risk group. Naïve B cells, macrophages M0, macrophages M1, and resting mast cells are closely associated with Ewing's sarcoma. METTL14 and YTHDF2 are strongly associated with multiple drug sensitivity. Immunohistochemical specific staining revealed higher expression of both METTL14 and YTHDF2 in Ewing's sarcoma than in the paraneoplastic tissues. The results of qRT-PCR showed that METTL14 expression was significantly higher in both ES cell lines than in the control cell line. The prognostic model constructed using m6A-related genes METTL14 and YTHDF2, can be a potential prognostic biomarker for Ewing's sarcoma, with the survival rate of cases in the high-risk group being much lower than that of the low-risk group.

Ewing's sarcoma (ES) represent a newly emerging subgroup of small round cell sarcomas. They share several morphological, immunohistochemical, molecular, and clinical features and have recently become the second most common type of skeletal tumor¹. This disease originates from mesenchymal stem cells, and it is the second most common sarcoma of the skeleton in young individuals and children. Further, this is also a highly aggressive type of mesenchymal tumor showing a very high metastatic rate²⁻⁴. The epigenetic tumor mechanisms have implications in mesenchymal tumors, chondroblastoma, epithelial sarcoma, and ES, with a higher prevalence of these diseases in young individuals⁵. Thus, possible disease prognostic biomarkers should be determined due to the high metastasis rate and poor prognosis.

The modifications in RNA methylation account for more than 60% of all RNA modifications, with those in N-6 methylation being the most prevalent among all higher biological lncRNAs and mRNAs. Previous studies have demonstrated that the complexity of cancer development can be determined by m6A methylation modifications by regulating cancer-related biological functions. The m6A modifications of the noncoding RNAs modulates the stability, cleavage, transport, and degradation of noncoding RNAs; and also monitors cancer cell proliferation, metastasis, homeostasis, and stem cell differentiation by altering the biological functions in cancer⁶⁻⁸. RNA m6A modifications regulate RNA splicing, allostereism, stability, and protein translation. The m6A is primarily catalyzed by RNA transferases METTL3, METTL14, and METTL16 (writers), cleared by demethylases FTO and ALKBH5 (scavengers), and interacts with m6A binding proteins such as YTHDF1 and IGF2BP1 (readers). RNA demethylases, methyltransferases, and m6A binding proteins are also upregulated in various human cancer tissues. This increased the expression of cancer cell proliferation, survival, and metastatic tumor proteins⁹.

Immune cell infiltration indicates the modification of the immune cell microenvironment, which possesses the tumorigenesis site. Most tumor cells express antigens mediating recognition by host CD8+ T cells. Based on

¹The First Clinical Affiliated Hospital of Guangxi Medical University, Nanning 530021, People's Republic of China. ²Department of Traditional Chinese Medicine, The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning 530000, People's Republic of China. ³These authors contributed equally: Jie Jiang and Qie Fan. ✉email: zhanxinli_215@163.com

Gene	conMean	treatMean	logFC	p-value
METTL3	5.699018	7.14749	0.326723	1.50E-13
METTL14	4.58475	5.824376	0.34526	1.08E-15
WTAP	5.598477	6.282423	0.166287	1.05E-13
KIAA1429	4.444641	5.170975	0.21837	4.87E-14
RBM15	5.27808	6.091391	0.206758	1.08E-15
ZC3H13	6.066223	7.29778	0.266659	1.57E-11
YTHDC1	4.593895	5.222197	0.184939	1.08E-15
YTHDC2	4.840991	5.124283	0.082048	0.009839
YTHDF1	7.71216	8.424596	0.127472	1.22E-09
YTHDF2	7.580526	8.96268	0.241632	1.08E-15
HNRNPC	6.54433	7.751271	0.244187	1.08E-15
FTO	7.779931	7.944228	0.03015	0.068138
ALKBH5	7.533025	6.182872	-0.28495	2.16E-15

Table 1. Differential expression of m6A-related genes. Table shows the differential expression of all m6A-related genes in Ewing's sarcoma.

the molecular features and cells of the tumor microenvironment, tumor escape is classified into two main types. The first type, including the infiltrating T cells, is characterized by most inflammatory chemokines and innate immune activation of type I interferons. In contrast, the second is phenotyped as lacking this T cell inflammatory phenotype, possibly through rejection by the immune system to resist immune attack¹⁰. Moreover, in recent years, an increasing number of studies have shown that tumor immune cell infiltration plays a critical role in cancer^{11–15}.

In the present study, we analyzed gene expression and survival data downloaded from the Gene Expression Omnibus (GEO) data using a combination of bioinformatics and laboratory validation to determine the role of m6A-related genes in ES. Furthermore, we aimed to provide new evidence by constructing prognostic models and analyzing immune cell composition and drug sensitivity to assist in clinical diagnosis and treatment.

Materials and methods

Data download and preliminary processing. The gene expression and clinical information data used for the present study were downloaded from the GEO (<https://www.ncbi.nlm.nih.gov/gds/>) database. Our selection criteria for datasets: 1. unique datasets with available clinical data and 2. all solid tumor specimens derived from Ewing's sarcoma. We downloaded two different datasets to make the analysis more logical and rigorous. We used GSE17674 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17674>)¹⁶, and GSE63157 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>) as the experimental and validation datasets, respectively¹⁷. We have corrected for batch effects using the *sva*_combat method. We used Perl 5 (v5.30.2) for all text processing. Compared to previous studies, we have constructed a prognostic model of Ewing's sarcoma using more sophisticated and accurate bioinformatics techniques for predicting prognosis. The turning language R was used for processing and graphing the data. All downloaded and completed data were normalized and log₂ processed.

Extraction of m6A-related genes, differential expression analysis, and correlation analysis. To analyze the differential expression of these genes, we used the “limma” packages to perform differential gene expression analysis. We compared the gene expression of Ewing's sarcoma samples with controls and set $P < 0.05$ as a significant differentially expressed gene. This was then visualized using the ‘ggplot’ package, the ‘pheatmap’ package and the ‘ggpubr’ package. The differential expression genes were visualized as heat maps and volcano plots. The differential expression of all m6A-related genes can be found in Table 1. Subsequently, the gene expression of m6A-related genes was extracted from 21,655 genes using the “limma” package. The “pheatmap” package was utilized to visualize them. The “ggpubr” and “reshape2” packages were utilized to obtain the m6A-related gene expression in ES and paraneoplastic tissues and visualized them as violin maps. Details of all m6A-related differentially expressed genes can be found in Table 1. Subsequently, the correlation of m6A-related differential genes was analyzed using the “corrplot” and “circlize” packages and visualized as correlation circle plots.

Construction of a prognostic model for Ewing's sarcoma. Several techniques were used to analyze the relationship between m6A differentially expressed genes and ES prognosis to determine those closely associated with prognosis. First, the least absolute shrinkage and selection operator (LASSO) regression analysis^{18,19} method was used for maximum model refinement. This advanced, streamlined modeling approach compresses the coefficients of insignificant variables to zero; thus, fulfilling the variable selection effect. Subsequently, a multivariate COX regression analysis of survival status and time was used along with gene expression to obtain a more accurate prognostic model. After these two precise screening steps, we constructed a logistic regression model for m6A-related Ewing's sarcoma. Finally, a risk score was assigned to each Ewing's sarcoma case and, based on its median value, cases greater than or equal to the median value were classified as high risk and those less than the median value as low risk.

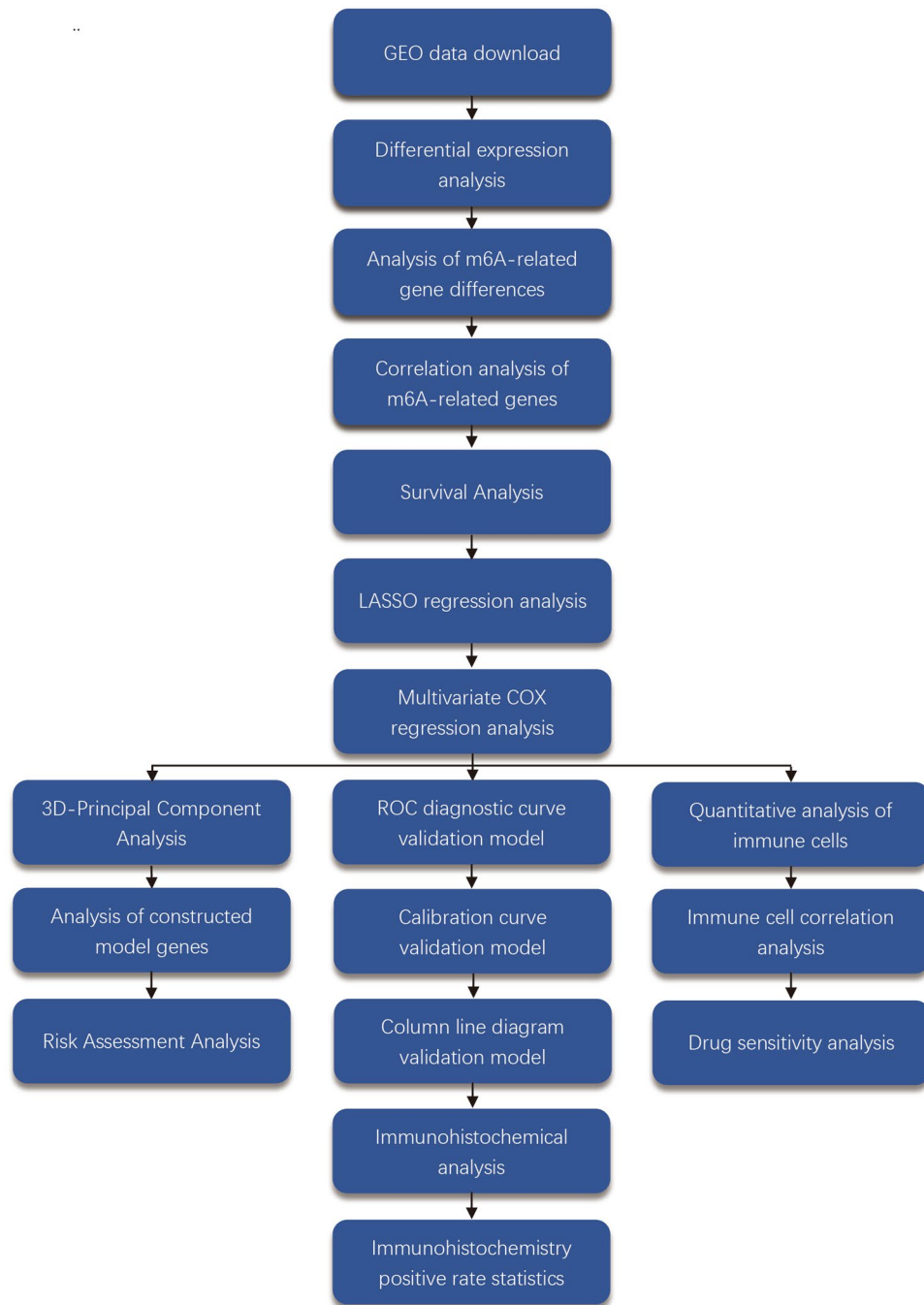


Figure 1. Workflow diagram. From the diagram, we can see the flow of our work clearly and briefly.

