Translational Oncology 25 (2022) 101499

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Original Research

Construction of the prognostic enhancer RNA regulatory network in osteosarcoma

Penghui Yan [a,1], Zhenyu Li [b,1], Shuyuan Xian [b,1], Siqiao Wang [b,c,1], Qing Fu [b], Jiwen Zhu [a], Xi Yue [a], Xinkun Zhang [a,d], Shaofeng Chen [e], Wei Zhang [f], Jianyu Lu [f], Huabin Yin [g,*], Runzhi Huang [a,b,f,**], Zongqiang Huang [a,***]

[a] Department of Orthopedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China
[b] Tongji University School of Medicine, Shanghai 200092, China
[c] Division of Spine, Department of Orthopedics, Tongji Hospital Affiliated to Tongji University School of Medicine, Shanghai 200065, China
[d] Department of Orthopedics, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China
[e] Department of Orthopedics, The First Affiliated Hospital of Naval Medical University, Shanghai 200433, China
[f] Department of Burn Surgery, The First Affiliated Hospital of Naval Medical University, Shanghai 200433, China
[g] Department of Orthopedics, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai 200065, China

ARTICLE INFO

Keywords:
Osteosarcoma
eRNA
CD8A
CEBPA
CD3E
Network

ABSTRACT

Background: Osteosarcoma (OS) is a common malignant tumor in osteoarticular system, the 5-year overall survival of which is poor. Enhancer RNAs (eRNAs) have been implicated in the tumorigenesis of various cancer types, whereas their roles in OS tumorigenesis remains largely unclear.
Methods: Differentially expressed eRNAs (DEEs), transcription factors (DETFs), target genes (DETGs) were identified using limma (Linear Models for Microarray Analysis) package. Prognosis-related DEEs were accessed by univariate Cox regression analysis. A multivariate model was constructed to evaluate the prognosis of OS samples. Prognosis-related DEEs, DETFs, DETGs, immune cells, and hallmark gene sets were co-analyzed to construct an regulatory network. Specific inhibitors were also filtered by connectivity Map analysis. External validation and scRNA-seq analysis were performed to verify our key findings.
Results: 3,981 DETGs, 468 DEEs, 51 DETFs, and 27 differentially expressed hallmark gene sets were identified. A total of Multivariate risk predicting model based on 18 prognosis-related DEEs showed a high accuracy (area under curve (AUC) = 0.896). GW-8510 was the candidate inhibitor targeting prognosis-related DEEs (mean = 0.670, $p < 0.001$). Based on the OS tumorigenesis-related regulation network, we identified that CCAAT enhancer binding protein alpha (CEBPA, DETF) may regulate CD8A molecule (CD8A, DEE), thereby promoting the transcription of CD3E molecule (CD3E, DETG), which may affect allograft rejection based on CD8+ T cells.
Conclusion: We constructed an eRNA-based prognostic model for predicting the OS patients' prognosis and explored the potential regulation network for OS tumorigenesis by an integrated bioinformatics analysis, providing promising therapeutic targets for OS patients.

Abbreviation: OS, osteosarcoma; TF, transcription factor; KEGG, kyoto encyclopedia of genes and genomes; ssGSEA, single sample gene-set enrichment analysis; GSVA, gene set variation analysis; CCLE, cancer cell line encyclopedia; GEPIA, gene expression profiling interactive analysis; GTEX, genotype-tissue expression; ROC, the receiver operator characteristic; AUC, area under the curve; FDR, false discovery rate; GO, gene oncology; DEE, differential expressed eRNA; CIBERSORT, cell type identification by estimating relative subsets of RNA transcripts; BP, biological process; CC, cell component; MF, molecular function; eRNA, enhancer RNA; ATAC-seq, assay for targeting accessible-chromatin with high throughout sequencing; Cdk2, cyclin kinase 2.

* Corresponding author at: Department of Orthopedics, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, 100 Haining Road, Shanghai, China, 200065, China.
** Corresponding author at: Department of Burn Surgery, The First Affiliated Hospital of Naval Medical University, Shanghai, 200433, China.
*** Corresponding author at: Department of Orthopaedics, The First Affiliated Hospital of Zhengzhou University, 1 East Jianshe Road,Zhengzhou, 450052, China.
E-mail addresses: yinhuabin@aliyun.com (H. Yin), 1810938@tongji.edu.cn (R. Huang), gzhuangzq@163.com (Z. Huang).
[1] Co-first authorship: Penghui Yan, Zhenyu Li, Shuyuan Xian and Siqiao Wang have contributed equally to this work.

https://doi.org/10.1016/j.tranon.2022.101499
Received 20 April 2022; Received in revised form 8 July 2022; Accepted 26 July 2022

Instruction

Osteosarcoma (OS) is a common malignant tumor in osteoarticular system, and it originates from primitive mesenchymal cells [1]. There will be 3610 new cases and 2060 new deaths of bone and joint cancer in 2021 estimated by the American Cancer Society [2]. However, the 5-year overall survival of OS patients in 0 to 14 and 15 to 19 is only 68% and 67%, respectively [2]. What' worse, the 5-year overall survival of OS patients with complication like pathologic fractures is only 46% in adults [3]. Moreover, the 5-year overall survival of metastasis and relapse OS is only 25% [4]. Many researchers devote to exploring the latent mechanism and therapy targets like IGF-1 receptor antagonist, HER2 receptor blocker and PDFG inhibitor [5]. However, the etiology is ambiguous up to date, and the poor quality of OS patients' survival attributes to metastases and pathologic fractures. Therefore, it is important to investigate the possible tumorigenic mechanism of OS and subsequently to identify the prognostic biomarkers and treatment targets underlying the interactions between potential factors. In this study, prognostic biomarkers that associate with tumorigenesis of OS were identified.

