Integrated Analysis of Multi-Omics Data to Establish a Hypoxia-Related Prognostic Model in Osteosarcoma

Ye Tong1, Xiaoqing Zhang2 and Ye Zhou1

1Department of Orthopaedics, Suzhou Hospital of Anhui Medical University, Suzhou, Anhui, China. 2Department of Laboratory, Bozhou People’s Hospital, Bozhou, Anhui, China.

ABSTRACT

BACKGROUND: Osteosarcoma (OS) is the most common malignant bone tumor in clinical practice, and currently, the ability to predict prognosis in the diagnosis of OS is limited. There is an urgent need to find new diagnostic methods and treatment strategies for OS.

MATERIAL AND METHODS: We downloaded the multi-omics data for OS from the TARGET database. Prognosis-associated methylation sites were used to identify clustered subtypes of OS, and OS was classified into 3 subtypes (C1, C2, C3). Survival analysis showed significant differences between the C3 subtype and the other subtypes. Subsequently, differentially expressed genes (DEGs) across subtypes were screened and subjected to pathway enrichment analysis.

RESULTS: A total of 249 DEGs were screened from C3 subtype to other subtypes. Metabolic pathway enrichment analysis showed that DEGs were significantly enriched to the hypoxic pathway. Based on univariate and multivariate COX regression analysis, 12 genes from the hypoxia pathway were further screened and used to construct a hypoxia-related prognostic model (HRPM). External validation of the HRPM was performed on the GSE21257 dataset. Finally, differences in survival and immune infiltration between high and low risk score groups were compared.

CONCLUSION: In summary, we proposed a hypoxia-associated risk model based on a 12-gene expression signature, which is potentially valuable for prognostic diagnosis of patients with OS.

KEYWORDS: Osteosarcoma, biomarkers, hypoxia, prognostic model, methylation

Background

Osteosarcoma is a highly aggressive malignancy that is often fatal if not detected and treated early.1 The 5-year survival rate of patients with osteosarcoma without distant metastases has now been increased to 60–70% percent in clinical practice using a mix of chemotherapy and surgical treatment options.2 However, for patients with metastatic osteosarcoma, the 5-year survival rate is less than 30%.3 The capacity to predict prognosis at the time of osteosarcoma diagnosis is currently limited. At the moment of diagnosis, the presence or absence of metastases is the most important predictor of outcome.4 However, markers of metastasis are absolutely meaningless in patients with just localized disease. As a result, in order to achieve further progress in treatment and prognosis, more generalized predictive methods are needed.

Hypoxia is a characteristic of the great majority of solid tumors, in which tumor cell proliferation results in hypoxia due to an aberrant lack of blood supply to the tumor microvasculature.5 The absence of oxidative phosphorylation in a chronic hypoxic environment causes a drop in ATP, forcing cells to employ anaerobic respiration as a method of survival, resulting in an accumulation of lactate in the cells and a decrease in pH.6 Although extreme hypoxia causes tumor necrosis, moderate hypoxia close to the tumor center promotes tumor angiogenesis, cancer cell survival, and stem cell characteristics, encouraging tumor development, metastasis, and medication resistance.7 DNA methylation plays a significant role in defining tissue-specific gene expression and chromosomal instability through epigenetic regulation of gene expression.8 Overall hypomethylation plays a key function in the tumor cell hypoxia response system to adapt to hypoxia.9 Aberrant methylation has been connected to carcinogenesis, and in many tumor forms, variations in methylation patterns between malignancies and benign tissues have been observed.10 Several biomarkers related with DNA methylation have been revealed for early cancer diagnosis or prognostic prediction due to the highly conserved feature of DNA methylation.11–14

In this study, we identified osteosarcoma methylation subtypes using prognostic methylation loci and further screened for differentially expressed genes (DEGs) between subtypes. Metabolic pathway enrichment analysis of the DEGs indicated that the hypoxic metabolic pathway was significantly enriched. Accordingly, we constructed a hypoxia-related prognostic model (HRPM) based on hypoxia pathway genes.
Material and Methods