The three-dimensional (3D)-principal component analysis, gene expression analysis, and risk scoring. 3D-Principal Component Analysis is a commonly used tool for data analysis. For a set of data that may be linearly correlated between different dimensions, PCA is able to transform this set of data into data that is linearly uncorrelated between the dimensions by means of an orthogonal transformation^{20–22}. After an ES prognostic model was constructed, a 3D-principal component analysis was performed to analyze whether the two could be distinguished based on high and low-risk groupings. The “scatterplot3d” package was utilized to analyze and visualize it. Subsequently, based on the constructed prognostic model, the “ggpubr” package was used to analyze gene expression in-depth. All cases were ranked serially from low to high risk, and the corresponding survival status for each patient was plotted. Moreover, the expression of the genes used to construct the prognostic model in the high-and low-risk groups was analyzed (Fig. 1A–C).

Survival analysis. Survival analysis was conducted for all cases to analyze the role of m6A-related genes in of sarcomas prognosis. These were analyzed using two different methods. First, based on the high and low expression values of each m6A-related gene, all patients were divided into high and low expression groups, and Kaplan–Meier survival curves were plotted. Subsequently, based on the constructed prognostic model, these curves were analyzed for all cases by dividing them into high-and low-risk groups. To verify the accuracy of the constructed prognosis, the same approach was utilized for GSE63157 to construct a prognostic model, yielding a Kaplan–Meier survival curve for the validated gene set.

Examination of the prognostic model of Ewing’s sarcoma. The accuracy of our constructed prognostic model of ES was tested using the receiver operating characteristic (ROC) diagnostic and calibration curves as well as the column line graph method for testing the model. First, the ROC diagnostic curves were plotted for the model using the “survival”, “survminer”, and “timeROC” packages. Second, the ROC diagnostic curves of clinical information were prepared using these three R package pairs. Finally, the “rms” package was utilized for the calibration curves and column plots.

Immune cell composition analysis and correlation analysis. In this study, all cases of ES were quantified to assess the relationship between the samples and immune cell composition using CIBERSORTs software^{23–25}. The immune cells of each sample of ES were set to 100%, and the composition of a total of 22 immune cells was analyzed and visualized as a composition map. Subsequently, the relationship between these immune cells was assessed, and a correlation analysis was performed using the “corrplot” package. In contrast, we performed an in-depth analysis of m6A-related gene expression and immune cell content in ES to analyze the role of immune cells at multiple levels and perspectives.

Drug sensitivity analysis. CellMiner software is based on 60 cancer cells listed by the National Cancer Institute’s Center for Cancer Research (NCI). The NCI-60 cell line is currently the most widely used cancer cell sample population for anti-cancer drug testing^{26,27}. It allows researchers to query data on the 22,379 genes identified in the NCI-60 cell line, as well as the 20,503 compounds analyzed (including 102 drugs approved by the US Food and Drug Administration). Here, a drug sensitivity analysis was conducted to investigate the relationship between two m6A genes and drug sensitivity for better guidance of clinical drug use. All data related to drug sensitivity was downloaded from the CellMiner database (version: 2021.1, database: 2.6). Subsequently, the programming language R (×64, version 4.0.2) was used to assess the relationship between m6A-related gene expression information and drug sensitivity using the “impute”, “limma”, “ggplot2”, and “ggpubr” packages.

Immunohistochemical analysis. The accuracy of the bioinformatics analysis was determined by performing immunohistochemical analysis on the two genes that were used for model construction. The pathological sections for immunohistochemistry were excised specimens obtained for pathological testing during surgery at the First Clinical Affiliated Hospital of Guangxi Medical University. The study followed the Declaration of Helsinki and was approved by the ethics department of the First Clinical Affiliated Hospital of Guangxi Medical University, and immunohistochemical staining analysis was performed on anonymous tissue specimens. Therefore, this study needs for the patient’s written informed consent is waived. The immunohistochemical analysis was performed for a total of 12 pathological sections for each gene in three pairs (three each of cancer and para cancer tissue sections). Antibodies for immunohistochemical specific staining were purchased from Proteintech (<https://www.ptgcn.com/>) (under item numbers 26158–1-AP and 24744–1-AP). All sections were dewaxed and hydrated by cutting 3 μm thick paraffin sections, baking them overnight at 64 °C, dewaxing in water, and immersing in distilled water. Subsequently, antigen repair was performed, endogenous peroxidase was blocked, water was shaken off the sections, and sections were circled at 3 mm from the tissue using an immunohistochemical grease pencil. The peroxidase blocker was added dropwise to the sections for 10 min at room temperature and washed thrice for 3 min each in phosphate-buffered saline and 95%, 100%, and 100% alcohols. The slices were then sealed using neutral environmental protection resin. All stained pathological tissue sections were placed under an inverted microscope for observation and image acquisition. Subsequently, a statistical analysis of the positive regions of specific staining was performed for each gene. The positive rates for all immunohistochemically stained images were calculated using Image J software. Moreover, the data were imported into IBM SPSS Statistics 25, where they were analyzed using a t-test of two-paired sample means. Finally, the results of the statistics were visualized using GraphPad Prism 8.

Quantitative reverse transcription PCR (qRT-PCR). The normal human 293 T cells used in this study were purchased from Shenzhen Aowei Biotechnology, human Ewing’s sarcoma cells A673 were purchased from SHANGHAI WHELAB BIOSCIENCE LIMITED and human Ewing’s sarcoma cells RD-ES were purchased from Zhejiang Meisen Cell Technology Ltd., Co. 293 T cells and A673 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) containing 10% FBS, 100 U/mL penicillin, and 100 mg streptomycin at 37 °C and 5% CO₂ atmosphere. A673 cells were cultured in RPMI medium 1640 containing 10% FBS, 100 U/mL penicillin, and 100 mg streptomycin at 37 °C and 5% CO₂ atmosphere. RNA extraction using Hipure Total RNA Mini kit (Magen, China) followed by quantitative real-time PCR (qRT-PCR) was used to purify the total intracellular RNA from the induced samples. Subsequently, 1000 ng of the extracted RNA was reverse transcribed into cDNA using a cDNA synthesis kit (Takara, China). qRT-PCR was used to detect the gene expression using the LightCycler 480 Sequence Detection System (Roche, Germany) and PCR Green Master Mix (Roche, Germany). The activation cycle of the polymerase included 10 min at 95 °C and 15 s at 95 °C, and 45 cycles such cycles were

Primer	Sequence (5'–3')
METTL14-F	GGCTATGACTCCTAATCACGCTTCC
METTL14-R	ATCCAAGTTCAAGTCCACACCACAG
GAPDH -F	CCACTCCTCCACCTTTGAC
GAPDH -R	ACCCTGTTGCTGTAGCCA

Table 2. Primer sequences for GAPDH and METTL14. Table shows the primer sequences for GAPDH and METTL14 during the primer design phase.

performed. Glyceraldehyde 3-phosphate dehydrogenase (GADPH, Abcam, USA) was used as the internal control, and the data analysis was performed using the 2- $\Delta\Delta$ CT method. The analysis for each gene was performed in triplicate. The primer sequences of the target genes are presented in Table 2.

Financial and competing interests disclosure. There were no conflicts of interest between all co-authors of this study. The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

Ethical disclosure. This study was approved by the Ethics Review Committee of the First Clinical Affiliated Hospital of Guangxi Medical University and was in accordance with the Declaration of Helsinki of the World Medical Association.

Data sharing statement. The datasets supporting the conclusions of this article are available in the GEO database (GSE21257: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21257>, and GSE63157: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63157>).

Results

Data download and preliminary processing. We downloaded two datasets from the GEO database, GSE17674 as the experimental set for constructing the prognostic model for ES and GSE63157 as the validation set for the prognostic model. Among them, GSE17674 comprised 44 ES and 18 normal samples. A total of 85 ES samples were included in GSE63157 as the validation set. We analyzed a total of 147 samples from 129 ES and 18 normal control samples. All gene expression data were normalized using R to remove inter-batch differential processing. The flow chart is illustrated in Fig. 2.

Differential expression analysis, extraction of m6A-related genes, and correlation analysis. We first performed differential expression analysis on all genes and obtained a total of 3947 differentially expressed genes according to the set cut-off values. Heat map and volcano plots are shown in Fig. 3A, B. Subsequently, we extracted these genes from all gene expression matrices for subsequent analysis to analyze the role of m6A genes in ES. We performed differential expression analysis of these m6A-related genes (Fig. 4A) and observed that the differences in 12 of these 13 genes were statistically significant ($P < 0.05$). We analyzed the expression of m6A-related genes in tumor and normal samples and found that the expression of all 12 genes was higher in tumor samples than in normal samples except for ALKBH5 (Fig. 4B). We found that there was a correlation between both these m6A-related genes, with some showing synergistic high expression (Fig. 4C) and some showing low synergistic expression. We have placed details of the m6A-related differentially expressed genes in Table 1.

Construction of a prognostic model for Ewing's sarcoma. We first performed LASSO regression analysis on these 13 m6A-related genes to maximize the number of genes in the prognostic model and obtain the most accurate and streamlined prognostic model. As observed from the results of the LASSO regression analysis (Fig. 5A and B), the most streamlined results can be achieved when the number of genes is four. From the multivariate COX regression analysis (Fig. 5C), we found that after analysis, only two genes, METTL14 and YTHDF2, were closely associated with the presence of survival. Details of the results of the multivariate COX regression analysis can be found in Table 3. We constructed a prognostic model for ES based on these two genes, and each case was rated for risk. The cases with risk scores greater than or equal to the mean were classified into the high-risk group and those less than the mean into the low-risk group.