With the development of the second- and third-generation sequencing assays, non-coding RNAs (ncRNAs) have gained a lot of attention due to their ability to regulate gene expression. Enhancer RNAs (eRNAs) are ncRNAs transcribed from enhancers, which indicates the activation of according enhancers [6]. Gene expression can be regulated by eRNAs via directing chromatin accessibility, chromatin stabilization, chromatin remodeling, epigenetic modification, and transcription factors (TFs) [7]. Recent studies have validated the potential role of eRNAs interacting with oncogenes and pathways activating in various human diseases [8,9]. And eRNAs activate the transcription of target genes, which potentially promotes the tumor genesis [10]. We proposed that eRNAs may play significant roles in the tumorigenesis of OS, indicating the clinical use of eRNA-associated treatments.

In this study, differentially expressed eRNAs (DEEs) were identified, the prognostic value of which was accessed by univariate Cox regression analysis. Besides, a multivariate model was constructed to evaluate the prognosis of OS samples. For exploring the potential signaling axis underlying the mechanism of OS tumorigenesis, differentially expressed target genes (DETGs), differentially expressed TFs (DETFs), immune cells, and hallmark gene sets significantly co-expressed with DEEs were determined using correlation analysis. In addition, an eRNA centric regulatory network was established to further decode the potential interplay between these factors. Besides, specific inhibitors targeting DEEs/DETFs were also filtered by connectivity Map (CMap) analysis. Finally, a series of multidimensional validation based on serval multi-omics databases were utilized to further validate the reliability of our findings. Our findings may promote the understanding of the mechanism underlying the tumorigenesis of OS, which allow a more selective application of therapeutic agents targeting specific eRNAs and their co-factors.

Methods

Data acquisition

RNA-sequencing (RNA-seq) data of 189 OS samples were downloaded from Target database (https://ocg.cancer.gov/programs/target) and Treehouse database (https://treehousegenomics.soe.ucsc.edu/public-data/#datasets) (74 from Target and 115 from Three house database). Raw count of RNA-seq data obtained from Target database and Treehouse database were quantified, respectively, which were then standardized using voom function in limma (Linear Models for Microarray Analysis) package. To eliminate the batch effect, these two batches of RNA-seq data were corrected using normalizeBetweenArrays function in limma package, which were then merged for differential expression analysis [11]. Moreover, RNA-seq profiling of 3 normal bone tissue was from Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra) (SRX3436393, SRX3436394, SRX3436395). Besides, demographic information like gender and age, clinical information like metastasis and primary site and endpoint data were also available from Target database. Therefore, the following survival analysis was performed based on the data from the target database.

Based on eRNA in cancer (eRic) database (https://hanlab.uth.edu/eRic/) [10], the normalized profiles of eRNA expression in OS were obtained. Chromatin-immunoprecipitation followed by sequencing (ChIP-seq) for Histone 3 Lysine 27 acetylation (H3K27ac) was used to annotate eRNAs in the form of the official gene symbol in line with their location in hg38 genome [12]. A total of 318 TFs, 50 Hallmark gene sets, and 80 target genes were obtained from Cistrome database (http://cistrome.org) [13], Molecular Signatures Database (MSigDB) v7.1 (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) [14], and eRic database (https://hanlab.uth.edu/eRic/) [10], respectively.

Differential expression analysis and functional enrichment analysis

DETGs and DEEs were identified between 3 normal bone tissue and 189 OS samples by Linear Models for Microarray Data (limma) package [11] and edge R algorithm [15]. When the absolute value of log2 Fold Change (FC) more than 1.0 and False Discovery Rate (FDR) value less than 0.05, genes and eRNAs can be defined as DEGs and DEEs, respectively. What's more, Gene Oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed to explore the biological processes and signaling pathways which enriched DEGs [16].

Construction of eRNA prognostic model

Univariate Cox regression analysis, based on the expression levels of the DEEs, was conducted on the 189 OS patients with the threshold of $p < 0.01$ to determine the DEEs that were related to prognosis of OS. least absolute shrinkage and selection operator (LASSO) Cox regression analysis was further performed on the screened DEEs to optimize the eRNAs and construct a prognostic model by utilizing the glmnet package in R [17]. Then risk score of each OS patient was calculated by utilizing the following formula based on the screened prognosis-related DEEs [18]:

$$Risk\ score = \beta_1 \times eRNA_1 + \beta_2 \times eRNA_2 + \ldots + \beta_n \times eRNA_n$$

In the equation, "β", "eRNA" and "n" represented the coefficient of each eRNA calculated using LASSO Cox regression, the expression level of eRNA in each sample, and the sequence number of corresponding eRNA, respectively. OS samples were divided into low-risk and high-risk groups according to the median risk score. Further, the efficiency and accuracy of the prognostic model were assessed using receiver operating characteristic (ROC) curve analyses. Then, to evaluate accuracy of the risk score, the Kaplan-Meier survival analysis was utilized. Gender, age, metastasis, and primary site were included in our work as the clinicopathological factors affecting the prognosis of OS patients as well as many other cancers. The independent prognosis predicting value of risk score was accessed by univariate and multivariate Cox regression analysis with the correction of the demographics and clinical information discussed above. The four factors and Risk Score were integrated in the multivariate Cox regression analysis as variables.

