Identifying Potential Prognostic Expression Quantitative Trait Methylation for Predicting Survival in Osteosarcoma

Guofeng Zhang (gfzhangknu@gmail.com)  
Yantai Affiliated Hospital of Binzhou medical University

Pengfei Dong  
Yantai Hospital of Traditional Chinese Medicine

Yonglin Zhu  
Yantai Affiliated Hospital of Binzhou Medical University

Jingsheng Wang  
Yantai Affiliated Hospital of Binzhou Medical University

Fengmin An  
Yantai Affiliated Hospital of Binzhou Medical University

Qingpeng Shi  
Yantai Affiliated Hospital of Binzhou Medical University

Apeng Zou  
Yantai Affiliated Hospital of Binzhou Medical University

Guosheng Wang  
Yantai Affiliated Hospital of Binzhou Medical University

Shanhui Li  
Yantai Affiliated Hospital of Binzhou Medical University

Research article

Keywords: DNA methylation, osteosarcoma, eQTM, survival, biomarker

DOI: https://doi.org/10.21203/rs.3.rs-41529/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Osteosarcoma (OS) is a complex cancer and depends on multiple biological processes and pathways. Therefore, there is an urgent need to identify reliable prognostic biomarkers for predicting clinical outcomes and helping personalize treatment of OS.

**Results:** An expression Quantitative Trait Methylations (eQTMs)-based risk score model was developed to infer the prognostic efficacy for OS based on gene expression and methylation level. Some OS-specific eQTMs were identified and divided to positive and negative eQTMs. We found that some OS-specific eQTMs were significantly associated with survival in OS according to the eQTM-based risk score. The global characteristics for these prognosis-associated eQTMs were depicted and analyzed. These prognosis-associated eQTMs are stable in multiple experiments and showed better performance in predicting survival than single gene expression, age and gender of OS. Functional analysis indicated that the genes in prognosis-associated eQTMs were associated with essential biology processes of cancer development. In addition, these genes were also drug targets for many kinds of drugs and may provide assistance for studying drug repurposing of OS.

**Conclusion:** These results highlight the potential of eQTMs as novel diagnostic, prognostic and therapeutic targets for OS.

Background

Osteosarcoma (OS) derives from primitive bone-forming mesenchymal cells and is the most common primary bone malignancy[1]. OS belong to a large family of tumor entities of mesenchymal origin which exhibit heterogeneous histological, genetic, and molecular features[2]. Although advances have been scored in surgical techniques, multi-agent systemic chemotherapy, precise radiotherapy and immunotherapy, the 5-year survival rate of a localized tumor remains at 60–70%, while that of metastasis and recurrence is less than 20%[3]. Some studies had explored the molecular mechanism of prognosis and effective method for early diagnosis [3]. Identifying genetic biomarkers of patient survival remains a major goal of large-scale cancer profiling studies. More information of guide therapy for OS could be provided follow better comprehension of the prognostic variables, which contributes to prolonging survival and enhancing quality of life.

DNA methylation is the main form of epigenetic modifications, and the global and local changes to DNA methylation are a seminal feature of cancer cells [3]. The methylation of chromatin components is highly conserved as it helps coordinate the regulation of gene expression. Many earlier studies reported the significant associations between DNA methylation and gene expression (expression Quantitative Trait Methylations, eQTMs), and these associations were both positive (pos) and negative (neg)[3]. There were also some studies focused on the roles of eQTMs in Osteosarcoma. Shi et al. reported a novel interplay between HOTAIR and DNA methylation in osteosarcoma cells indicates a new therapeutic strategy[4]. Epigenetic silencing of the Wnt antagonist APCDD1 by promoter DNA hyper-methylation contributes to
OS cell invasion and metastasis were also reported [3]. However, global specific eQTMs and their roles in OS had not been widely identified and characterized.

