1. Introduction

Osteosarcoma driven by bone-forming mesenchymal cells belongs to the most frequent primary tumor of bone in children and adolescents [1]. Children and adolescents comprised 70.62% of osteosarcoma cases between 1999 to 2017 [2]. It is counted that 4.7/1000000 people in children and adolescents suffer from osteosarcoma [3]. Osteosarcoma is characterized by acute local pain caused by the serious imbalance between formation and degradation in bone tissue [4]. Age, gender, height, socio-economic status, genetics and environmental condition are named as the significant risk factors of osteosarcoma [5]. Before chemotherapy, surgery was the only therapeutic strategy for patients with osteosarcoma with 20% of event-free survival [6]. Currently, patients suffering from osteosarcoma can be treated with surgery, chemoradiotherapy, and neoadjuvant chemotherapy based on the complex schedule of doxorubicin, cisplatin, ifosfamide, as well as high-dose methotrexate with leucovorin rescue [7]. Despite marked advancement in the numerous drugs and therapeutic strategies of osteosarcoma, the survival outcome of patients with osteosarcoma has not greatly improved for nearly four decades [6, 8]. Unfortunately, the particular molecular mechanism of osteosarcoma remains largely unknown. Understanding tumor pathogenesis at the molecular level provides signature biomarkers of the diagnosis and prognosis and potential targets for osteosarcoma therapeutic strategies.

Exogenous and endogenous genotoxic events, such as radiation, chemicals, reactive oxygen species (ROS) and topological alterations, evoke DNA damage that alters the transcription and translation of genetic information [9]. During tumorigenesis, cancer cells are featured by frequent proliferation that results in genomic instability and mutation accumulation [11]. To decrease...
the potential proliferation-induced risk of genomic alterations, the DDR mechanism is activated in these cells with malignant behaviors. On the one hand, DDR mediates the negative modulation of tumor microenvironment to genomic instability, which contributes to proliferation, apoptosis and malignant transformation in cancer cells [12]. On the other hand, innate immunity is activated by DDR deficiencies due to the overlaps between the DNA repair pathway and cytosolic DNA sensing pathway, and also, DDR can regulate lymphocyte development in adaptive immunity via DDR pathways containing mismatch repair (MMR), base excision repair (BER), alternative end-joining (A-EJ), non-homologous end joining (NHEJ) [13]. Thus, DDR function as the key role of tumorigenesis.

Changes in DDR are associated with an increased risk of osteosarcoma. DDR inhibitors have been predicted to be the potential therapeutic method for improving the survival outcome of osteosarcoma [14]. Nevertheless, the potential of DDR-related genes (DDRGs) in osteosarcoma is unclear. In the present research, we will screen and identify the signature DDRGs in osteosarcoma using bioinformatics technology on the basis of the microarray datasets from GEO, cBioPortal as well as TCGA. Subsequently, we plan to process risk analysis, function enrichment analysis along prediction analysis of chemical drugs to evaluate the potential of DDRGs. These DDRGs obtained from our investigation will greatly reveal the pathophysiological mechanism of osteosarcoma at the molecular level in addition to providing the novel DDR-related biomarkers contributing to osteosarcoma prognosis.

2. Materials and methods

2.1. Data source

Gene expression profiles together with clinical information of osteosarcoma patients were downloaded from TCGA-TARGET (https://www.ogc.cancer.gov/) as well as GEO (https://www.ncbi.nlm.nih.gov/geo/) database. Survival information of 85 osteosarcoma patients in TCGA-TARGET was adopted to obtain prognostic DNA damage response (DDR) signature as well as construct a risk score model. GSE16088, including 14 osteosarcoma and 6 control samples, was utilized for identifying differentially expressed DDR genes (DEDDRGs) to further screen prognostic DDR signature. GSE39055, including 37 osteosarcoma patients with survival data, was used to be an external validation set to test the reliability of the constructed risk score model. 543 DDR genes were downloaded from cBioPortal database (http://www.cbioportal.org/).

2.2. Consensus clustering and principal components analysis

On the basis of the expression profiles of DDR genes, osteosarcoma patients in TCGA cohort were categorized into subtypes using “ConsensusClusterPlus” R package with the settings maxK = 10, clusterAlg = “hc” and distance = “pearson”. Cumulative distribution (CDF) together with consensus matrices was adopted to calculate the number of subtypes. Thereafter, the Principal Components Analysis (PCA) was plotted to evaluate the effect of consensus clustering. Moreover, the distribution of clinical characteristics among different subtypes was analyzed by chi-square test. Kaplan-Meier curves were performed to analyze the survival of patients in different subtypes.