Clinical correlation analysis and cell type identification by estimating relative subsets of RNA transcripts (CIBERSORT) analysis

Nonparametric test was applied to analyze the correlation between prognosis-related DEEs and clinical information that included first event and metastasis. Then, the profiling of gene expression was input into

CIBERSORT algorithm, and the infiltration proportion of eight different immune cell types was identified. The sum of estimated infiltration proportion of all immune cells in each OS patient was 1. Besides, the correlation analysis was utilized to evaluate the correlation between these immune cell types.

Identification of differentially expressed TFs, hallmark gene sets, and immune cells/functions

Limma package and edge R algorithm were utilized to identify differential expressed TFs and hallmark gene sets between normal bone and OS samples. Besides, Gene Set Variation Analysis (GSVA) [19] was utilized to quantify the expression of 50 hallmark gene sets in all samples. Specifically, TFs and hallmark gene sets with FDR < 0.05 and the absolute value of log2 FC > 1.0. Further, immune infiltration patterns of 29 types of immune cells/functions in OS and normal bone samples were investigated by utilizing single-sample gene set enrichment analysis (ssGSEA) based on specific gene markers [20,21].

Regulatory network of prognosis-related DEEs for OS tumorigenesis

Prognosis-related DEEs, DETFs, and DETGs were retrieved from the above screening; then differentially expressed hallmark gene sets were quantified as continuous variables using GSVA, and immune cells and immune functions were investigated by utilizing CIBERSORT and ssGSEA, respectively. Then, correlation analysis was conducted among the factors discussed above, which were shown in different colors. Specifically, purple indicated the immune cell types by CIBERSORT, blue indicated the hallmark gene sets by GSVA, indigo blue indicated the immune cells/functions by ssGSEA, yellow indicated potential upstream DETFs of prognosis-related DEEs, and pink indicated potential DETGs of prognosis-related DEEs. The interaction pairs between those prognosis-related DEEs and DETFs, DETGs, immune cell types, hallmark gene sets, and immune functions were used to construct the regulatory network for OS oncogenesis. In this network, thresholds were set as R (correlation coefficient) > 0.85 and $p < 0.05$ between prognosis-related DEEs and DETGs; $R > 0.70$ and $p < 0.05$ between prognosis-related DEEs and DETFs; $R > 0.50$ and $p < 0.05$ between prognosis-related DEEs and infiltrating immune cells; $R > 0.50$ and $p < 0.05$ between prognosis-related DEEs and immune gene sets; $R > 0.60$ and $p < 0.05$ between prognosis-related DEEs and hallmark gene sets. Furthermore, the Pearson co-expression analysis was also conducted to estimate the correlations between the six components in the regulatory network.

Clinical characteristic and prognostic analysis of prognosis-related DEEs

Enhancers control cellular identities and play as biomarkers for various cancers [22]. Therefore, we reasoned that eRNA profiles, which was easily detectable, could be clinically used as biomarkers. Machine learning for estimating patients' prognosis was currently believed to be more robust compared with traditional methods in multiple settings and disease conditions [23–25]. Here, we obtained the ncRNA profiling by array and clinical phenotypes of 91 OS patients (accession number: GSE39058) to validate the prognostic value of our candidate eRNAs [26].

Identification of potential small-molecule inhibitors

In the Connectivity Map (CMap) database (https://portals.broadinstitute.org/cmap/) [27], DEG maps were used to predict the correlations between small-molecule drugs and multiple diseases. The positive score was the same as the reference gene expression profile, while the negative score might be the opposite. Herein, CMap was utilized to identify small-molecule drugs which may target DEGs in OS based on the expression profiles of DEGs. Specifically, the database was used to screen enrichment fractions $<$ -0.85 and $p < 0.05$, and small-molecule drugs with negative scores were considered as potential therapeutic molecules for OS therapy.

Assay for targeting accessible-chromatin with high throughout sequencing (ATAC-seq) and ChIP-seq validation

The profiling of eRNAs was obtained to analyze the state of chromatin [28]. And ATAC-seq data for prognosis-related DEEs (accession number: GSE139190 and GSE139099) were used to determine the chromatin accessibility of prognosis-related DEEs in OS tissue as well as in the pan-cancer level. Furthermore, to determine whether prognosis-related DEEs and DETF binding overlap at active enhancers, comparative analyses of our ChIP-seq data (accession number: GSE134744) and the enhancer-related histone marks H3K27ac were conducted.