There is an urgent need to identify reliable prognostic biomarkers for predicting clinical outcomes and personalizing treatment for OS. Inferring and predicting the survival by computational model is a key factor for studying the mechanism and therapy in OS. Some studies identify gene, microRNA and long non-coding RNA expression signatures that predict the survival risk [3]. DNA methylation is an independent prognostic marker of survival in cancer [3]. Guo et al. also suggested that a four-DNA methylation biomarker is a superior predictor of survival of patients with cutaneous melanoma[5]. Combined analysis of DNA methylation and gene expression profiles of OS also could identify several prognosis signatures[6]. However, if the eQTMs could become as potential prognostic signatures in OS had not been globally studied.

In the present study, we aimed to identify eQTMs and depict their roles in predicting the survival of OS patients. We first identified specific eQTMs by integrating gene expression and methylation level for OS. The eQTM-based risk score model was constructed to infer the prognostic efficacy of each gene-methylation relations. According to the eQTM-based risk score, we found that panels of same genes regulated by the diverse methylations were significantly associated with patient survival. The global characteristics for prognosis-associated eQTMS were described in OS. The 17 prognosis-associated eQTMS are stable in multiple experiments and showed better performance in predicting survival than single gene expression, age and gender of OS. Functional analysis showed that the genes in prognosis-associated eQTMs were associated with essential biology processes of cancer development. In addition, these genes were also drug targets for many kinds of drugs and may provide assistance for studying drug repurposing of OS. In summary, our findings revealed that the eQTM-based risk score model can provide helpful information on OS prognosis stratification and discovery of therapeutic biomarkers.

Methods

Clinical, gene expression and methylation profile dataset of OS

The gene expression, methylation profile and related clinical information of 86 OS patients were obtained from the TARGET (Therapeutically Applicable Research To Generate Effective Treatments, https://ocg.cancer.gov/programs/target) data portal, which included 17070 mRNAs and 34166 methylation sites. Target program applies a comprehensive genomic approach to determine molecular changes that drive childhood cancers. The goal of the program is to use data to guide the development of effective, less toxic therapies. The case selection criteria and sample details could be found at https://ocg.cancer.gov/programs/target/projects/osteosarcoma.

Identification Of Specific Eqtms For Os Patients
First, we mapped the methylation sites for each gene based on genomic sites and obtained some gene-methylation relation profiles. Second, only the genes with more than three methylation sites were extracted for follow-up analysis. Third, the Pearson correlation coefficients (PCCs) were calculated for each gene-methylation profile based on gene expression and methylation level profiles in OS. The gene-methylation relations were considered as specific eQTMs if the absolute values of PCCs were more than 0.3. The specific eQTMs were divided positive and negative eQTMs based on the direction of correlation between gene and methylations.

**Construction of a comprehensive eQTM-based risk score model to identify prognosis-related eQTMs in OS**

In order to identify prognosis-related eQTMs in OS, a comprehensive and computational pipeline was constructed based on survival, gene expression and methylation level data. First, the OS patients were randomly divided into two groups. One group was used as modeling group and another group was used as validation group. Second, a multivariate Cox regression model was performed for the methylations related with a same gene in a specific eQTM in modeling group. Third, the risk score for each patient was calculated according to the linear combination of the expression values weighted by the coefficient from multivariate Cox regression analysis:

\[
\text{risk score} = \sum_{i=1}^{n} \text{cox}_i \times \text{meth}_i
\]

where \(\text{cox}_i\) is the Cox regression coefficient of a methylation and \(n\) is the number of methylations regulated by the same gene. Methylation \((\text{meth}_i)\) is the methylation value of methylation site \(i\) in the corresponding patient. The median risk score was used as a cut-off point to divide the patients into high and low risk groups. Finally, Kaplan-Meier survival analysis was performed for the two groups, and statistical significance was assessed using the log-rank test. The survival results were considered significant when \(P < 0.05\). All analyses were performed within the R 3.3.3 framework.

**Evaluation of the performance about the eQTMs-based risk score model for prognosis in OS**

In order to evaluate the performance about the eQTMs-based risk score model for prognosis in OS, we performed eQTM-based risk score model ten times and extracted the prognosis-related eQTMs that were always significant. In addition, the OS patients were divided into two groups based on the median value of age and Kaplan-Meier survival analysis was performed for the two groups. The association between gender and survival was also validated by same way.