2.3. Identification of key DEDDRs involved in osteosarcoma

DEDDRGs between subtype with the best survival and subtype with the worst survival were identified by “limma” R package with the threshold of p-value <0.05 as well as |log(FC)| >0.5. Likewise, DEDDRGs between 14 tumor and 6 control samples in GSE16088 dataset were identified with the same criteria. Then key DEDDRs involved in osteosarcoma were obtained by overlapping those DEDDRG.

2.4. Establishment and validation of the risk score model

We first performed univariate Cox regression to screen DEMAGs significantly linked to survival (p-value <0.1). Moreover, LASSO algorithm was performed to acquire the robust prognostic signature. Then multivariate Cox regression was implemented to calculate the coefficients of each prognostic signature to construct the risk score model. Based on the median value of the risk score, osteosarcoma patients were separated into high-risk and low-risk groups, respectively. The overall survival of both groups was analyzed with the help of Kaplan-Meier analysis. For assessing the performance of the risk score model, the receiver operating characteristics (ROC) curve could plotted with the help of the "survivalROC" in the "R" package. Besides, the risk score model was tested in an external GSE39055 dataset. Additionally, univariate and multivariate Cox regression analyses were implemented to certify independent prognostic factors for osteosarcoma patients. The nomogram was established to predict 1-, 3- as well as 5-year survival of osteosarcoma patients. The performance of the nomogram was assessed with of the calibration curves.

2.5. GSVA analysis

To investigate the molecular mechanisms of the prognostic gene signature, the enrichment score of KEGG pathways in osteosarcoma patients was calculated by “GSVA” R package. The reference gene set “c2.cp.kegg.v7.4.symbols.gmt” was downloaded from MSigDB database (www.gsea-msigdb.org/gsea/msigdb/). Then the relationship between risk score and KEGG pathways was analyzed by virtue of Pearson correlation.

2.6. Characterization of patients in low- and high-risk groups

For characterizing the patients in low- together with high-risk groups, we compared: (i) their sensitivity to anti-cancer drugs using pRRophetic algorithm. (ii) their expressions of immune checkpoint molecules, including PDCD1, CD274, CTLA4, ICOS, HAVCR2, CD47, SIRPA and TNFRSF9. (iii) their immune activity and immune cell infiltration by ssGSEA method.

3. Results

3.1. Identification of three osteosarcoma subtypes based on DDRs

Based on the expressions of DDRGs, consensus clustering was performed and K = 3 was identified with the optimal clustering stability (Figure 1A-1C). PCA analysis further demonstrated that osteosarcoma patients were divided into 3 distinct subtypes (Figure 1D). The distribution of clinical characteristics (age, gender and race) in each
subtypes were displayed in the heatmap (Figure 1E). Besides, it was observed that patients in cluster 1 possessed the best survival (Figure 1F, 1G), implying that DDRGs have a vital role in the prognosis of osteosarcoma patients.

3.2. Identification of 43 key DEDDRGs in osteosarcoma patients

A total of 80 DEDDRGs was identified between cluster 1 and cluster 3, including 40 up-regulated as well as 40 down-regulated genes in cluster 1 in comparison with cluster 3 (Figure 2A-2B). Meanwhile, we found total 296 DEDDRG between osteosarcoma and control samples, including 174 up-regulated as well as 122 down-regulated genes in osteosarcoma samples in comparison with control ones (Figure 2C-3D). After overlapping those DEDDRGs, 43 key DEDDRGs involved in osteosarcoma were identified in the current study (Figure 2E).

Thereafter, we measured the prognostic value of 43 key DEDDRGs. Through univariate Cox regression analysis, CALD1, CDK6, CSF1R, EGFR, ERBB4, FGFR3, GATA2, ID2, SLC7A8, SOCS1 and SUGCT were found to be closely linked to prognosis of osteosarcoma (Figure 2F). Subsequently, those 11 genes were input into LASSO algorithm, and CDK6, CSF1R, EGFR, ERBB4, GATA3 and SOCS1 were further identified as the robust prognostic signatures in osteosarcoma (Figure 2G). Next, via multivariate Cox regression, the coefficient of each prognostic signature received calculation that used for the following construction of risk score model.