Data acquisition

We downloaded gene expression, methylation and copy number variation (CNV) data for OS from the TARGET database (https://oarc.cancer.gov/programs/target). Clinical information of OS patients was also obtained from the TARGET database, and samples lacking survival information were excluded. Next, we matched only the sample labels shared between gene expression data, methylation data, CNV data and clinical data, and finally retained 83 OS samples. The GSE21257 dataset contains 53 samples of pre-chemotherapy osteosarcoma patients, we obtained their gene expression matrix and corresponding clinical information from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo) database, which were used for model validation.

Osteosarcoma methylation subtype clustering analysis

Before clustering, we first performed univariate Cox regression analysis of the methylation levels (Beta values) of all CpG sites in the DNA methylation data of 83 patients. CpG sites associated with patient prognosis were retained (P value < .05). Next, unsupervised consensus clustering of these prognosis-related CpG sites was performed based on the K-means algorithm using the ConsensusClusterPlus R package.15 The t-SNE method was performed using the Rtsne package to explore the distribution of different clusters.16 Subsequently, the relationship between clustered subtypes and overall patient survival was analyzed using patient-matched survival data using the survival R package.

Subtype immune infiltration analysis

We applied 3 independent algorithms to assess the abundance of tumor microenvironment (TME) cell infiltration. The CIBERSORT R package quantified the abundance of 22 immune cell types;17 the MCPCounter algorithm was used to identify 10 immune cell lineages;18 and the stromal score, immune score and tumor purity were calculated for all samples by the ESTIMATE algorithm.19 In addition, we calculated the DNA methylation of tumor-infiltrating lymphocytes (MeTIL) for all samples based on the PCA method.20 Finally, we compared the differences in immune infiltration abundance, immune scores and MeTIL in different subtypes.

Copy number variation and immunotherapy analysis

Whole genome gene copy number variations (CNV) profiles and chromosome arm/gene-level variations were detected by GISTIC 2.0 (http://archive.broadinstitute.org/cancer/cga/gistic).21 The focal length cutoff was set to 0.5, the confidence interval was set as 0.9, and all other operational parameters were set to their default values. Meanwhile, we used OncoPrint plot to show the overview of CNVs in each OS subtype. To assess the probability of response to immunotherapy in individuals with OS, the Tumor Immune Dysfunction and Exclusion (TIDE) (http://tide.dfci.harvard.edu/) algorithm was used to predict the response to immunotherapy in each patient. We further compared the differences in immunotherapy between patients of different subtypes.

Differential expression analysis and enrichment analysis

Based on the results of the subtype survival analysis, we screened for differential genes (DEGs) between poor prognosis subtypes and good prognosis subtypes. Differential expression analysis was performed using the limma package22 and the screening thresholds were set as follows: adjusted P-value less than .05 and absolute value of Log2 fold change greater than 1. Gene set enrichment analysis (GSEA) of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) in the ranked list of DEGs were performed using the clusterProfiler R package.23 Adjusted P value < .05 was considered statistically significant.

Construction and validation of a hypoxia-related prognostic model

Based on the results of the pathway enrichment analysis, we acquired hypoxia-related genes (HRGs) from the GSEA (https://www.gsea-msigdb.org/gsea/index.jsp) website (hallmark gene sets: h.all.v7.5.1.entrez.gmt). A total of 200 HRGs were obtained. Subsequently, univariate Cox regression analysis was performed on all HRGs to identify hypoxia-related prognostic genes associated with survival (P value < .05). Further, we used hypoxia-related prognostic genes to establish a multivariate Cox regression model, and based on the bidirectional stepwise regression method, we finally established a hypoxia-related prognostic model (HRPM). Additionally, Risk scores were calculated for each patient by using the formula of HRPM. All patients were then divided into a high-risk group and a low-risk group based on the median risk score. Survival plot, time-dependent receiver operating characteristic (ROC) curve, and patient risk heat map were plotted separately for HRPM. The GSE21257 dataset was used to validate the HRPM. Furthermore, we performed a multifactorial independent prognostic analysis of risk scores and clinical information including gender, age, and metastatic status. Finally, we compared the differences between high and low risk groups in terms of immune cell infiltration and immune score.