**Functional enrichment analysis for the genes in prognosis-related eQTMs of OS**

With the Enrichr tool online web server using default parameters, functional enrichment was performed for genes across core clusters [3]. We obtained enriched GO (Gene Ontology) terms \((P < 0.01, FDR < 0.05)\), KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways \((P < 0.05, FDR < 0.05)\).

**Drug repurposing candidates for the genes in prognosis-related eQTMs of OS**
The drug and gene targets information were obtained from DrugBank (version 5.1.1, released 2018-07-03), which is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information (https://www.drugbank.ca).

Results

Identification and characterization of specific eQTMs in OS

A integrated pipeline was constructed to identify specific eQTMs for OS (Fig. 1A). First, the methylation sites were mapped into each gene. Then, PCC values were calculated based on methylation level and gene expression in OS. Specific eQTMs were obtained and divided into positive and negative eQTMs. At last, a number of eQTM network were constructed. The numbers of eQTMs, positive eQTMs, negative eQTMs, genes and methylation sites were changed with the threshold of PCCs (Fig. 1B). The number of positive eQTMs was always higher than the number of negative eQTMs. The genes related with top number of methylation sites were PRDM16, SDK1, SHANK2, PRKCZ, PTPTN2, DIP2C and so on (Fig. 1C). The methylation sites related with top number of genes were cg15923943, cg25962358, cg01202150, cg11954680 and so on (Fig. 1D). PRDM16- and cg15923943-based network was constructed (Fig. 1E, F). The PR domain containing the 16 (PRDM16) gene, may act as a bidirectional switch between brown fat and skeletal muscle in mice. The copy number of PRDM16 molecules of patients with OS was higher than that of the normal [3]. All above results showed there were associations between eQTMs and OS.

Construction of the eQTMs-based risk score model for prognosis in OS

In order to identify prognosis-related eQTMs for OS, a comprehensive eQTMs-based risk score model was constructed (Fig. 2). There were four general steps in the workflow of the eQTM-based risk score model (Fig. 2A-D). In step 1, all the OS samples were randomly divided into two groups including modeling group and validation group to avoid over-fitting (Fig. 2A). In step 2, Multivariate Cox regression analysis was performed for the methylations related with a gene in modeling group (Fig. 2B). In step 3, a risk score formula was developed by integrating the methylation levels and corresponding coefficients of these methylation sites in validation group (Fig. 2C). A risk score was given for each patient based on this formula. In step 4, the OS patients were ranked by their risk scores and divided into two risk groups including high- and low-risk groups by the median risk score. Further survival analysis was performed to evaluate the prognostic significance between two risk groups (Fig. 2D).

Application of the eQTMs-based risk score model for prognosis in OS

We performed above eQTMs-based risk score model ten times and extracted the eQTMs which are significantly associated with survival at least seven times to ensure the accuracy. At last, we obtained 17 prognosis-related eQTMs for OS patients. There were 51.8% eQTMs with more than three methylation sites and 0.37% of them were significantly associated with survival in all the OS-specific eQTMs (Fig. 3A). The genes in this 17 prognosis-related eQTMs for OS patients were BMP4, CNNM2, DAPP1, DNAJC15, EN1, FADS2, FAM60A, ITSN1, JAM3, KIRREL3, MEIS1, NCOA4, NGEF, PSD4, PSTK, TTC12 and WWTR1-
AS1 (Fig. 3B). Gene BMP4 ($P = 4.14e-4$), DNAJC15 ($P = 4.46e-4$) and EN1 ($P = 0.011$) were shown as examples and strongly associated with survival in OS (Fig. 3C).