3.3. Construction and validation of the risk score model by prognostic DDR signature

Based on the coefficients of NUDT1 and PDGFB in Table S1, the risk scores of each patient were calculated. Based on the median of the risk scores, the patients in the TCGA training set were separated into high- and low-risk groups (Figure 3A). The expressions of prognostic DDR signature in both groups are displayed in Figure 3B. Furthermore, an apparent survival difference (p <0.05) could be also discovered between the 2 groups (Figure 3C). Besides, ROC curves displayed that the risk score model possessed high accuracy in forecasting the survival of osteosarcoma patients with areas under the curves (AUC) > 0.7 (Figure 3D). The consensus outcomes could be also acquired in the GSE39055 dataset (Figure 3E-3H).

3.4. Establishment of the DDR-related nomogram in osteosarcoma and a close relationship between the risk score and multiple signaling

Next, the relation between risk score and clinical features compartmentalized by age, gender, race, cluster and metastasis status was explored. It was discovered that the risk score was different between different clusters and metastasis status (Figure 4A, 4B), while no difference of risk score was detected between different ages, gender or race (Figure 4C-4E). Moreover, using univariate analysis, we discovered that risk score along with metastasis status were significantly related to prognosis (Figure 4F). Subsequent multivariate analysis unveiled that the risk score and metastasis status remained markedly linked to prognosis (Figure 4G), implying that they could be independent prognostic factors in osteosarcoma. Moreover, we adopted risk score and metastasis status to establish a nomogram for forecasting the 1-, 3-, as well as 5-year survival of osteosarcoma patients.

Fig. 1. Consensus clustering and survival analysis for osteosarcoma subclassification. A, Delta area curve of cumulative distribution function (CDF); B, CDF curve at K=2-10; C, Consensus matrix at K=7; D, Principal component analysis for 3 clusters; E, Distribution of clinical features in 3 clusters; F, Kaplan-Meier curves of cluster 1 together with cluster 2; G, Kaplan-Meier curves of cluster 1 together with cluster 3.

Fig. 2. Identification of DDRGs based on differential expression analysis. A, Volcano plots of DDRGs from TCGA-target database (cluster 1 vs cluster 3). Red dots, upregulated DDRGs. Blue dots, downregulated DDRGs. Grey dots, non-differential genes; B, Heatmap of DDRGs from TCGA-target database; C, Volcano plots of DDRGs from GEO (Tumor vs normal). Red dots, upregulated DDRGs. Blue dots, downregulated DDRGs. Grey dots, non-differential genes; D, Heatmap of DDGRs from GEO; E, Venn diagram of overlapping DDGRs between TCGA-target and GEO; F, The p-value and hazard ratio for univariate COX analysis of DDRGs. Red blocks, hazard ratio>1. Grey blocks, p-value<0.05. Blue line, confidence interval; G, Least absolute shrinkage as well as selection operator (LASSO) of DDRGs.

Fig. 3. Construction and validation of survival risk model based on DDRGs in training set. A, Risk score and survival time of DDRGs; B, Heatmap of DDRGs in 2 groups; C, Kaplan-Meier curves of 2 groups; D, Receiver operating characteristic curve (ROC) showing prognostic value of risk model in 1 year - , 3- as well as 5-year survival; E, Risk score and survival time of DDRGs; F, Heatmap of DDRGs in 2 groups; G, Kaplan-Meier curves of 2 groups; H, Receiver operating characteristic curve (ROC) showing prognostic value of risk model in 1 year , 3- as well as 5-year survival.
DNA damage in osteosarcoma.

Fig. 4. Correlation analysis between risk score and osteosarcoma clinical features, and the correlation between risk score and KEGG pathways analyzed by GSEA. Correlation between risk score and subclassification group (A), metastasis status (B), age (C), gender (D) and race (E). NS, non-statistically difference; E, The p-value and hazard ratio for univariate COX analysis of risk score and clinical features. Red blocks, hazard ratio<1. Grey blocks, p-value<0.05. Blue line, confidence interval; G, The p-value and hazard ratio for multivariate COX analysis of risk score and metastasis status. Red blocks, hazard ratio>1. Grey blocks, p-value>0.05. Blue line, confidence interval; H, Nomogram constructed based on risk score and metastasis status for the prediction of overall survival (OS) at 1-, 3- as well as 5-year for osteosarcoma; I, Calibration curves of nomogram at 1-, 3- and 5-year; J. Clustering of correlation coefficients between risk score and KEGG pathways; K, The correlation heatmap between the KEGG pathway and the risk score. The horizontal axis represents the sample, and the risk score increases in turn from left to right. ***, p<0.001.