Results

Tumor methylation subtype clustering analysis

From a total of 385291 methylation sites, we identified 33519 potential prognostic methylation sites. As shown in Figure 1B,
it shows a graph of the consensus cumulative distribution function (CDF) plot for each cluster to help the user determine the choice of the number of subtypes when the CDF reaches a maximum, at which point the consistency and clustering confidence is maximized. In addition, Figure 1C is used to represent the relative change in area under CDF curve for $k=2-6$. (D) TSNE plot shows the distribution of the 3 clustered subtypes. (E) Each DNA methylation subtype’s survival curves.

**The tumor microenvironment and immune cell infiltration**

The immune cell components infiltrating the OS samples were analyzed using the CIBERSORT method and the MCPcounter algorithm, respectively (Figure 2A and C). Meanwhile, in the TARGET cohort, differences in the abundance of 3 subtypes of immune cell infiltration were estimated. The infiltration abundance of dendritic cells resting and macrophages M1 was considerably different among the 3 subtypes, according to the findings based on CIBERSORT (Figure 2B). Furthermore, with the help of the MCPcounter algorithm, we found that CD8+ T cells, fibroblasts and neutrophils were significantly differentially infiltrated in subtypes (Figure 2D). The C3 subtype had the lowest stromal score and the highest tumor purity, according to the ESTIMATE algorithm, whereas the immune score did not differ substantially among the 3 subtypes (Figure 3B, C, and D). We derived MeTIL score using the methylation signature of tumor-infiltrating lymphocytes, and integrated it with immune checkpoint and the immune microenvironment to map the tumor subtype immune landscape (Figure 3A). Additionally, the C3 subtype had the lowest MeTIL score of all the subtypes (Figure 3E).

**CNV and TIDE analysis**

Raw CNV data from 83 samples were analyzed with GISTIC 2.0, and the locations of significant amplifications and deletions were visualized for each chromosome (Figure 4A and B). Simultaneously, we plotted the gistic scores of overall copy
Figure 2. Analysis of immune cell infiltration in osteosarcoma subtypes. (A) CIBERSORT shows the infiltration proportion of 22 immune cell subpopulations. (B) MCPCounter shows the infiltration proportion of 10 immune cell subpopulations. (C) Differences in immune cell infiltration of 22 immune cell types in 3 tumor subtypes estimated using the CIBERSORT algorithm. (D) Estimation of differences in immune cell infiltration of 10 immune cell types in 3 tumor subtypes using the MCPCounter algorithm. ns: no significant; * P < 0.05; ** P < 0.01; *** P < 0.001.
Figure 3. Comparison of immune microenvironment among osteosarcoma subtypes. (A) Immune landscape of osteosarcoma subtypes. (B) Differences in immune scores among osteosarcoma subtypes. (C) Differences in stromal scores among osteosarcoma subtypes. (D) Differences in tumor purity among osteosarcoma subtypes. (E) Differences in MeTIL scores among osteosarcoma subtypes. ns: no significant; *P < .05. **P < .01. ***P < .001.
number on the chromosomes (Figure 4C). In addition, we found that 14q11.2 had the highest amplification frequency, while 13q14.2 had the highest deletion frequency (Figure 4D). The TIDE algorithm was used to predict the responsiveness to immunotherapy in 3 subtypes of osteosarcoma. The results showed that C3 subtypes responded better to immunotherapy than C1 and C2 (Figure 5A). Consistent with the expected results, C3 had lower TIDE scores, dysfunction scores and exclusion scores than the other subtypes (Figure 5B, C, and D). Low TIDE scores indicate low tumor immune dysfunction, low immune escape capacity and therefore high immunotherapy response.