**The Global Characteristics For Prognosis-associated Eqtms In Os**

We further to analyze the characteristics of prognosis-associated eQTMS to explore the mechanism for eQTMs in OS. There were complex and multiple regulated patterns including positive and negative eQTMs in most prognosis-associated eQTMs (Fig. 4A). For example, all the eQTMs were negative for gene BMP4, CNNM2, DNAJC15, EN1, FADS2, FAM60A, NCOA4 and WWTR1-AS1. Some genes, such as DAPP1, ITSN1, JAM3, MEIS1, NGEF, PSTK and TTC12 had mixed regulated patterns including positive and negative eQTMs. All the eQTMs were positive in only gene KIRREL3 and PSD4. Totally, there were 47.06% negative, 11.76% positive and 41.18 mixed eQTMs in all prognosis-associated eQTMs of OS (Fig. 4B). There were six methylation sites related with gene BMP4 to form a eQTM network (Fig. 4C). Although all the methylation sites were negative correlated with gene BMP4, the levels of correlation were diverse. The cg24526899 had the most strong negative correlation with gene BMP4 in OS patients ($\text{Corr} = -0.53$, Fig. 4D). Gene TTC12 was correlated with most methylation sites to form eQTM network in OS (Fig. 4E). There were five positive and 12 negative eQTMs of gene TTC12 and they also showed diverse correlation levels in OS patients. The methylation level of cg05127217 was significantly associated with gene expression of TTC12 ($\text{Corr} = -0.58$, Fig. 4F). The methylation level of cg14632485 was significantly associated with gene expression of TTC12 ($\text{Corr} = 0.47$, Fig. 4G). All above results indicated that the regulation patterns of prognosis-associated eQTMS in OS were complex.

**The performance of eQTMs-based risk score model for prognosis in OS**

In order to ensure the accuracy and avoid over-fitting, we performed the eQTMs-based risk score model ten times. Only the eQTMs were significantly related with survival at least seven times would be extracted for follow analysis. Totally, 17 prognosis-associated eQTMs were obtained in OS patients. There were slight differences in each time for these 17 prognosis-associated eQTMs (Fig. 5A, 5B). Gene DNAJC15 and FADS2 were significant in nine of ten times for performing eQTMs-based risk score model (Fig. 5C). We also only used the median of gene expression to divide the samples to two groups and validated the associations between single gene and survival. We found the eQTMs could better distinguish the prognostic risk group than single gene in all the 17 prognosis-associated eQTMs of OS (Fig. 5D). In addition, we also explored the associations between age, gender and survival. We found the age ($P = 0.323$) and gender ($P = 0.442$) were both not related with survival (Fig. 5E, 5F). The results indicated that eQTMs may be the better prognostic biomarker candidates than single gene, age and gender.

**The functional and drug target analysis shows the roles of prognosis-associated eQTMS in OS**

We perform GO enrichment analysis based on all the genes in the prognosis-associated eQTMs of OS. We find these genes are enrichment in some important GO terms associated with cancer development.
(Fig. 6A). For example, some GO terms such as positive regulation of programmed cell death, positive regulation of apoptotic process, negative regulation of vasculature development and so on were discovered. These genes were also enrichment in some key pathways such as signaling pathways regulating pluripotency of stem cells (Fig. 6B). Previous study reported that human OS cell lines contained subpopulations of cells with stem-like attributes[7]. In addition, we also construct a drugs and genes network which is extracted from drug enrichment result (Fig. 6C). The drugs in the FADS2-based network contain Alpha-Linolenic Acid, Omega-6 fatty acids and Evening primrose oil which all could influence metabolic. The drug Streptozocin and Fostamatinib were also found. Fostamatinib has been investigated for the treatment and basic science of Rheumatoid Arthritis and Immune Thrombocytopenic Purpura (ITP).

Discussion

Predicting the survival is a major challenge of studying the mechanism of and treatment for OS. To identify potential prognosis-associated eQTMs for predicting survival in OS, an eQTM-based risk score model based on gene expression and methylation level was built to infer the prognostic efficacy of each eQTM. 17 prognosis-associated eQTMs were identified and analyzed.