Fig. 5. Immune microenvironment alteration between low- and high-risk groups. A, Immune cell infiltration in 2 groups; B, The expressions of immune checkpoints in 2 groups; C, Response to chemical drugs in 2 groups. *, p<0.05. **, p<0.01. ****, p<0.0000.

toesarcoma patients (Figure 4H). The calibration curves for the 1-, 3-, as well as 5-year demonstrated that the predicted overall survival of 3- and 5-year were very close to the actual observed overall survival (Figure 4I), implying the clinical use of the nomogram.

To elucidate the molecular mechanisms of these prognostic signatures in osteosarcoma, we performed GSVA and calculated the Pearson correlation between risk score and 186 pathways. We found that the risk score had moderate correlations with signaling pathways relevant to immune, cell metabolism, proliferation, adhesion, apoptosis, and migration (Figure 4J, 4K), including primary bile acid biosynthesis, pyrimidine metabolism, glycine serine and threonine metabolism, phenylalanine metabolism, DNA replication, protein export, MAPK signaling pathway, cytokine receptor interaction, chemokine signaling pathway, cell cycle, endocytosis, apoptosis, focal adhesion, cell adhesion molecules CAMs, complement and coagulation cascades, Toll-like receptor signaling pathway, Nod-like receptor signaling pathway, JAT/STAT signaling pathway, natural killer cell-mediated cytotoxicity, T cell receptor signaling pathway, B cell receptor signaling pathway, Fc gamma receptor-mediated phagocytosis and leukocyte transendothelial migration.

3.5. The immune microenvironment and drug sensitivity were different between low- and high-risk group

In consideration of the above results, we examined the immune microenvironment of patients in low- and high-risk groups. The results of ssGSEA revealed that the enrichment scores of CCR, CD8+ T cells, checkpoint, macrophages, neutrophils, T cell co-inhibition, T cell co-stimulation, Th2 cells, TIL, Treg as well as type II IFN presented elevation in low-risk group (Figure 5A). Additionally, the expressions of checkpoint molecules CD274, CD47, HAVCR2, SIRPA as well as TNFRSF9 were significantly different between the two groups (Figure 5B). These results indicated the different immune microenvironments between 2 groups. It has been reported that the immune microenvironment affects the therapeutics of cancer patients [15]. Herein, patients in low-risk group could be more sensitive to Midostaurin, LFM.A13, Bryostatin.1, Embelin, Pazopanib, CHIR.99021, Z.LLNle.CHO, WH.4.023, CMK, AS601245, JNK.Inhibitor.VIII, FH535, GSK269962A, MG.132, Cl.1040, Bexarotene, FTI.277, PLX4720, Imatinib, WO2009093792 and CCT007093 (Figure 5C).

4. Discussion

DDR-involved genomic instability induces the mutation accumulation that elevates the malignant clonal evolution of cancer cells [16]. Recently, DDR has gained increasing attention in the tumorigenesis of several cancers, particularly in osteosarcoma. However, the profile of DDR-related genes (DDRGs) in osteosarcoma remains elusive. In the current study, we screened and identified the signature DDRGs to describe the role of DDR in osteosarcoma progression.