3D-principal component analysis, gene expression analysis, and risk scoring. From the constructed 3D-principal component analysis plot, we observed (Fig. 5D) that the cases in the high-risk group were distributed in the positive axis direction of the PC3 axis, while most of the cases in the low-risk group were distributed in the negative axis direction of the PC3 axis. Moreover, we found that the gene expression value of METTL14 was lower in the high-risk group than that in the low-risk group (Fig. 5E), with a statistically significant difference. The gene expression value of YTHDF2 was higher in the high-risk group than that in the

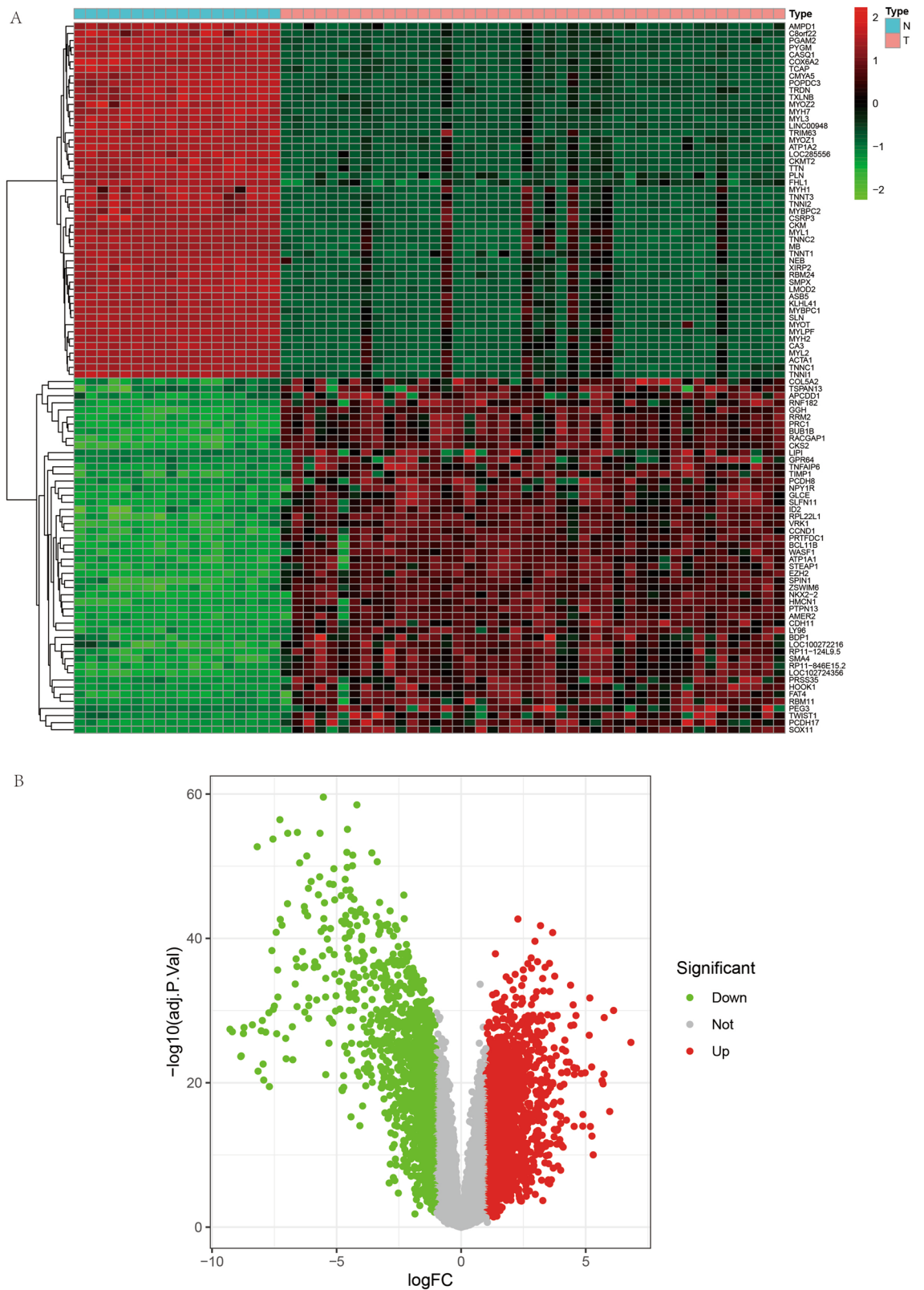


Figure 2. Heat map and volcano plots of differentially expressed genes. (A) shows a heat map of the differentially expressed genes. (B) shows the volcano map of differentially expressed genes.

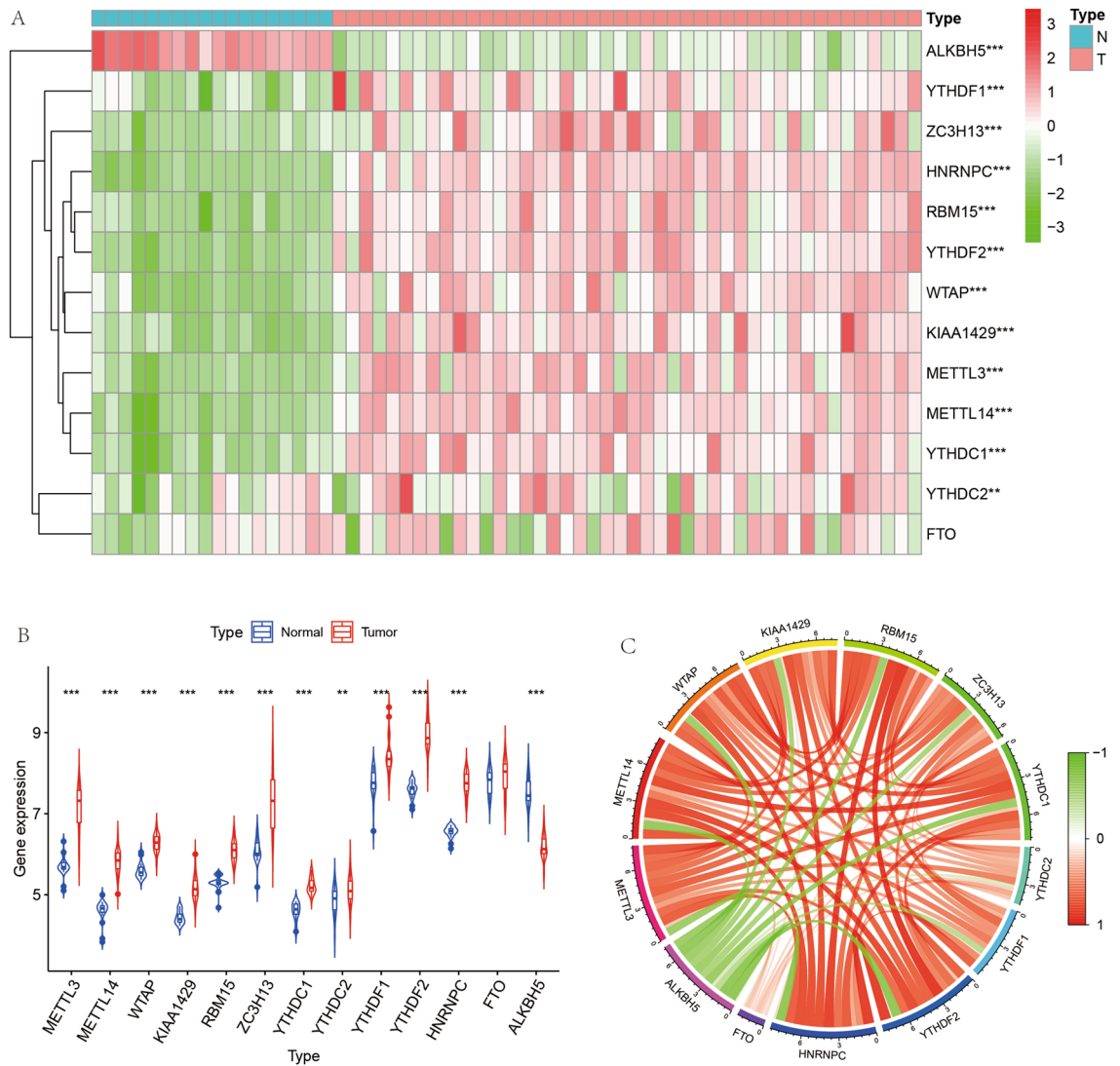


Figure 3. Heat map, violin map and correlation circle map of m6A-related genes. **(A)** shows a heat map of differential expression of m6A-related genes, with red indicating highly expressed genes and green indicating lowly expressed genes. **(B)** represents the expression of m6A-related genes in paraneoplastic tissues and in Ewing's sarcoma. Figure **C** represents the correlation plot between m6A-related genes, with the linkage between quantitative genes in red for synergistic high expression and in green for synergistic low expression.

low-risk group (Fig. 5F), with a statistically significant difference. Each case was ranked consecutively based on the risk score from the lowest to the highest to obtain the risk graph (Fig. 1A–C).

Survival analysis. We first divided all cases into the high and low expression groups of this gene for Kaplan–Meier survival curves based on the mean expression of each m6A gene to utilize the survival data of ES completely. We observed that except for METTL14, YTHDC1, and YTHDF2 (Fig. 6A–M), the differences between the other ten genes in these two groups were not statistically significant ($P > 0.05$). Moreover, the results of our constructed prognostic model for m6A-associated ES showed that cases in the high-risk group possessed a worse survival rate than those in the low-risk group (Fig. 6N), with a statistically significant difference ($P < 0.05$). Also, we constructed a prognostic model using the same method for GSE63157, yielding similar results (Fig. 6O), with cases in the high-risk group having worse survival than those in the low-risk group ($P < 0.05$).

Examination of the prognostic model of Ewing's sarcoma. The ROC diagnostic curve indicated (Fig. 7A) that the constructed prognostic model predicts the 1-year, 3-year, and 5-year survival rates with an accuracy above 50%. Also, as depicted in Fig. 7B, the area under the curve of the ROC curve for the diagnosis of risk, survival status, age, and sex with clinical information is greater than 0.5. The calibration plot depicts (Fig. 7C) that the starting point and focus of the predicted survival largely overlap, which further illustrates the accuracy of our constructed prognostic model. In contrast, we constructed a line graph (Fig. 7D) and were able to predict the survival rate of patients more accurately using the line graph.