As genomic technologies expand, multiple data types may serve as informative biomarkers, and bioinformatic strategies have evolved around these different applications. In previous studies, many factors such as age, gender, gene expression and methylation level were all could became prognostic biomarkers for cancer patients [3]. For example, Li et al. discovered and validated DNA methylation markers for overall survival prognosis in patients with thymic epithelial tumors [3]. DNA methylation was associated with survival in non-metastatic clear cell renal cell carcinoma [3]. Specially, many computational approach were built to predict survival risk. Lee et al. proposed a data-driven construction method for survival risk-gene networks as well as a survival risk prediction method using the network structure[8]. Dong et al. used machine learning to analyze DNA methylation data to build a model with three risk categories for predicting survival of hepatocellular carcinoma patients[9]. Prognostic risk model was constructed in glioblastoma multiform based on mRNA/microRNA/long non-coding RNA analysis using random survival forest method[10]. In our present study, a novel computational approach based on eQTMs by integrating gene expression and methylation level in OS.

We ensured that the accuracy of identifying the prognosis-associated eQTMs in OS from the following aspects: 1) To avoid over-fitting problem, all the OS patients were randomly divided into two groups including modeling and validation group. The modeling group was used to train the cox regression coefficients and the validation group was used to construct risk score model. 2) The eQTM-based risk score model was performed ten times and only the eQTMs which were significant in at least seven times could be considered as prognosis-associated eQTMs of OS. 3) The comparsion of prediction efficiency between single gene and eQTMs were performed. We used single gene expression to predict the survival risk and found all the eQTMs had better performance than single gene in OS survival risk prediction. 4) The age and gender also serve as informative biomarkers to predict survival and also obtained poor
performance than eQTMs in OS. All above results could indicate that eQTMs had the strong performance based on tests of accuracy, reliability, and robustness in OS survival risk prediction.

Conclusions

In summary, we constructed the eQTM-based risk score model to infer the prognostic efficacy by integrating gene expression and methylation levels. 17 prognosis-associated eQTMs of OS were extracted and validated. Further analysis indicated that these eQTMs showed better performance than existing clinical and molecular signatures such as single gene expression, age and gender of OS survival risk prediction. Our systematic analysis revealed that the eQTM-based risk score model can provide helpful information in the discovery of prognostic biomarkers in OS.

Abbreviations

OS
Osteosarcoma; eQTMs: expression Quantitative Trait Methylations; TARGET: Therapeutically Applicable Research To Generate Effective Treatments; PCCs: Pearson correlation coefficients; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: false discovery rate

Declarations

Funding

Not applicable.

Disclosure of interest

The authors declare that they have no competing interest.

Acknowledgments

Not applicable.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication
Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Consent for publication**

Not applicable.

**Authors’ contributions**

ZGF conceived and designed the experiments, FMA, SQP, ZAP, WGS and LSH analysed the data, and DPF, ZYL and WJS wrote the manuscript.

**References**

1. Ottaviani G, Jaffe N. The epidemiology of osteosarcoma. Cancer Treat Res. 2009;152:3–13.
2. Heymann D. Metastatic osteosarcoma challenged by regorafenib. Lancet Oncol. 2019;20(1):12–4.
3. !!!! INVALID CITATION !!!!
4. Fu LL, Li CJ, Xu Y, Li LY, Zhou X, Li DD, Chen SX, Wang FG, Zhang XY, Zheng LW. Role of IncRNAs as Novel Biomarkers and Therapeutic Targets in Ovarian Cancer. Crit Rev Eukaryot Gene Expr. 2017;27(2):183–95.
5. Guo W, Zhu L, Zhu R, Chen Q, Wang Q, Chen JQ. A four-DNA methylation biomarker is a superior predictor of survival of patients with cutaneous melanoma. *Elife* 2019, 8.
6. Tian W, Li Y, Zhang J, Li J, Gao J. Combined analysis of DNA methylation and gene expression profiles of osteosarcoma identified several prognosis signatures. Gene. 2018;650:7–14.
7. Wang L, Park P, Lin CY. Characterization of stem cell attributes in human osteosarcoma cell lines. Cancer Biol Ther. 2009;8(6):543–52.
8. Lee M, Han SW, Seok J. Prediction of survival risks with adjusted gene expression through risk-gene networks. *Bioinformatics* 2019.
9. Dong RZ, Yang X, Zhang XY, Gao PT, Ke AW, Sun HC, Zhou J, Fan J, Cai JB, Shi GM. Predicting overall survival of patients with hepatocellular carcinoma using a three-category method based on DNA methylation and machine learning. J Cell Mol Med. 2019;23(5):3369–74.
10. Wang X, Wang Y, Li L, Xue X, Xie H, Shi H, Hu Y. A lncRNA coordinates with Ezh2 to inhibit HIF-1alpha transcription and suppress cancer cell adaption to hypoxia. *Oncogene* 2019.