External validation based on multidimensional data

Considering the data limitation in our work, verification based on multiple databases was performed to further demonstrate the reliability of our findings, which was essential for optimizing the universality and authenticity of our results. Therefore, CellMarker [10] databases and PathCard (https://pathcards.genecards.org/) were utilized to explore the top three gene markers of specific immune cells and top five genes of hallmark pathways.

In addition, Gene Expression Profiling Interactive Analysis (GEPIA) [29], Oncomine [30], UALCAN [31], Linkedomics [32], SurvExpress [33], cBioportal [34], Genotype-Tissue Expression (GTEx) [35], and UCSC xena [36] were used to validate our key results in transcriptome level, Cancer Cell Line Encyclopedia (CCLE) [37] were used to validate our results in cellular level, The human protein atlas [38] was used to verify gene expression in tissue level. And the Protein-Protein Interaction (PPI) network was plotted by String database [39]. The ChIP-seq analysis was utilized in Cistrome data browser [13] based on seven studies [40–44], the chromatin location and targeted drugs for identified prognosis-related DEEs were explored in eRic database [10].

Single-cell RNA sequencing transcriptome analysis

The single-cell RNA sequencing (scRNA-seq) data of human OS (accession number: GSE162454) were downloaded from Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE162454), including 6 OS patients [45]. All data were integrated using "IntegrateData" function and analyzed by utilizing the R toolkit Seurat (http://satijalab.org/seurat/). Those single cells were extracted for the subsequent analysis which had more than 100,000 transcripts expressing. After the top 2000 variable genes were identified by "vst" method, "FindConservedMarkers", and "FindMarkers" function, the marker genes of each cell type were determined. The tumor stem cell markers were also utilized to determine the tumor stem cells. Data dimensionality was reduced by utilizing principal component analysis (PCA), and the top 20 principal components (PCs) were extracted for the further clustering analysis and Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) analysis. "CellCycleScoring" function and cell cycle-related gene markers were used to determine the cell cycle stages of each cell type. Eventually, "iTALK" package [46] was utilized to determine the ligand-receptor interactions between different cell types, and the "edgebundleR" package (https://github.com/garthtarr/edgebundleR) was utilized to visualize the intercellular communication.

Statistics analysis

In this study, two-side p value less than 0.05 was considered as statistically significant in all analysis processes. And R software was utilized to analyze (www.r-project.org; version 3.6.1; Institute for Statistics

and Mathematics, Vienna, Austria) (Package: limma, edgeR, ggplot2, survminer, survival, rms, randomForest, pROC, glmnet, pheatmap, timeROC, vioplot, corrplot, ConsensusClusterPlus, forestplot, survivalROC, beeswarm, edgeR, chromVAR, Biostrings, BSgenome.Hsapiens.UCSC.hg38, ChIPseeker, TxDb.Hsapiens.UCSC.hg38.knownGene, clusterProfiler, org.Hs.eg.db, karyoploteR, GSVA, GSEABase, stringr, GEOquery, dplyr, ComplexHeatmap, RColorBrewer).

Results

Differential expression analysis

All analysis processes were illustrated in the flow chart (Fig. 1). Based on edge R algorithm, 3981 DEGs were identified between normal and OS samples in the heatmap and volcano plot (Fig. 2A, B). GO and KEGG enrichment analyses were conducted using R's cluster Profiler software package. The most significant GO items of biological processes (BPs), cellular components (CCs), and molecular functions (MFs) were regulation of hemopoiesis, extracellular matrix, and receptor ligand activity, respectively (Fig. 2C). Cytokine-cytokine receptor interaction