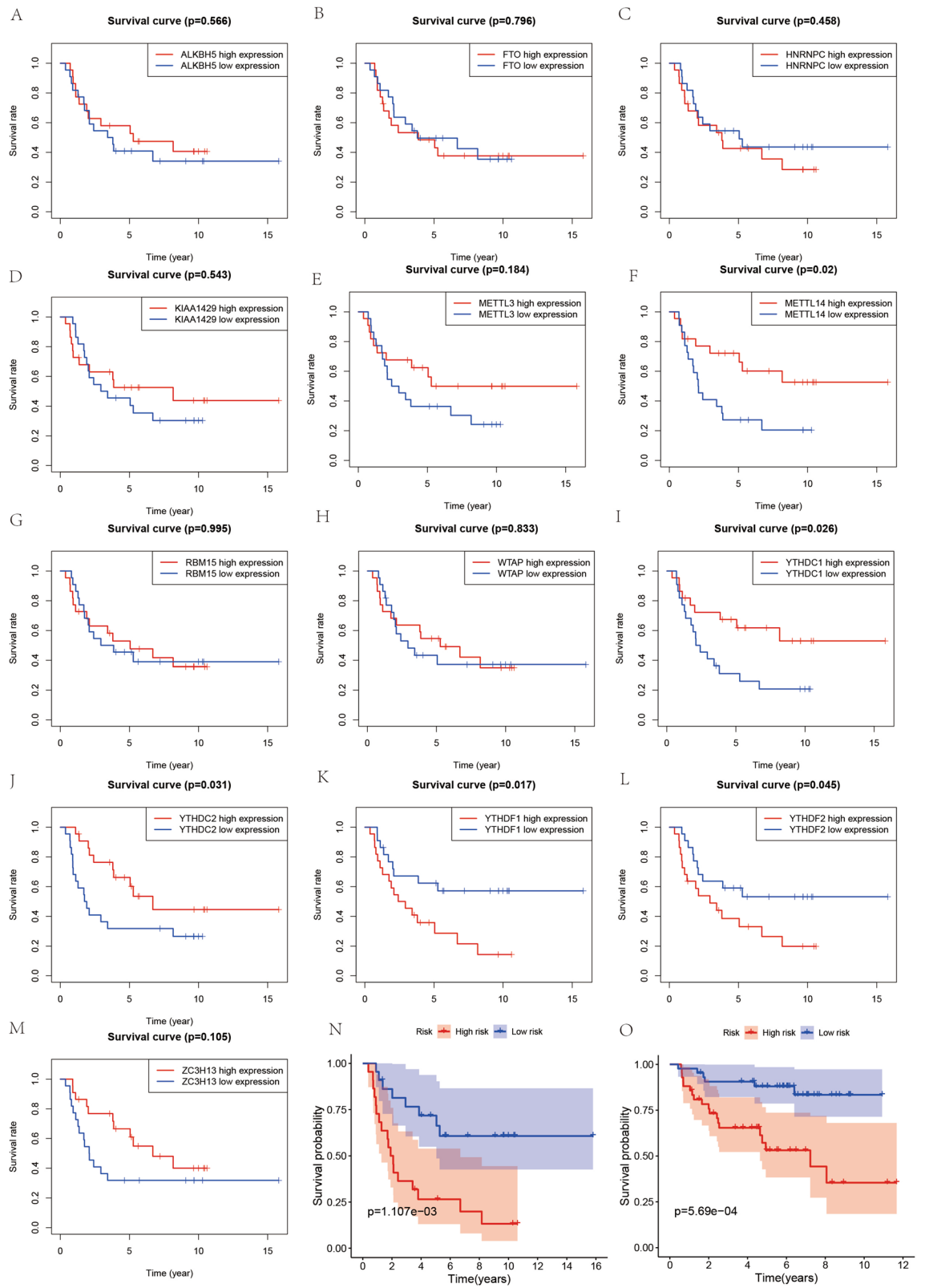


Figure 4. Survival analysis. (A–L) and (M) represent the results of survival analysis based on high and low expression of genes in Ewing’s sarcoma. (N) represents the results of the constructed prognostic model. (O) represents the prognostic model validation of the validation set.

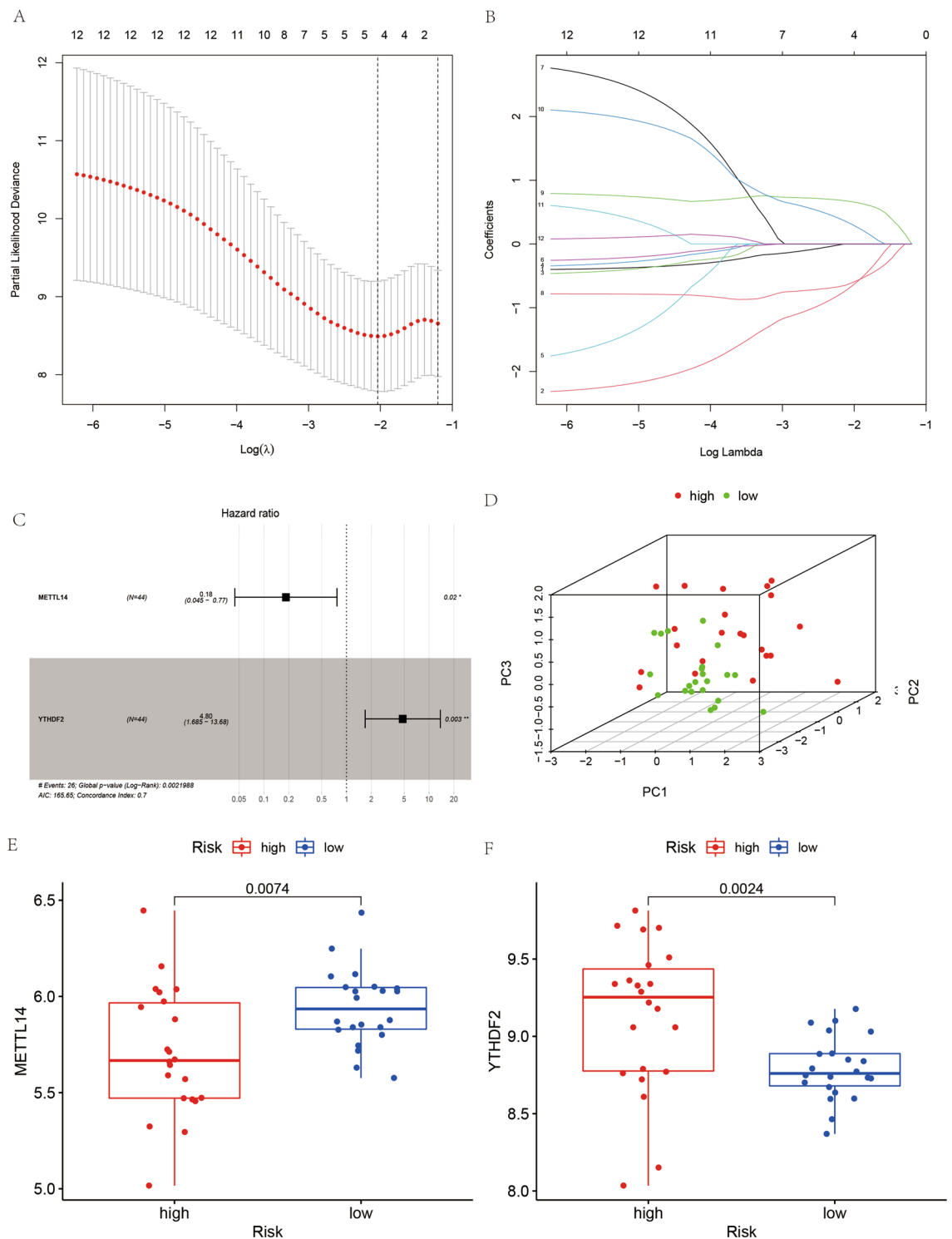


Figure 5. Diagram of the construction process of the prognostic model. (A) and (B) represent the results of the LASSO regression analysis. (C) represents a plot of the results of the multivariate COX regression analysis. (D) shows the 3D-principal component analysis plot, with red points indicating high-risk cases and green points indicating low-risk cases. (E) and (F) indicate the expression of METTL14 and YTHDF2 in the high-risk group and in the low-risk group, respectively.

ID	Coef	HR	HR.95L	HR.95H	p value
METTL14	-1.6877	0.184944	0.044527	0.768168	0.02018
YTHDF2	1.569026	4.801968	1.685304	13.68234	0.003314

Table 3. Results of the multivariate COX regression analysis. Table shows the detailed results of the multivariate COX regression analysis.

Immune cell composition analysis and correlation analysis. We used CIBERSORT software to quantify the composition of immune cells in both ES and normal cases. This is illustrated in Fig. 8A, with each sample comprising 22 types of immune cells. Moreover, there is a correlation between the expression of these 22 immune cells (Fig. 8B), with the red line synergizing high expression and green synergizing low expression. We performed a differential analysis of immune cell components in all cases (Fig. 9A) and found significant differences in the expression of naïve B cells, macrophages M0, macrophages M1, and resting mast cells in ES and normal controls ($p < 0.05$). Also, we analyzed the relationship between METTL14 and YTHDF2 with these immune cells (Fig. 9B–E) and found that METTL14 showed a negative correlation with naïve B cells, macrophages M1, and resting mast cells, and a positive correlation with macrophages M0 ($P < 0.05$). In contrast (Fig. 9F–I), YTHDF2 showed a negative correlation with naïve B cells and macrophages M1 and a positive correlation with macrophages M0 and resting mast cells ($P < 0.05$).

Drug sensitivity analysis. To better explore the relationship between Ewing's sarcoma and drug sensitivity, a drug sensitivity analysis was performed. Using drug sensitivity analysis, we found that METTL14 and YTHDF2 were quite closely related to the drug sensitivity of multiple drugs (Fig. 10). Among these, METTL14 showed a positive correlation with the drug sensitivity of several drugs, including Nelarabine, Hydroxyurea, and Clofarabine, and a negative correlation with the Mithramycin, Depsipeptide, and Actinomycin. In contrast, YTHDF2 was closely associated with the sensitivity of several drugs such as Dasatinib, Vorinostat, and Brigatinib. A positive correlation indicates that the sensitivity of the corresponding drug increases with the increase in the gene expression value. A negative correlation indicates that the sensitivity of the corresponding drug is weakened with the increase in the gene expression value.