In the present investigation, we divided patients with osteosarcoma into three clusters to demonstrate the correlation between DDRGs and osteosarcoma progression via screening DDRGs. The significant difference in osteosarcoma survival among the three clusters suggests DDR contributes to the development of osteosarcoma. When cancer cells are under the rapid replication phase, DDR pathway-activated cancer-related genes such as p53 and sirtuin-1 (SIRT1) regulate, at the epigenetic level, expressions of genes that monitor cell cycle, senescence, proliferation, apoptosis and DNA repair, thereby driving osteosarcoma progression [17, 18]. Therefore, profiling the key genes of DDR network may contribute to understanding osteosarcoma pathogenesis. Importantly, we identified six DDRGs (CDK6, CSF1R, EGFR, ERBB4, GATA3 and SOCS1) linked with osteosarcoma prognosis. These six DDRGs are closely associated with malignant behaviors in cancer cells. CDK6 is one of the cyclin-dependent ki-
nases that mainly supervises the replication and repair of DNA molecules, indicating CDK6 plays a significant role in DDR [19]. DDR-related activation of CDK6 contributes to malignant tumor events and is correlated with cancer subtype and survival outcome. In tumor immunity, CSF1R is greatly phosphorylated by DNA damage to repress the recruitment of tumor-infiltrating myeloid cells [20, 21]. EGFR is found to elevate DDR through inducing ROS production in the tumor process [22]. ERBB4, a member of the EGFR subfamily, enhances MDM2 phosphorylation to maintain p53 stabilized when DNA is damaged [23]. Additionally, SOCS1 directly phosphorylates p53 to activate DDR in cancer cells while DDR triggers the activation of AMPK/PGC1α pathway causing increased GATA3 expression, which enhances mitochondrial biogenesis that maintains the viability and biofunction of CD4 T cells [24]. In response to DNA damage, the six signature genes target downstream signaling that controls oxidative stress, immune cells and the activation and stability of p53, which contributes to the development of osteosarcoma.

To verify the prognosis value of the identified six DDRGs, we established the risk model according to COX analysis. Then, the risk score of osteosarcoma samples was calculated on the basis of six DDRG expressions. We noticed that the risk score was associated with the survival rate of osteosarcoma patients. As the independent prognostic factor, the risk score contributes to the forecast osteosarcoma prognosis. Furthermore, we discovered that this risk model could forecast 3-years-survival as well as 5-years-survival of osteosarcoma. Collectively, the six DDRGs are valuable biomarkers of osteosarcoma prognosis, showing the feasibility of predicting the survival outcome of osteosarcoma based on expressions of the six DDRGs.

Interestingly, we found the risk score was correlated to six function strategies including immunity, cellular metabolism, proliferation, apoptosis, adhesion and migration, suggesting that the six prognosis-related DDRGs might encode these cellular processes in response to DDR-involved osteosarcoma progression. Our findings were supported by previous works, which verified the six DDRGs induce changes in tumor immunity, p53-mediated tumorigenesis [25] and ROS-involved metabolism [26], which results in the disorder of replication and metastasis in osteosarcoma cells.

With the risk score as the classification standard, osteosarcoma patients were separated into a high-risk group and a low-risk group to assess the relationship between six biomarkers and the immune microenvironment. Notably, significant differences were discovered in immunocyte infiltration, immunocompetence and expressions of immune-associated genes between 2 groups. Differentially distributed immune profile affects the occurrence and development of malignant tumors. Immunocyte plays a key role in osteosarcoma. During tumorigenesis, there are several changes in the polarization of macrophages, the recruitment of myeloid-derived suppressor cells (MDSCs), and the infiltration of T lymphocytes, which finally cause immunosuppression and the release of pro-growth factors around cancer cells [27]. Differences in immune microenvironment further suggest six signature DDRGs control immunity-involved tumor progression in osteosarcoma, and provide the possibility of immune checkpoint inhibitors combined with targeted therapy of DDRGs. More significantly, we discovered that low-risk group easily responded to 22 chemotherapeutic drugs, indicating that chemotherapeutic strategies can be adopted referring to the risk score based on expressions of the six DDRGs.

Several limitations exist in this investigation. Osteosarcoma samples in databases are slightly small, resulting in the potential bias for our findings. Although we used samples in GEO as the external validation, the prognostic role of the six biomarkers should be determined by experimental evidences. Moreover, the exact mechanism of these biomarkers requires descriptions in follow-up investigations in vivo and in vitro.

In conclusion, we first demonstrated 6 DDRGs is the prognostic biomarkers in osteosarcoma, including CDK6, CSF1R, EGFR, ERBB4, GATA3 and SOCS1. The risk model based on expressions of six biomarkers will assist the prognosis assessment of osteosarcoma patients. Also, the six gene can be the potential molecular targets of anti-tumor therapeutic strategies for osteosarcoma.

Informed consent
The authors report no conflict of interest.

Availability of data and material
We declared that we embedded all data in the manuscript.

Authors’ contributions
LT conducted the experiments and wrote the paper; TZ and LS analyzed and organized the data; LM conceived, designed the study and revised the manuscript.

Funding
None.

Acknowledgements
We thanked Suzhou Dushu Lake Hospital (Dushu Lake Hospital Affiliated to Soochow University) for approving our study.

References

DNA damage in osteosarcoma.


114
DNA damage in osteosarcoma.