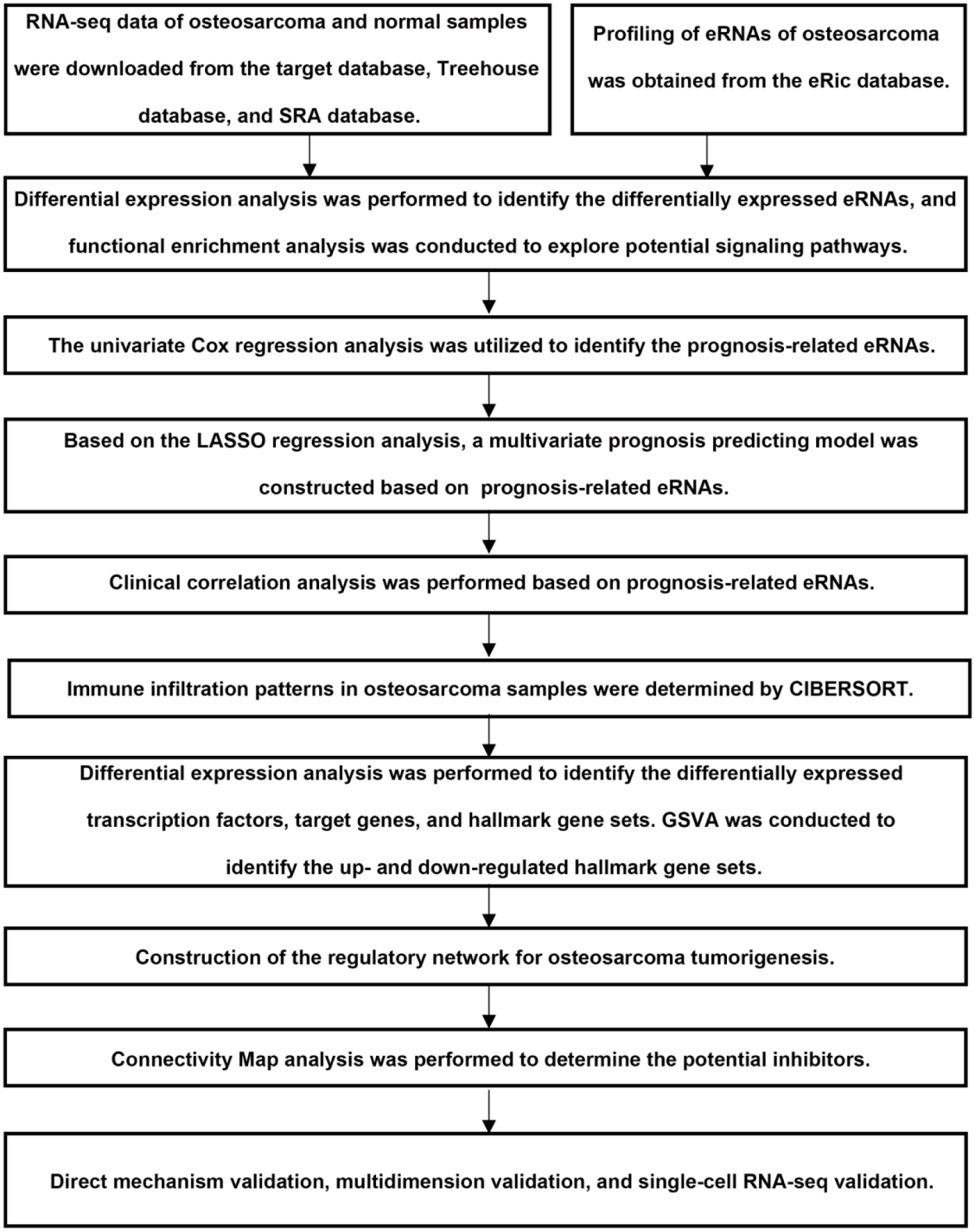


Fig. 1. The flow chart of whole analysis.

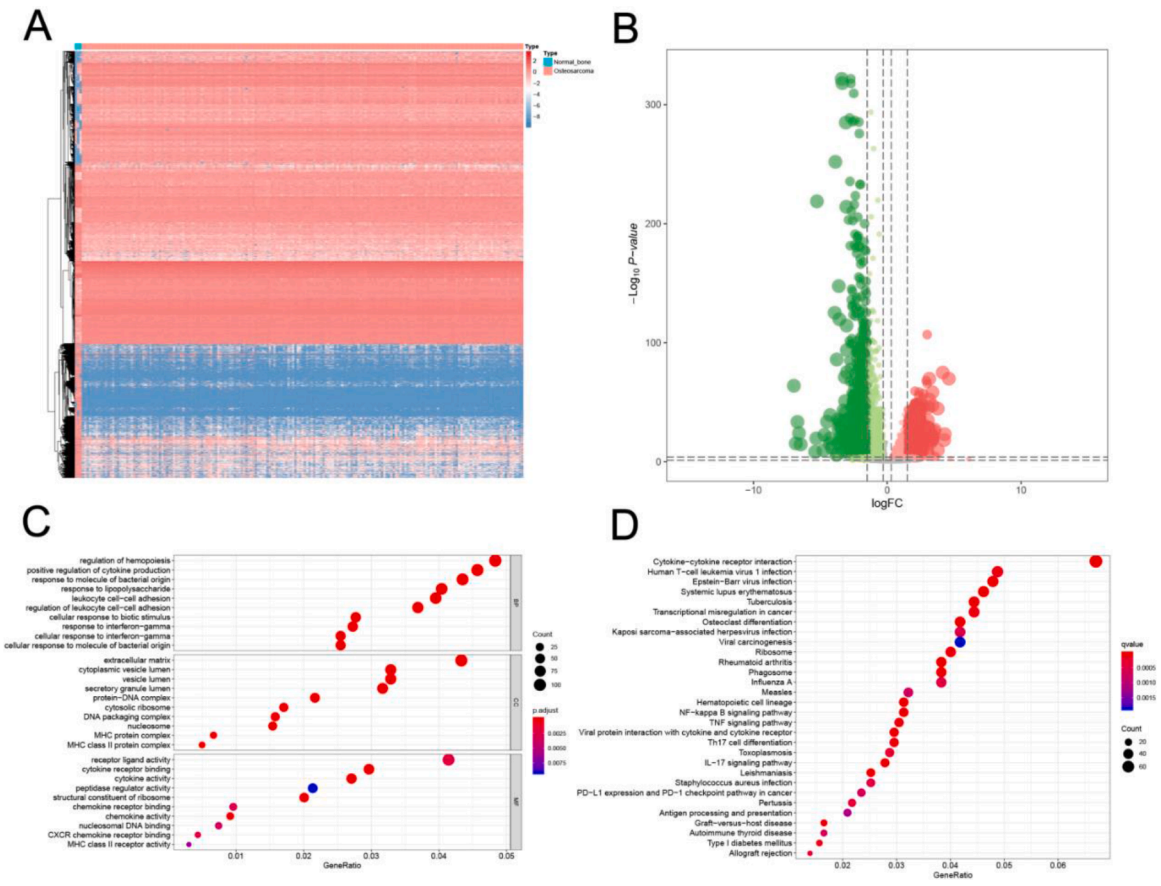


Fig. 2. The differential analysis of all genes. The heatmap plot (A), volcano plot (B), Gene Oncology (GO) enrichment analysis (C) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (D).