Immunohistochemical analysis. We performed laboratory analysis of six (three cases of ES and three cases of paraneoplastic tissue) of each gene for a total of 12 immunohistochemical specific stains. After laboratory manipulation, we obtained all the specific stained sections, which we placed in the same inverted microscope for observation. We found that the specific expression of both METTL14 and YTHDF2 was higher in ES than in the paraneoplastic tissues (Fig. 11A1–D2). We were able to find from the pictures of the specific staining by immunohistochemistry that the specific expression of both METTL14 and YTHDF2 in Ewing's sarcoma was higher than their respective expression in the control group. This is consistent with our previous analysis and further enhances the credibility of the prognostic model we have constructed. We again performed statistical analysis of the slices using Image J software for positive rate statistics, using the IBM SPSS Statistics 25 t-test for two-paired sample means. The positive rates of both METTL14 and YTHDF2 (Fig. 11E, F) were observed to be higher in ES than in the paraneoplastic tissues ($P < 0.05$).

Quantitative reverse transcription PCR (qRT-PCR). We performed in vitro qRT-PCR cellular assays for cellular level validation of the METTL14 gene in order to verify the reliability of the bioinformatics analysis. From Fig. 11G, we can learn that the expression of METTL14 gene in the control cell line 293 T was much lower than that in the human Ewing's sarcoma A673 cell line ($P < 0.05$); in addition, its expression in the control 293 T was also much lower than that in the human Ewing's sarcoma RD-ES ($P < 0.01$).

Discussion

N6-RNA methylation modifications appear as a highly abundant dynamic regulation throughout eukaryotic transcription. With an increasing number of m6A-related genes observed to be dysregulated in tumors, m6A alters the course of cancer cell development²⁸. The m6A regulators have been shown to promote the development of various types of cancer possibly, and a study by Anita et al. found that overexpression of YTHDF3 and YTHDF1 can lead to poor prognosis in patients with breast cancer²⁹. Moreover, m6A-related genes were not only dysregulated in expression in breast cancer but also in gastric cancer. Wang et al. demonstrated that high expression of METTL3 promoted glycolysis and tumor angiogenesis in gastric cancer, suggesting METTL3 as a potential therapeutic target for human gastric cancer³⁰. However, the role of m6A-related genes in ES is poorly reported, and the low prognosis and high metastatic nature of this malignant disease critically require investigators to explore the role of m6A-related genes in ES.

Methyltransferase Like 14 (METTL14) is a protein-coding gene that is most closely associated with periosteal chondrosarcoma and Miyoshi muscular dystrophy 3. The upregulation of METTL14 has been demonstrated to lead to the decrease of PERP level through m6A modification which promotes the metastasis and growth of pancreatic cancer; thus, METTL14 can be a potential therapeutic target for pancreatic cancer³¹. METTL14 has also been reported in leukemia; METTL14 exerts oncogenic effects by regulating its mRNAs (MYB and MYC) through m6A modifications³². Liu et al. found that approximately 70% of endometrial tumors exhibit reduced m6A methylation, which may be owing to mutations in METTL14 or reduced expression of METTL3

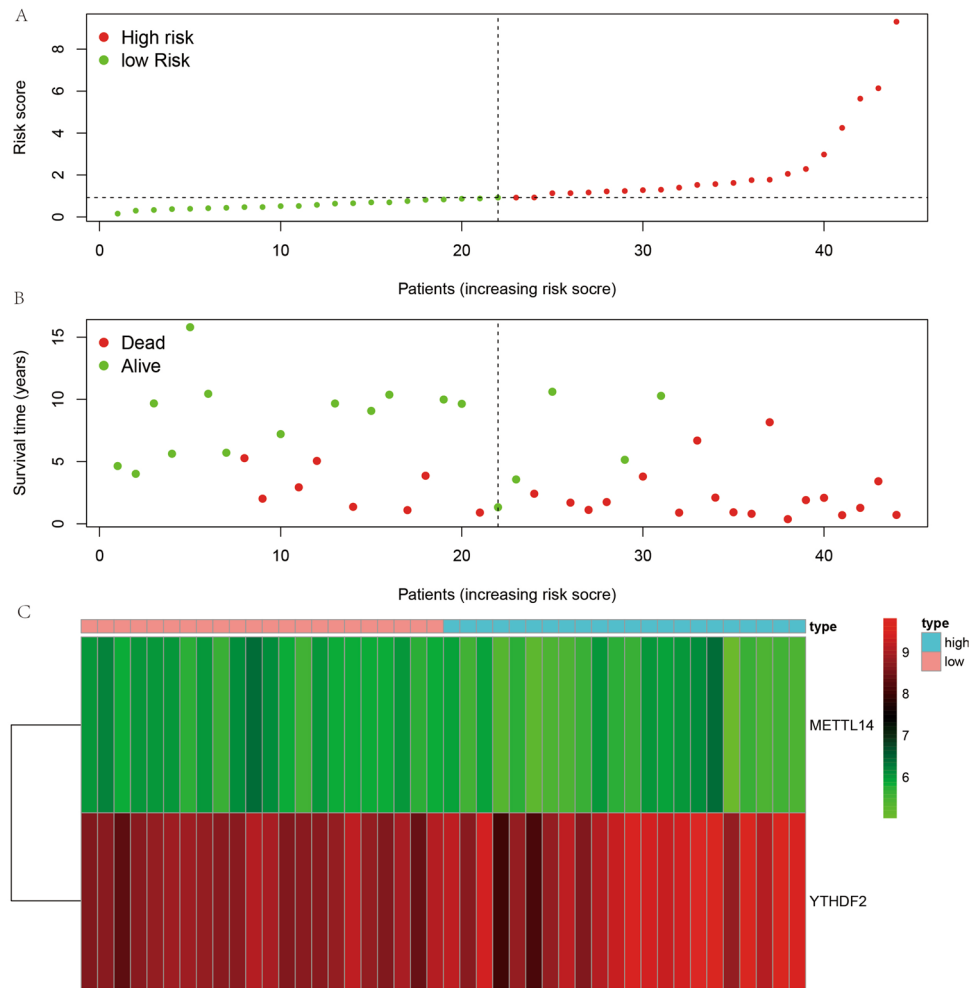


Figure 6. Risk graph for Ewing's sarcoma cases. **(A)** represents the risk score chart based on the risk of the cases in order from ground to high. **(B)** represents the graph of survival and death of the cases. **(C)** represents the expression of METTL14 and YTHDF2 in high and low risk groups.

(another component of the methylation transferase complex), and that these changes lead to a possible increase in endometrial tumorigenicity³³. Thus, the above study is consistent with our results. Our results show that the prognostic models constructed for m6A-related genes, including METTL14, have much lower survival rates in their high expression group than in the low-risk group. More importantly, we found that METTL14 showed a negative correlation with naïve B cells, macrophages M1, and resting mast cells and a positive correlation with macrophages M0 ($p < 0.05$), and these immune cells are closely related to cancer development^{34–37}. This provides new evidence for immunotherapy of ES. Moreover, we performed a correlative drug sensitivity analysis of METTL14 and found that this gene showed a quite close correlation with the sensitivity of multiple drugs, providing new reference evidence for the treatment of ES.

YTH N6-Methyladenosine RNA Binding Protein 2 (YTHDF2), a member of the superfamily containing the YTH structural domain, is a typical structural domain of eukaryotes, whose structural domain is generally located in the middle of the protein sequence and may have a function similar to that of RNA nodules. It has been found that YTHDF2 can regulate m6A methylation of OCT4 mRNA to promote liver phenotype and cancer metastasis in cancer stem cells³⁸. Dixit et al. found that the m6A-related gene YTHDF2 is aberrantly expressed in glioblastoma compared to normal neural stem cells and could be a therapeutic target for the treatment of MYC signaling in glioblastoma³⁹. This is consistent with the results of our study. Our results showed that the prognostic model constructed by m6A-related genes, including YTHDF2, had a much lower survival rate in the high expression group than that in the low-risk group. Moreover, we found that this YTHDF2 is particularly closely related to naïve B cells, macrophages M0, macrophages M1, and resting mast cells, which are immune cells. This may provide a new reference for guiding the immunotherapy of ES. Also, we found a close correlation between the sensitivity of YTHDF2 and several drugs such as Dasatinib, Vorinostat, and Brigatinib, which provides new reference information for the drug treatment of ES.

Here, we downloaded two datasets from the GEO database, constructed m6A-related ES, prognostic models using LASSO and multivariate COX regression analysis, and analyzed the prognosis of ES by differential expression analysis, 3D-principal component analysis, gene expression analysis, survival analysis, risk assessment of ES