**Figures**
Figure 1

Identification and characterization of specific eQTMs in OS. (A) The workflow of identifying the specific eQTMs for OS. (B) The size of circles represented number of global, positive, negative eQTMs, gene and methylation sites based on diverse criteria. (C) The green radar chart showed the number of methylations for eQTMs related with most methylations. (D) The red radar chart showed the number of genes for
eQTMs related with most genes. (E) The PRDM16-based eQTM network. (F) The cg15923943-based eQTM network.

Figure 2

Workflow of the eQTM-based risk score model for evaluating the prognostic ability of eQTMs in OS. (A) Dividing the OS patients to modeling group and validation group at random. (B) Performing the COX regression analysis for the eQTM network in modeling group. (C) Calculating the risk score based on gene expression and methylation profiles. (D) Patients were ranked by their risk scores and divided into two risk groups. (E) The PRDM16-based eQTM network. (F) The cg15923943-based eQTM network.
expression and methylation profile in validation group. (D) Dividing the patients intro high- and low-risk groups based on risk scores.

Figure 3

Application of the eQTMs-based risk score model for prognosis in OS. (A) The first pie chart showed the percent of eQTMs with more than three methylation sites. The second pie chart showed the percent of prognosis-related eQTMs in all the eQTMs with more than three methylation sites. (B) The P-values of
prognosis-related eQTMs for OS. (C) The Kaplan-Meier curve for the overall survival of high- and low-risk groups. The difference between the two curves was evaluated by a two-sided log-rank test. The eQTM-based risk score distribution. The patient survival status of the prognosis-related eQTMs for OS.

Figure 4

The global characteristics for prognosis-associated eQTMS in OS. (A) The number of methylation sites for each prognosis-related eQTMs in OS. Red and blue represented positive and negative eQTMs. (B) The pie chart showed the percent of all positive, all negative and mix eQTMs for all prognosis-related eQTMs.
in OS. (C) The BMP4-based eQTM network. The thickness of edges represented correlated level. (D) The point plots showed the correlation between gene expression of BMP4 and methylation level of cg24526899 in OS. (E) The TTC12-based eQTM network. (F) The point plots showed the correlation between gene expression of TTC12 and methylation level of cg05127217 in OS. (G) The point plots showed the correlation between gene expression of TTC12 and methylation level of cg14632485 in OS.

Figure 5
The performance of eQTM-based risk score model for prognosis in OS. (A) Heat map showed the P-values of prognosis-related eQTMs at ten times. Red represented bigger P-values. (B) The bar plot represented the minimum P-values of these prognosis-related eQTMs in an eQTM-based risk score model. (C) The bar plot represented the minimum P-values of gene in ten times for eQTM-based risk score model. (D) The bar plot showed the P-values of survival for single gene and eQTM in OS patients. (E) The survival plot of age in OS patients. (F) The survival plot of gender in OS patients.

Figure 6
Functional analysis for prognosis-related eQTMs in OS. (A) GO terms and (B) KEGG pathways enriched for genes in prognosis-related eQTMs in OS, ranked by -\log_{10}(P\text{-value}). (C) Drug and gene network for prognosis-related eQTMs in OS.