was the most critical KEGG pathway, in which most DEGs were enriched (Fig. 2D). Similarly, 468 DEEs were identified, and their expression level was demonstrated in the heatmap and volcano plot (Fig. 3A, B).

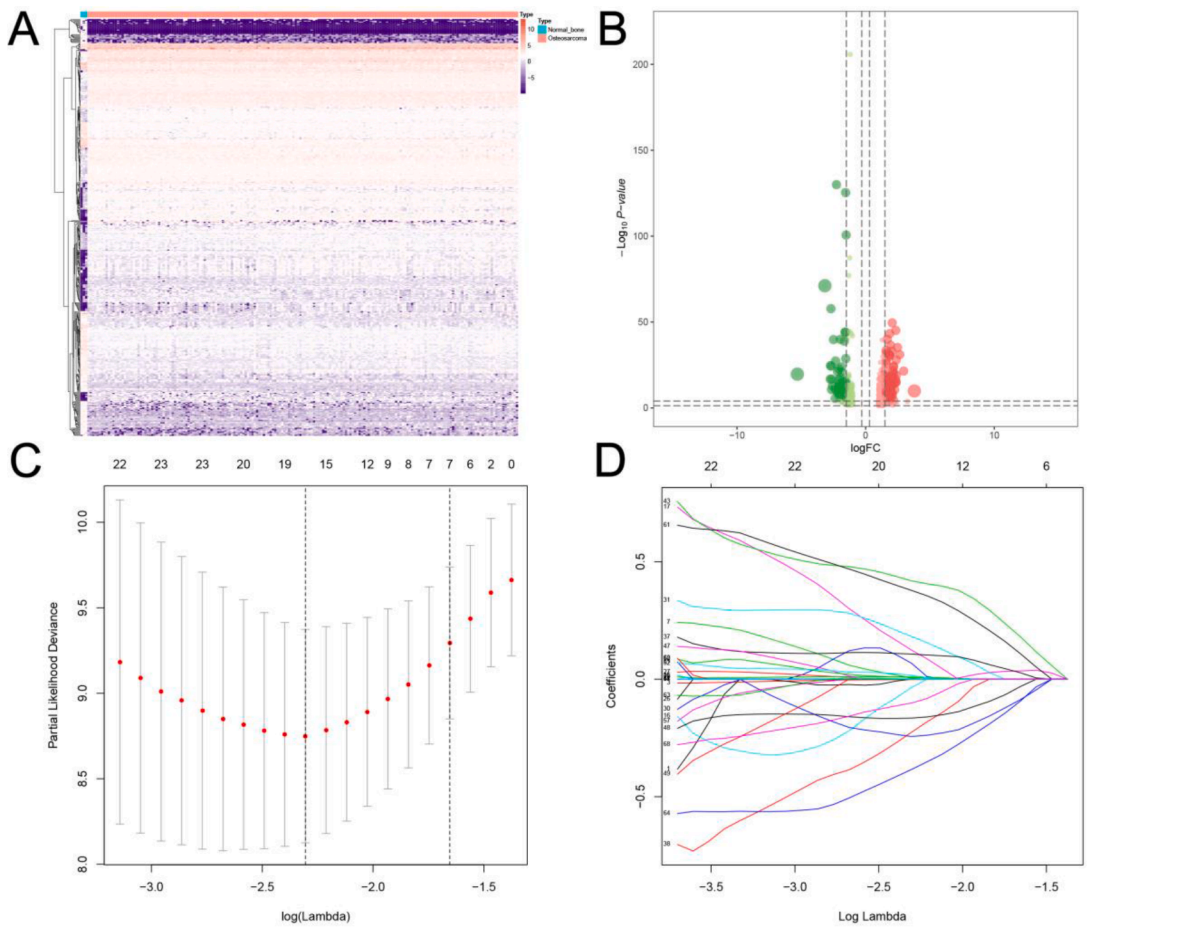
Univariate cox regression and multivariate cox model construction