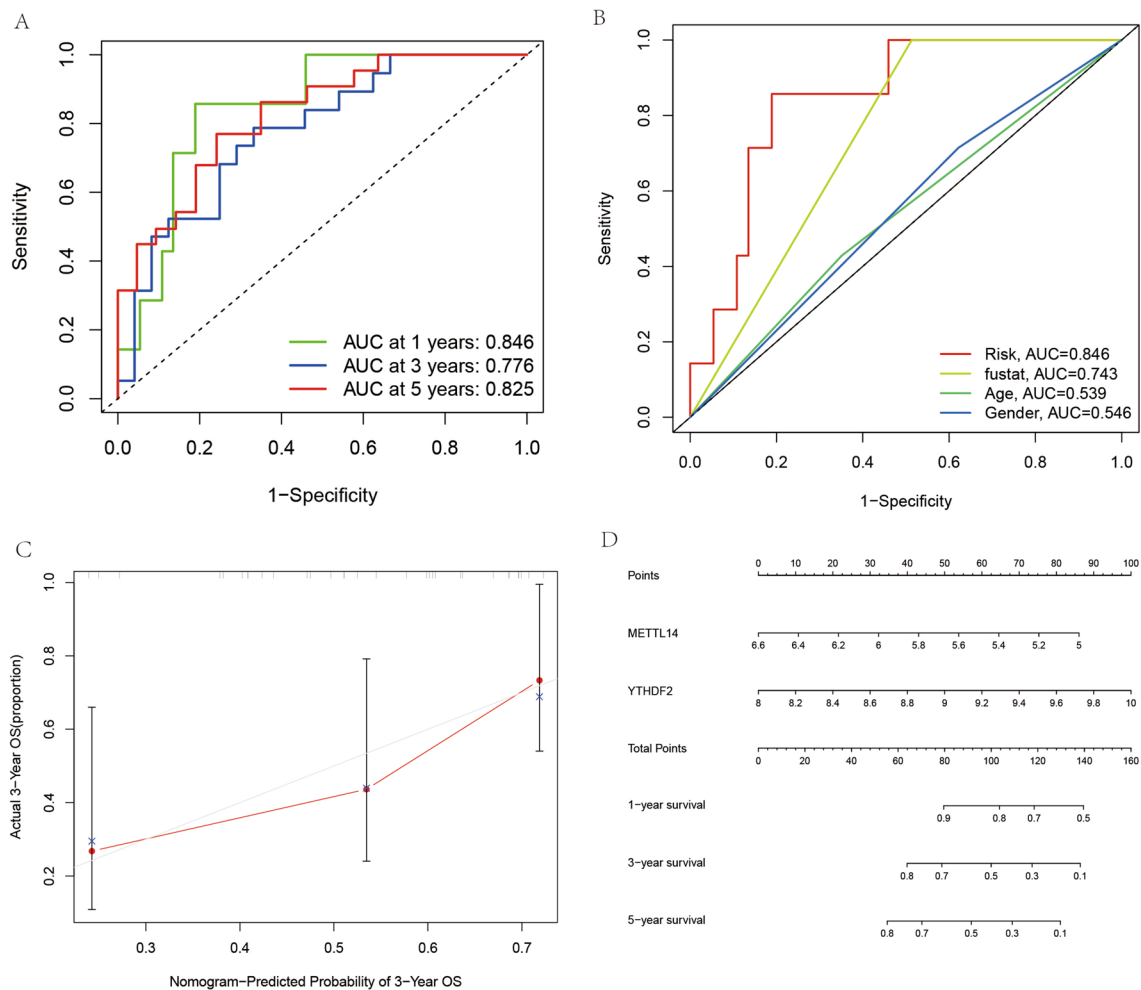


Figure 7. A test plot of the prognostic model of Ewing's sarcoma. **(A)** represents the accuracy of predicting the probability of survival at 1,3 and 5 years. **(B)** indicates the accuracy of the diagnostic clinical information. **(C)** represents the calibration curve for predicting patient survival. **(D)** represents the column line plot for predicting the probability of patient survival.

cases, ROC diagnostic curve, calibration curve analysis, and column line graph prediction analysis. The roles of m6A-related genes METTL14 and TYHDF2 in ES were analyzed using precise and sophisticated bioinformatics techniques, such as immune cell composition, immune cell correlation, immune cell-gene correlation, and drug sensitivity analyses. METTL14 and TYHDF2 serve as biomarkers of Ewing's sarcoma, and we can use the expression of these two genes in Ewing's sarcoma to predict the survival risk and prognosis of Ewing's sarcoma patients, providing new and useful information to guide clinical diagnosis and treatment. Finally, we tested our results using two different levels of laboratory methods. First, we first performed immunochemical specific staining of Ewing's sarcoma specimens derived from human tissue, and our analysis showed that both genes were expressed more in Ewing's sarcoma than in the control group, and the differences were statistically significant ($P < 0.05$). Secondly, we performed cell cultures of Ewing's sarcoma and control cells and the results of the qRT-PCR experiments performed showed that METTL14 expression was higher in both Ewing's sarcoma cell lines than in the control cell lines, with statistically significant differences ($P < 0.05$). All experimental results support the results of our bioinformatics analysis. We therefore have reason to believe that METTL14 and TYHDF2 can serve as prognostic biomarkers for Ewing's sarcoma associated with m6A.

Similar to others, our study also had certain limitations. First, the sample size was inadequate. We used a total of 147 samples for analysis and validation, which was insufficient. Second, we did not perform an adequate analysis of the specific tumor staging of ES.

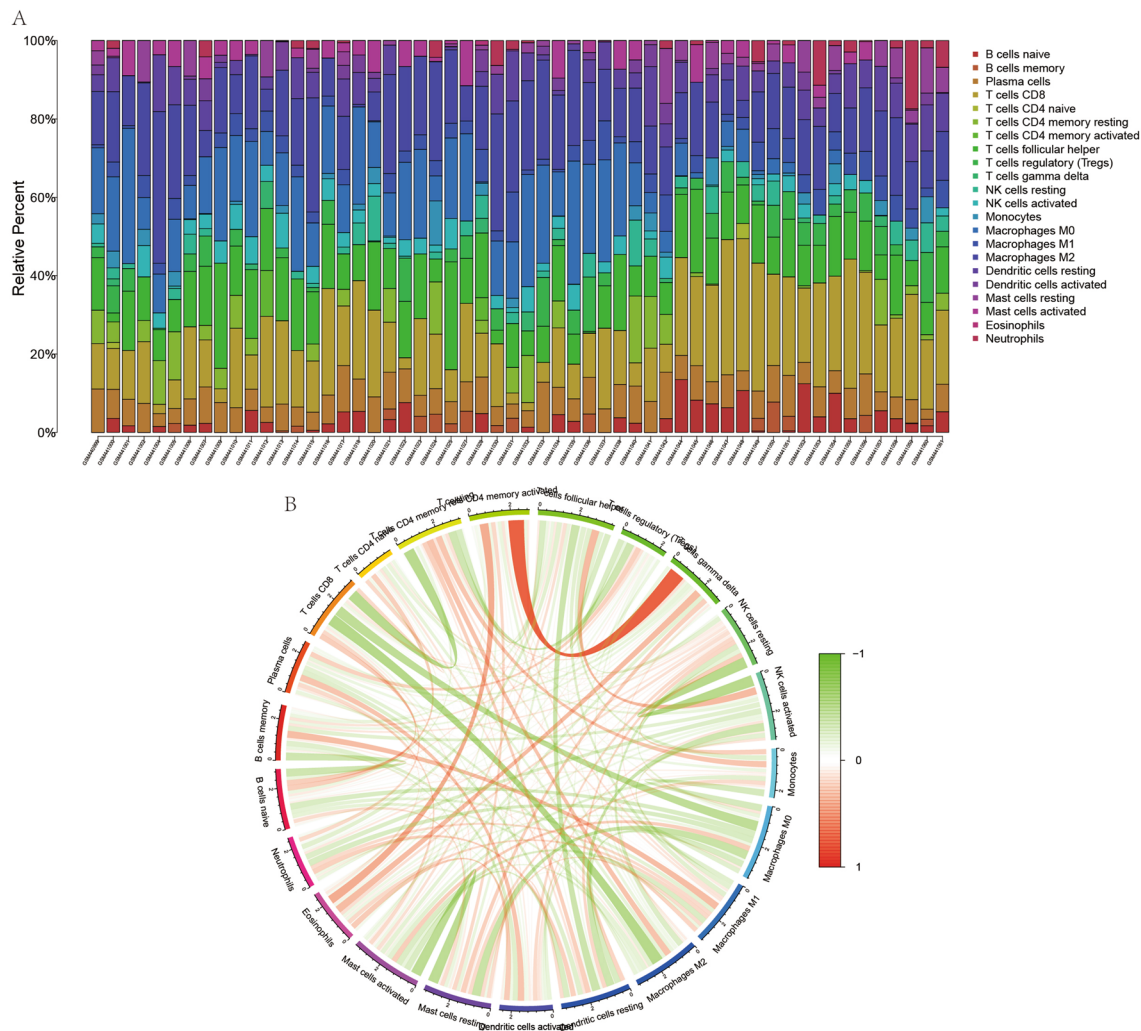


Figure 8. Immune cell composition and correlation analysis. **(A)** represents the analysis of immune cell composition for all cases, with each color indicating one immune cell. **(B)** represents the correlation analysis between immune cells, with red connecting lines indicating synergistic high expression and green links indicating synergistic low expression.

Conclusion

A prognostic model constructed using the m6A-related genes METTL14 and TYHDF2, whose survival rates for cases in the high-risk group were much lower than those in the low-risk group, can be a potential prognostic biomarker for ES.