**Differential gene screening and analysis**

The above results suggest that the C3 subtype has a poorer prognosis and a better response to immunotherapy, thus we compared it to other subtypes (C1 and C2) and screened for DEGs. A total of 249 DEGs were identified, with 146 genes that were down-regulated and 103 genes that were up-regulated (Figure 6A). Following that, we used GSEA to conduct gene ontology and metabolic pathway enrichment analyses of the DEGs. The GO enrichment analysis results for biological processes showed that GO items such as translational initiation (GO:0006413), extracellular matrix organization (GO:0030198) and protein targeting to ER (GO:0045047) were enriched (Figure 6B). In the metabolic pathway enrichment analysis, we found that the hypoxia pathway (NES: 1.65) was activated while the oxidative phosphorylation pathway (NES: −1.78) was inhibited (Figure 6C).

**Hypoxia-related prognostic model construction and evaluation**

A total of 29 hypoxia-related prognostic genes were screened from 200 hypoxia-related genes by univariate COX regression. Subsequently, these genes were used to construct a multivariate COX regression model, and a HRPM including 12 genes was finally obtained by a bidirectional stepwise regression method (Figure 7A). The risk score for the HRPM can be calculated using the following formula: risk score = ANXA2 * 0.203 + CASP6 * 0.366 + CAVIN1 * 2.486 + DCN * 0.671 + FAM162A * 4.871 + HMOX1 * 0.674 + MAFF * 2.098 + PDK1 * 0.282 + RBPJ * 0.431 + SDC3 * 0.586 + STC2 * 1.463 + TES * 0.566. Based on the median risk score, patients were divided into high-risk and low-risk categories.
Survival risk heat map showed that higher mortality in OS patients with higher risk scores (Figure 7B). The survival analysis revealed that the mortality rate in the high-risk score group was considerably greater than in the low-risk score group (Figure 7C). The AUCs for 1-year, 2-year, and 3-year of their time-dependent ROC curves were 0.939, 0.931, and 0.909, respectively (Figure 7E). The HPRM was subsequently validated in the GSE21257 cohort, and we observed that the model performed well in the validation set (Figure 7D and F). Furthermore, a multivariate regression model was built using the patient’s clinical information and risk score, and the result indicate that risk score was an independent prognosis factor (Figure 8A).

**Discussion**

Previous studies have shown that OS is closely associated with aberrant genetic and epigenetic changes, leading to abnormal expression of oncoproteins or methylation of antioncogenes.24 Meanwhile, a large number of studies have confirmed the reliability of clustering subtype analysis of solid tumors based on methylation sites.25-27 In this study, 33519 methylation sites were found to be statistically associated with prognosis in OS patients. Subsequently, we categorized OS into 3 clustering subtypes based on prognosis-related methylation sites, with the C3 subtype possessing the poorest prognosis of the 3 subtypes, according to the survival analysis. Meanwhile, we observed that the C3 subtype had the lowest CD8 T cell infiltration, but the highest macrophage M1 infiltration. Immune infiltration in the tumor serves a crucial function in tumor progression.28 Previous studies have linked an increase in the number of CD8 T cells infiltration tumors to a favorable anti-tumor immune response, suggesting that it could be used as a prognostic indicator,29,30 and that CD8 T lymphocyte infiltration improves the prognosis of patients with osteosarcoma.31 Macrophages, which are polarized M1 and M2 macrophages, are a prominent component of the tumor microenvironment.32 The formation of the tumor microenvironment encourages M1 to M2 conversion. Early tumor tissue is dominated by M1 macrophages, which limit angiogenesis and stimulate tumor immunity.33 Contrary to expectations, we discovered that M1 macrophages of the C3 subtype exhibited the highest infiltration abundance among the 3 subtypes. In osteosarcoma, Buddingh et al. discovered that macrophages’ antitumor efficacy outweighed their
probable tumor-supporting role. Tumor-infiltrating lymphocytes (TILs) have been shown to have a profound impact on the prognosis of a variety of cancers. Previous studies have utilized the methylation signature of TILs compared to gene expression-based immune markers to measure TILs distribution and to predict survival and tumor immune response. Therefore, we evaluated the differences in MeTIL scores among the 3 osteosarcoma subtypes. We observed significantly lower scores for the C3 subtype than for the other 2 subtypes. In general, elevated levels of tumor-infiltrating lymphocytes are thought to be associated with a better prognosis. In particular, the synergistic effect of tumor stromal and TILs has an enhanced prognostic impact. Based on the TIDE website, we evaluated the efficacy of immunotherapy against 3 OS subtypes, and the results suggested that the C3 subtype responded well to immunotherapy. TIDE has shown potential in published clinical trials for predicting patient response to immunotherapy, but more studies are needed to prove its utility in OS.