A total of 468 DEEs were screened by univariate Cox regression analysis, and 72 prognosis-related DEEs were identified. In addition, with the aim of avoiding over-fitness, the Lasso regression was performed to determine lambda coefficients of the model (Fig. 3C, D). 18 key prognosis-related DEEs were integrated into multivariate Cox regression, which were critical for model fitting according to p-value $<$ 0.05, which were shown in the forest map (Fig. 3E). Samples were divided into low- and high-risk groups according to the median of risk score. And the risk scatter plot (Fig. 4A), risks line plot (Fig. 4B) showed the distribution of risk score among all OS patients. Kaplan–Meier survival curve showed risk score for overall survival had prognostic value for OS patients with the survival time of the high-risk group was significantly shorter than that of the low-risk group (Fig. 4C, $p < 0.001$). Besides, the ROC curve illustrated that the area under curve (AUC) in the prognostic model was 0.896, which validated that the risk score model exhibited a stable performance (Fig. 4D). Furthermore, through principal component analysis (PCA) of low-risk and high-risk groups based on respective median risk score, it also showed that OS patients in different risk groups were clearly distributed into two directions (Fig. 4E). Then, we conducted univariate Cox regression analyses and multivariate Cox regression analyses to determine whether clinical parameters (gender, age, metastasis, and primary site) and the risk score were independent prognostic factors of OS patients' overall survival. Finally, the univariate (HR = 90.635, 95%CI (16.334-502.922), $p <$ 0.001) (Fig. 4F) and multivariate (HR = 1.006, 95%CI (1.003–1.009), $p < 0.001$) (Fig. 4G) Cox regression analyses showed the risk score was an independent prognostic predictor for OS patients. The data also showed that metastasis was an independent prognostic predictor in the univariate (HR = 4.613, 95%CI (2.020–10.537), $p < 0.001$) (Fig. 4F) and multivariate (HR = 3.580, 95%CI (1.411–9.084), $p = 0.007$) (Fig. 4G) Cox regression analyses.

Clinical correlation analysis and the immune responses

Here, we compared the differences between expression of prognosis-related DEEs and clinical pathologic characteristics. The boxplots illustrated the clinical correlation analysis of 18 prognosis-related DEEs among first event (Fig. 5A) and metastasis (Fig. 5B). The expression of ACKR3, LINC00598, BMP8B, and TDRP for the censored OS patients were significantly lower than those for the dead patients, whereas the expression of UBB, AMACR, CDK6, CTNNBIP1, and CD8A was significantly higher in censored patients compared to dead patients (all $p <$ 0.05). The same phenomenon was identified for metastasis. The expression of PDPN, RPS15A for the non-metastatic OS patients were significantly lower than those for the metastatic patients, whereas the expression of AMACR, CDK6, and CTNNBIP1 was significantly higher in non-metastatic OS patients compared to metastatic OS patients (all $p <$ 0.05).

Further, the relationship between the expression of prognosis-related DEEs and tumor-infiltrating immunocytes was explored, and a summary of the cell compositions in OS patients and normal samples was depicted using CIBERSORT algorithm. The proportions of 8 types of immune cells in OS patients and normal samples were illustrated in the bar plot, encompassing B cells, cancer associated fibroblasts, CD4+ T cells, CD8+



E

	pvalue	Hazard ratio
UBB	0.00641	0.94525(0.90774−0.98430)
ACKR3	0.04565	1.19445(1.00344−1.42182)
LINC00598	0.01313	5.73065(1.44225−22.77024)
CD274	0.01308	0.17581(0.04454−0.69392)
MTG1	0.00184	2.28459(1.35844−3.84217)
RPS7	0.02650	1.06025(1.00685−1.11648)
SAMD4A	0.00508	1.86173(1.20535−2.87553)
BMP8B	0.00055	1.41916(1.16358−1.73087)
CTTNBP2NL	0.00148	0.23166(0.09403−0.57074)
FAAH	0.00014	2.82563(1.65679−4.81908)
AMACR	0.00026	0.30066(0.15771−0.57319)
CDK6	0.00156	0.52742(0.35480−0.78404)
SLC39A14	0.00022	1.78824(1.31332−2.43489)
PDPN	0.00901	1.25671(1.05867−1.49180)
TDRP	0.00031	1.87996(1.33368−2.65000)
CTNNBIP1	0.00031	0.29735(0.15380−0.57489)
RPS15A	0.04270	1.07674(1.00243−1.15655)
CD8A	0.00849	0.32083(0.13759−0.74808)

Fig. 3. The differential analysis of eRNAs. The heatmap plot (A), volcano plot (B), Lasso regression (C-D) and univariate Cox regression analysis (E).