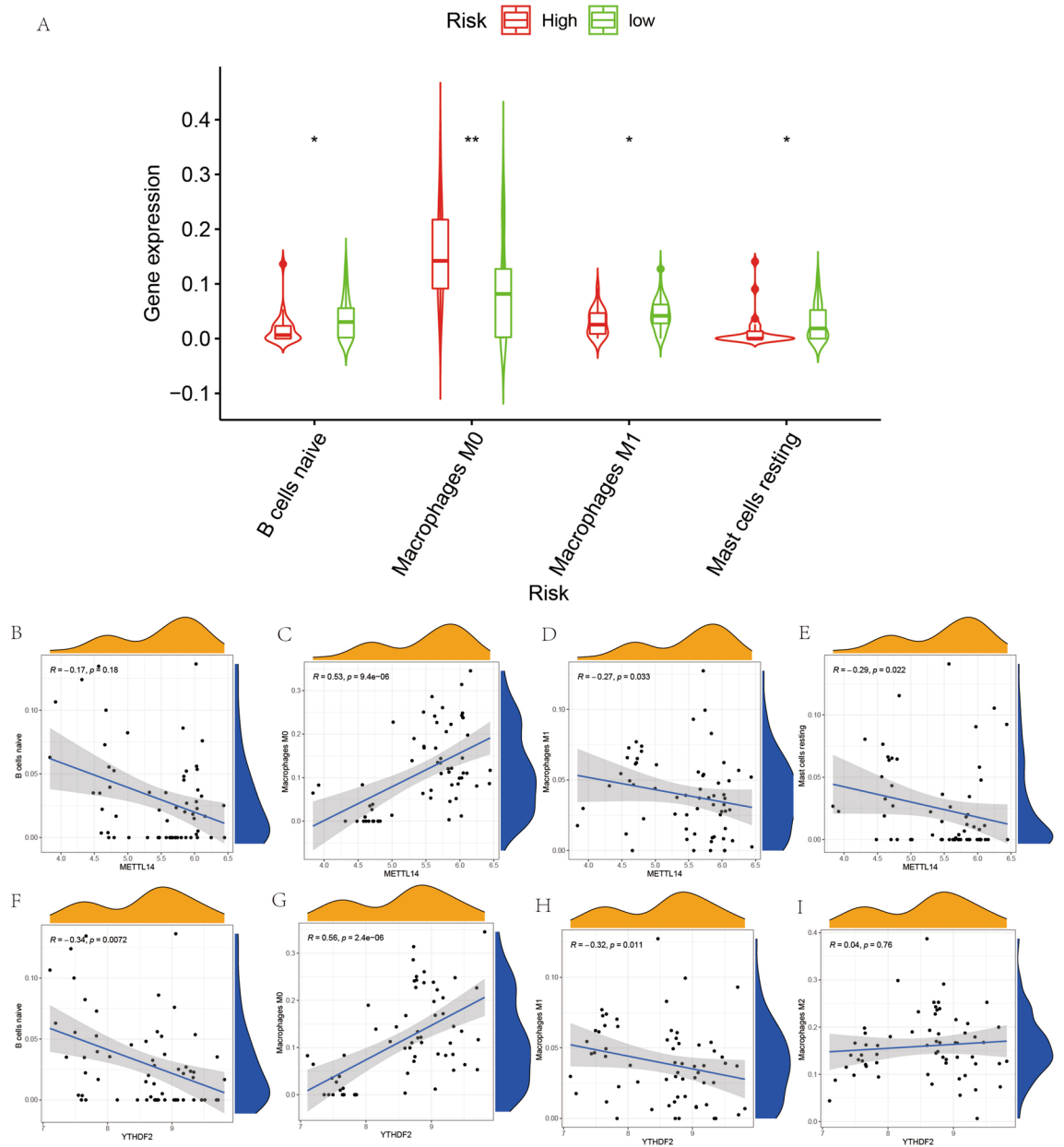


Figure 9. Expression of immune cells in Ewing’s sarcoma and paraneoplastic tissues. Figure A shows the expression of four immune cells in Ewing’s sarcoma and paraneoplastic tissues, “*” indicates $P < 0.05$, “**” indicates $P < 0.01$. (B–E) represent the correlation analysis obtained between METTL14 and this B cells naive, Macrophages M0, Macrophages M1 and Mast cells resting. (F–H) and (G) represent the correlation analysis obtained between YTHDF2 and this B cells naive, Macrophages M0, Macrophages M1 and Mast cells resting.

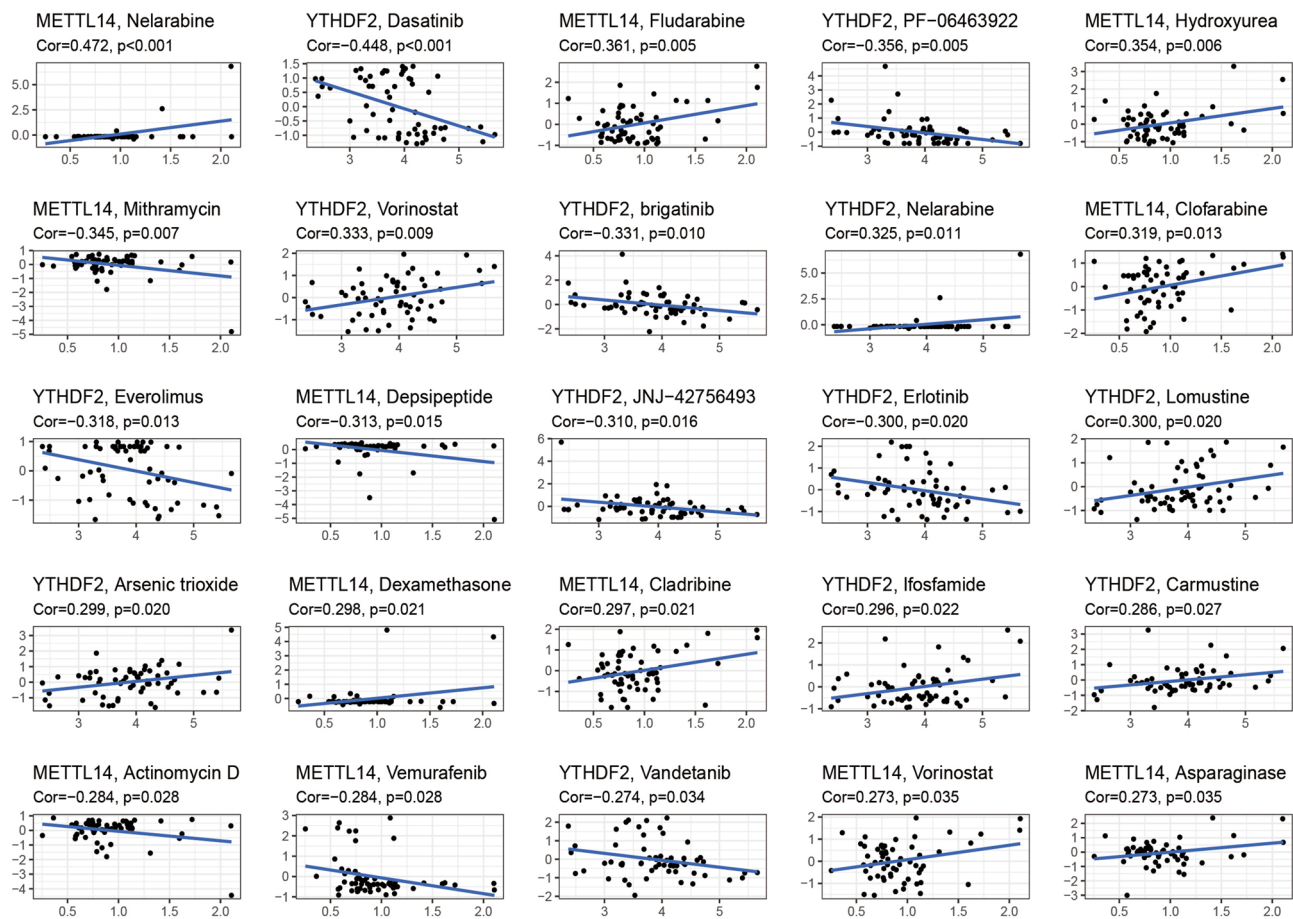


Figure 10. Drug sensitivity analysis. The figure shows the relationship between METTL14 and YTHDF2 each with different drug sensitivities. If $Cor > 0$, the high expression of this gene showed a positive correlation with drug sensitivity. If $Cor < 0$, the high expression of this gene showed a negative correlation with drug sensitivity.

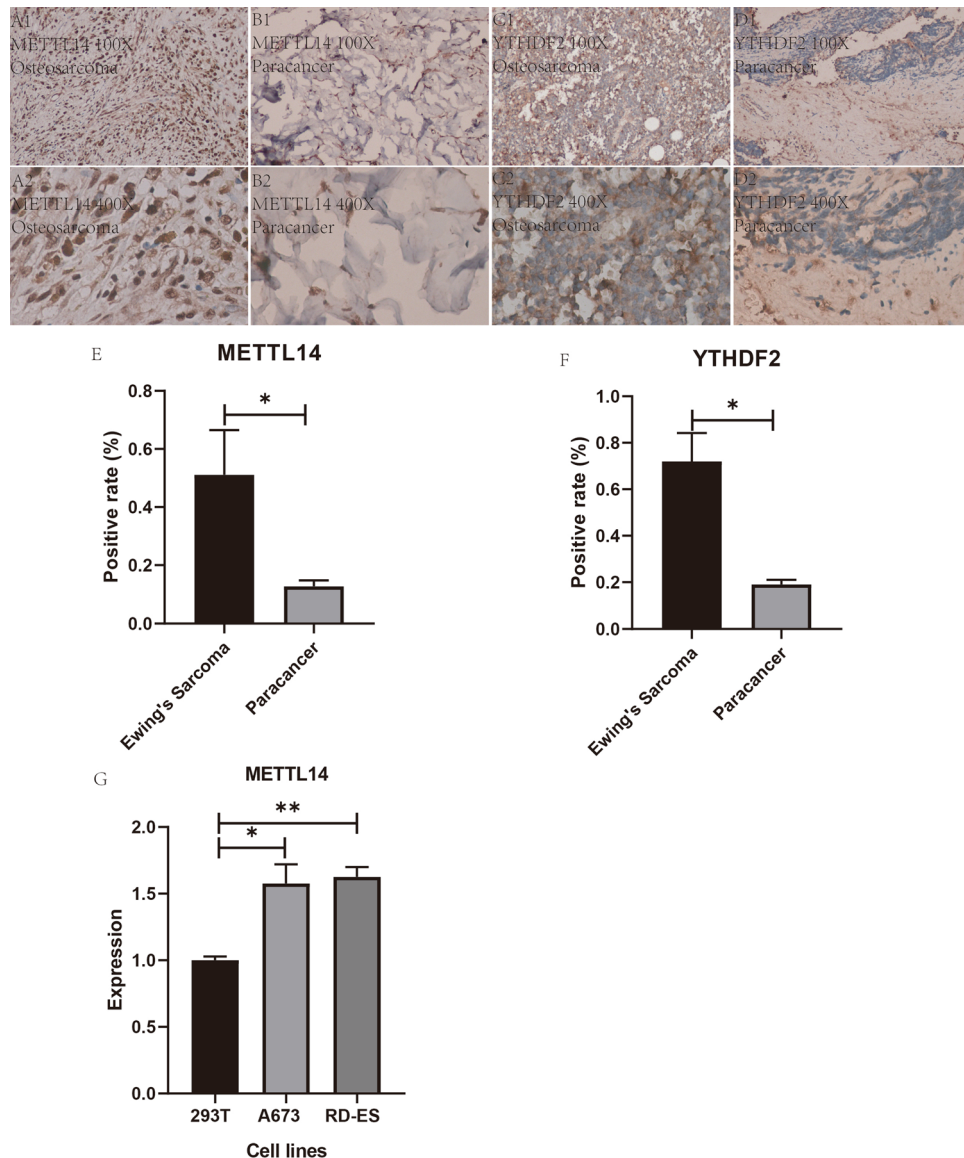


Figure 11. Immunohistochemistry, positivity statistics and qRT-PCR results. (A1, A2, B1, B2) represent immunohistochemical specific staining maps of METTL14 in Ewing's sarcoma tissue and in paraneoplastic tissue at 100× magnification and 400× magnification. (C1, C2, D1, D2) represent immunohistochemical specific staining maps of YTHDF2 in Ewing's sarcoma tissue and in paraneoplastic tissue at 100× magnification and 400× magnification. (E) indicates the statistical plot of the positive rate of immunohistochemistry for METTL14. Figure F represents the statistical graph of the positive rate of immunohistochemistry for YTHDF2, "*" indicates $P < 0.05$. (G) shows the expression of METTL14 in two Ewing's sarcoma cell lines and one control cell line, which can be found to be significantly different in ES and control. "*" indicates $P < 0.05$. "**" indicates $P < 0.01$.