Next, we screened for genes that were differentially expressed between the C3 subtype and other subtypes. The pathway...
Figure 7. Construction and validation of a hypoxia-related prognostic model. (A) Heat map of the differential gene expression patterns. (B) GO enrichment analysis of differential gene expression. (C) Survival curves for different risk score groups in the training set. (D) Survival curves for different risk score groups in the validation set. (E) Time-dependent overall survival ROC for training set (TARGET). (F) Time-dependent overall survival ROC for validation set (GSE21257).
enrichment analysis suggested that these DEGs were substantially related with hypoxia pathway activation and oxidative phosphorylation pathway inhibition. Studies have shown that patients with lower oxygen levels are not only more tolerant to radiation than patients with higher oxygen levels, but also have a greater likelihood of recurrence. Hypoxia is also linked to chemotherapeutic resistance. Anti-angiogenic medications cut off the tumor’s blood supply, causing or exacerbating a hypoxic microenvironment in the tumor, which is the primary cause of resistance to anti-angiogenic therapy after the first response, resulting in treatment failure. This treatment-induced hypoxia has been proven to be the catalyst for secondary anti-vegf therapeutic resistance. Thus, we speculate that patients with C3 subtype have a worse prognosis, which could be connected to chemotherapy resistance due to increased hypoxia pathway activation.

We obtained 200 hypoxia pathway genes from the MSigDB database, and screened 12 genes by univariate and multivariate COX regression analysis to construct HRPM and calculate the risk scores of the patients. According to the median risk score, patients were separated into high and low risk groups, and we found that patients in the high risk group had a lower overall survival rate. Furthermore, the ROC analysis revealed that the risk score model is highly reliable. We externally validated the model in the GSE21257 dataset and also obtained more satisfactory results. In a multivariate regression analysis, we included age, sex, metastasis, and risk score, and the result indicated that risk score could be an independent prognostic factor. Notably, CD8 T cell infiltration was markedly lower in individuals with high risk scores than in patients with low risk scores. Furthermore, patients in the high-risk group had lower immune score and higher tumor purity than those in the low-risk group.

Overall, in this study, we proposed a novel 12-gene signature for OS prognosis encompassing ANXA2, CASP6, CAVIN1, DCN, FAM162A, HMOX1, MAFF, PDK1, RBPJ, SDC3, STC2, and TES, which can be used to predict high-risk groups of patients with osteosarcoma. The contribution of HRPGs as biomarkers of tumor progression as well as potential therapeutic targets might be confirmed by further studies.

Conclusions
Overall, we constructed a hypoxia-associated gene signature based on multi-omics data to assess the prognosis of patients with OS. It may serve as a potential prognostic marker for the prognosis diagnosis of OS patients.
Acknowledgements
We thank the TCGA, GEO, and TARGET databases.

Declaration of Figures’ Authenticity
All figures submitted have been created by the authors who confirm that the images are original with no duplication and have not been previously published in whole or in part. All of the figures were drawn by the first author, Ye Tong.