Received: 22 September 2021; Accepted: 24 January 2022
Published online: 29 April 2022

References

- Renzi, S., Anderson, N. D., Light, N. & Gupta, A. Ewing-like sarcoma: an emerging family of round cell sarcomas. *J Cell Physiol* **234**, 7999–8007. <https://doi.org/10.1002/jcp.27558> (2019).
- Morales, E. *et al.* Role of immunotherapy in Ewing sarcoma. *J. Immunotherapy Cancer* <https://doi.org/10.1136/jitc-2020-000653> (2020).
- Kondo, T. Current status and perspectives of patient-derived models for Ewing's sarcoma. *Cancers (Basel)* <https://doi.org/10.3390/cancers12092520> (2020).
- Ozaki, T. Diagnosis and treatment of Ewing sarcoma of the bone: a review article. *J. Orthop. Sci.* **20**, 250–263. <https://doi.org/10.1007/s00776-014-0687-z> (2015).
- Nacev, B. A. *et al.* The epigenomics of sarcoma. *Nat. Rev. Cancer* **20**, 608–623. <https://doi.org/10.1038/s41568-020-0288-4> (2020).

6. Ma, S. *et al.* The interplay between m6A RNA methylation and noncoding RNA in cancer. *J. Hematol. Oncol.* **12**, 121. <https://doi.org/10.1186/s13045-019-0805-7> (2019).
7. Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol* **20**, 608–624. <https://doi.org/10.1038/s41580-019-0168-5> (2019).
8. He, L. *et al.* Functions of N6-methyladenosine and its role in cancer. *Mol Cancer* **18**, 176. <https://doi.org/10.1186/s12943-019-1109-9> (2019).
9. Lan, Q. *et al.* The critical role of RNA m(6)A methylation in cancer. *Cancer Res* **79**, 1285–1292. <https://doi.org/10.1158/0008-5472.Can-18-2965> (2019).
10. Gajewski, T. F., Schreiber, H. & Fu, Y. X. Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* **14**, 1014–1022. <https://doi.org/10.1038/ni.2703> (2013).
11. Zuo, S., Wei, M., Wang, S., Dong, J. & Wei, J. Pan-Cancer Analysis of immune cell infiltration identifies a prognostic immune-cell characteristic score (ICCS) in lung adenocarcinoma. *Front. Immunol.* **11**, 1218. <https://doi.org/10.3389/fimmu.2020.01218> (2020).
12. Baba, Y. *et al.* Tumor immune microenvironment and immune checkpoint inhibitors in esophageal squamous cell carcinoma. *Cancer Sci.* **111**, 3132–3141. <https://doi.org/10.1111/cas.14541> (2020).
13. Deng, X. *et al.* Profiles of immune-related genes and immune cell infiltration in the tumor microenvironment of diffuse lower-grade gliomas. *J Cell Physiol* **235**, 7321–7331. <https://doi.org/10.1002/jcp.29633> (2020).
14. Ringel, A. E. *et al.* Obesity shapes metabolism in the tumor microenvironment to suppress anti-tumor immunity. *Cell* **183**, 1848–1866.e1826. <https://doi.org/10.1016/j.cell.2020.11.009> (2020).
15. Zhang, Q. F. *et al.* CDK4/6 inhibition promotes immune infiltration in ovarian cancer and synergizes with PD-1 blockade in a B cell-dependent manner. *Theranostics* **10**, 10619–10633. <https://doi.org/10.7150/thno.44871> (2020).
16. Savola, S. *et al.* High expression of complement component 5 (C5) at tumor site associates with superior survival in Ewing's sarcoma family of tumour patients. *ISRN Oncol* **2011**, 168712. <https://doi.org/10.5402/2011/168712> (2011).
17. Volchenboum, S. L. *et al.* Gene expression profiling of ewing sarcoma tumors reveals the prognostic importance of tumor-stromal interactions: a report from the children's oncology group. *J. Pathol. Clin. Res.* **1**, 83–94. <https://doi.org/10.1002/cjp.2.9> (2015).
18. McEligot, A. J., Poynor, V., Sharma, R. & Panangadan, A. Logistic LASSO regression for dietary intakes and breast cancer. *Nutrients* <https://doi.org/10.3390/nu12092652> (2020).
19. Lin, Q., Zhao, Z. & Liu, J. S. Sparse sliced inverse regression via lasso. *J. Am. Stat. Assoc.* **114**, 1726–1739. <https://doi.org/10.1080/01621459.2018.1520115> (2019).
20. Page, R., Yoshida, R. & Zhang, L. Tropical principal component analysis on the space of phylogenetic trees. *Bioinformatics* **36**, 4590–4598. <https://doi.org/10.1093/bioinformatics/btaa564> (2020).
21. Ali, A., Margetts, B. M. & Zainuddin, A. A. Exploration of the principal component analysis (PCA) approach in synthesizing the diet quality of the Malaysian population. *Nutrients* <https://doi.org/10.3390/nu13010070> (2020).
22. Codesido, S. *et al.* Network principal component analysis: a versatile tool for the investigation of multigroup and multiblock datasets. *Bioinformatics* **37**, 1297–1303. <https://doi.org/10.1093/bioinformatics/btaa954> (2021).
23. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* **1711**, 243–259. https://doi.org/10.1007/978-1-4939-7493-1_12 (2018).
24. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457. <https://doi.org/10.1038/nmeth.3337> (2015).
25. Zhou, R. *et al.* Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I–III colon cancer. *Cancer Immunol Immunother: CII* **68**, 433–442. <https://doi.org/10.1007/s00262-018-2289-7> (2019).
26. Luna, A. *et al.* CellMiner cross-database (CellMinerCDB) version 1.2: exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Res* **49**, D1083–d1093. doi:<https://doi.org/10.1093/nar/gkaa968> (2021).
27. Zhang, Y. *et al.* Identification of prognostic miRNAs associated with immune cell tumor infiltration predictive of clinical outcomes in patients with non-small cell lung cancer. *Front Oncol* **11**, 705869. <https://doi.org/10.3389/fonc.2021.705869> (2021).
28. Yang, G., Sun, Z. & Zhang, N. Reshaping the role of m6A modification in cancer transcriptome: a review. *Cancer Cell Int.* **20**, 353. <https://doi.org/10.1186/s12935-020-01445-y> (2020).
29. Anita, R., Paramasivam, A., Priyadharsini, J. V. & Chitra, S. The m6A readers YTHDF1 and YTHDF3 aberrations associated with metastasis and predict poor prognosis in breast cancer patients. *Am. J. Cancer Res.* **10**, 2546–2554 (2020).
30. Wang, Q. *et al.* METTL3-mediated m(6)A modification of HDGF mRNA promotes gastric cancer progression and has prognostic significance. *Gut* **69**, 1193–1205. <https://doi.org/10.1136/gutjnl-2019-319639> (2020).
31. Wang, M. *et al.* Upregulation of METTL14 mediates the elevation of PERP mRNA N(6) adenosine methylation promoting the growth and metastasis of pancreatic cancer. *Mol Cancer* **19**, 130. <https://doi.org/10.1186/s12943-020-01249-8> (2020).
32. Weng, H. *et al.* METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA m(6) A modification. *Cell Stem Cell* **22**, 191–205.e199. <https://doi.org/10.1016/j.stem.2017.11.016> (2018).
33. Liu, J. *et al.* m(6)A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nat Cell Biol* **20**, 1074–1083. <https://doi.org/10.1038/s41556-018-0174-4> (2018).
34. Seifert, M. & Küppers, R. Human memory B cells. *Leukemia* **30**, 2283–2292. <https://doi.org/10.1038/leu.2016.226> (2016).
35. Ivashkiv, L. B. IFN γ : signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. *Nat Rev Immunol* **18**, 545–558. <https://doi.org/10.1038/s41577-018-0029-z> (2018).
36. Wen, P. *et al.* Pan-cancer analysis of radiotherapy benefits and immune infiltration in multiple human cancers. *Cancers (Basel)* **12**, doi:<https://doi.org/10.3390/cancers12040957> (2020).
37. Rahal, O. M. *et al.* Blocking interleukin (IL)4- and IL13-mediated phosphorylation of STAT6 (Tyr641) decreases M2 polarization of macrophages and protects against macrophage-mediated radioresistance of inflammatory breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **100**, 1034–1043. <https://doi.org/10.1016/j.ijrobp.2017.11.043> (2018).
38. Zhang, C. *et al.* YTHDF2 promotes the liver cancer stem cell phenotype and cancer metastasis by regulating OCT4 expression via m6A RNA methylation. *Oncogene* **39**, 4507–4518. <https://doi.org/10.1038/s41388-020-1303-7> (2020).
39. Dixit, D. *et al.* The RNA m6A reader YTHDF2 maintains oncogene expression and is a targetable dependency in glioblastoma stem cells. *Cancer Discov.* **11**, 480–499. <https://doi.org/10.1158/2159-8290.Cd-20-0331> (2021).

Acknowledgements

The authors are grateful to Dr. Xin Li Zhan (Spine and Osteopathy Ward, The First Affiliated Hospital of Guangxi Medical University) for his kindly assistance in all present study stages.

Author contributions

J.J., Q.F., H.Q., C.L. and X.Z. designed the study. T.L., L.C., S.H., X.S. analyze the data. J.C., T.C., H.L. and Y.Y. digital visualization. J.J. wrote and revised the manuscript. J.J., Q.F., H.Q., and C.L. performed the laboratory operations. C.L. and X.Z. revised the manuscript. All co-authors participated in the laboratory operation. All authors read and approved the final manuscript.

Funding

This study was supported by the Youth Science Foundation of Guangxi Medical University, Grant/Award Numbers: GXMUYFY201712; Guangxi Young and Middle-aged Teacher's Basic Ability Promoting Project, Grant/Award Number: 2019KY0119.

Competing interests

There were no conflicts of interest between all co-authors of this study. The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

Additional information

Correspondence and requests for materials should be addressed to X.Z.

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

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



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

© The Author(s) 2022