ORCID iD
Ye Tong https://orcid.org/0000-0003-1044-7984

REFERENCES
1. Leichter AL, Sullivan MJ, Eccle MR, et al. MicroRNA expression patterns and signalling pathways in the development and progression of childhood solid tumours. *Mutat Cancer 2017;16:15.
2. Zhang Y, Wang F, Li M, et al. Self-stabilized hyaluronate nanogel for intra-cellular delivery of doxorubicin and cisplatin to osteosarcoma. *Adv Sci 2018;5:1700821.
3. Zhang K, Dong C, Chen M, et al. Extracellular vesicle-mediated delivery of miR-101 inhibits lung metastasis in osteosarcoma. *Thrombosis 2020;10:411-425.
4. Wang X, Qiao G, Liang X, et al. Targeting the CKB/CBX4 axis for metastasis in osteosarcoma. *Nat Commun 2020;11:1141. doi:10.1038/s41467-020-14870-4
5. Qiao Y, Yang F, Xie T, et al. Engineered algae: a novel oxygen-generating system for effective treatment of hypoxic cancer. *Sci Adv 2020;6:eaba5996.
6. Pijlman B, Ksavc V, Dzhamsane R, et al. Cancer etiology: a metabolic disease originating from life’s major evolutionary transition? *Oxid Med Cell Longev 2019;2019:1-16.
7. Han F, Li C-F, Cai Z, et al. The critical role of AMPK in driving Akt activation under stress, tumorigenesis and drug resistance. *Nat Commun 2018;9:4728.
8. Alam H, Tang M, Maitituoheti M, et al. KMT2D deficiency impairs super-enhancers to confer a glycolytic vulnerability in lung cancer. *Cancer Cell 2020;37:599-617.e7.
9. Zhu J, Wang K, Li T, et al. Hypoxia-induced TET1 facilitates trophoblast cell migration and invasion through HIF1α signaling pathway. *Sci Rep 2017;7:8077.
10. Zhao SG, Chen WS, Li H, et al. The DNA methylation landscape of advanced prostate cancer. *Nat Genet 2020;52:778-789.
11. Deng Y, Yuan W, Ren E, et al. A four-methylated LncRNA signature predicts survival of osteosarcoma patients based on machine learning. *Genomics 2021;2013:113-785-794.
12. Rosenblum JM, Wijetunga NA, Fazzari MJ, et al. Predictive properties of DNA methylation patterns in primary tumor samples for osteosarcoma relapse status. *Epigenetics 2015;10:31-39.
13. Chung H-H, Kuo C-C, Hsiao C-W, et al. A novel prognostic DNA methylation panel for colorectal cancer. *Int J Mol Sci 2019;20:4672.
14. Lin R, Li C, Liu Z, et al. Genome-wide DNA methylation profiling identifies epigenetic signatures of gastric cardiac intestinal metaplasia. *J Transl Med 2020;18:292.
15. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics 2010;26:1572-1573.
16. Van Der Maaten L. Accelerating t-SNE using tree-based algorithms. *J Mach Learn Res 2014;15:3221-3245.
17. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets and their association with tumor immune regulatory gene expression in muscle invasive bladder cancer subtypes. *J Immunother Cancer 2019;7:148.
18. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *J Immunother Cancer 2015;3:257-275.
19. Bergers G, Hanahan D. Modes of resistance to anti-angiogenic therapy. *Nat Rev Cancer 2008;8:592-603.
20. Loges S, Schmidt T, Carmeliet P. Mechanisms of resistance to anti-angiogenic therapy. *Nat Rev Cancer 2017;17:2110-2119.
21. thibaudin M, Chaux B, Briasoulis E, et al. Metronomic vinorelbine: anti-angiogenic activity in vitro in normoxic and severe hypoxic conditions, and severe hypoxia-induced resistance to its anti-proliferative effect with reversal by Akt inhibition. *Int J Oncol 2015;47:455-464.