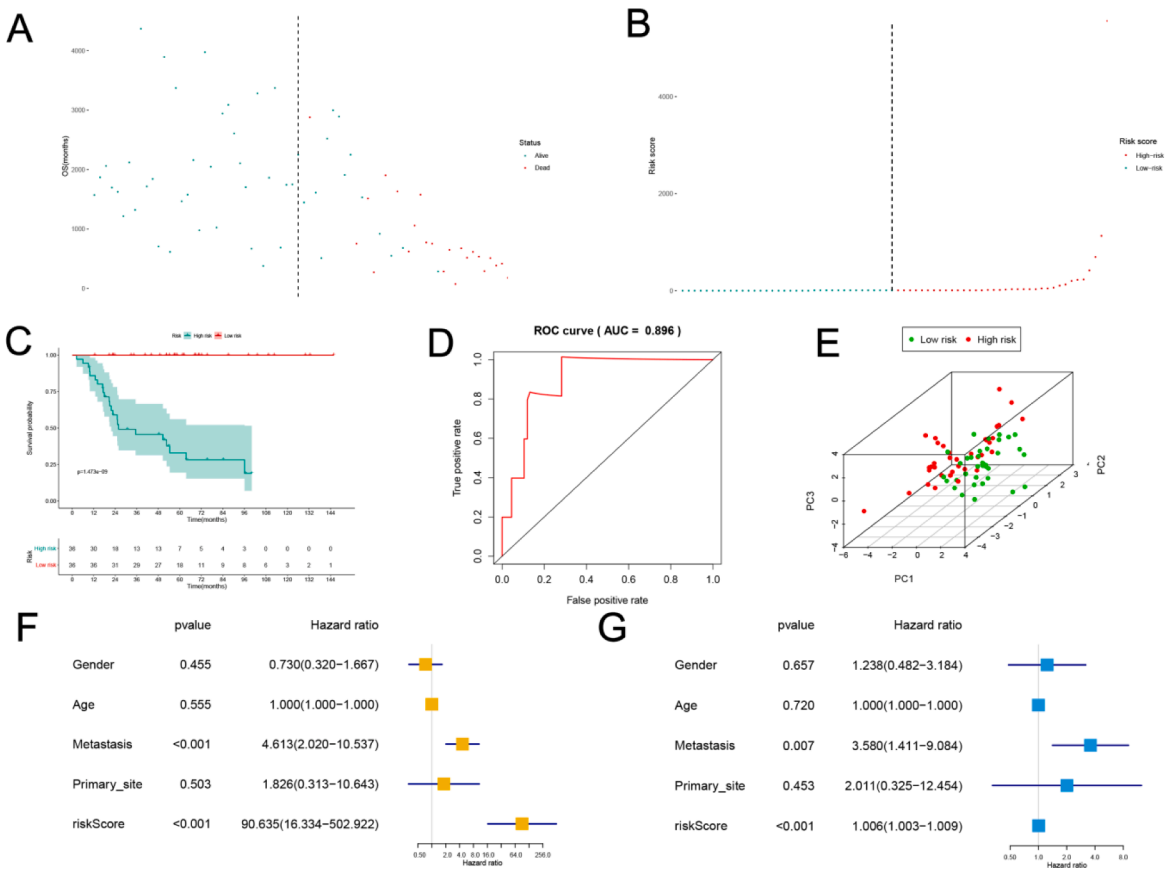


Fig. 4. The multivariate Cox regression analysis of prognosis eRNA. The risk scater plot (A), risk line plot (B), Kaplan-Meier curve (C), ROC curve (D), risk score PCA (E), univariate (F) and multivariate (G) Cox regression analysis.

T cells, endothelial cells, macrophages, NK cells, and uncharacterized cells (Fig. 5C). Compared with normal samples, infiltration degrees of endothelial cells ($p < 0.05$) were increased in OS patients, which suggested that these immune cells had a significantly prognostic value for OS (Fig. 5D). Moreover, immune cells were co-analyzed (Fig. 5E). Importantly, infiltration of cancer associated fibroblasts was negatively related to that of CD4+ T cells (R = -0.73), whereas infiltration of macrophages was positively correlated with infiltration of endothelial cells ($R = 0.70$).

Identification of the prognosis-related DEEs co-expressed DETFs, pathways, and DETGs

To explore DETFs, the edge R algorithm was performed. A total of 51 DETFs was selected, the expression of which was illustrated in the heatmap (Fig. 6A) and volcano plot (Fig. 6B). TFAP2A, BATF, CBX2, IRF5, LYL1, CEBPA, CBX8 and HOXB7 were lowly-expressed in normal tissue, as compared with the tumor samples. Similarly, differentially expressed hallmark gene sets were also filtered by the edge R algorithm, and a total of 27 differentially expressed hallmark gene sets were identified, which were shown in the heatmap (Fig. 6C) and volcano plot (Fig. 6D). What' more, the correlation of GSVA score of hallmark gene sets and OS was investigated (Fig. 6E). Additionally, immune cells and immune functions were explored by ssGSEA to identify the correlations between the OS patients and normal samples with tumor immune characteristics. Specifically, 29 immune cells or immune functions were incorporated to deconvolve the abundance of diverse immune responses in OS patients and normal samples (Fig. 5F). Eventually, 63 DETGs were screened by the edge R algorithm, which were shown in the heatmap plot (Fig. 6G) and volcano plot (Fig. 6H).

The heatmap illustrated the expression of prognosis-related DEEs, DETFs, and DETGs in Fig. 7A. A total of six different dimension regulatory network was constructed based on 4 prognosis-related DEEs, 27 DETFs, 25 DETGs, 8 immune cells by CIBERSORT, 28 immune gene sets by ssGSEA, and 42 hallmark pathways by GSVA, which illustrated the potential regulatory relationships among these factors (Fig. 7B). Interestingly, four key prognosis-related DEEs (CD8A, CDK6, FAAH, and SAMD4A) exhibited significant co-expression patterns in the six different dimension regulatory network. It indicated that these prognosis-related DEEs may play important roles in the tumorigenesis of OS. Furthermore, the interaction coefficients among these components were shown by the heatmap using Pearson correlation analysis (Fig. 7C). Specifically, our data showed that, CD8B ($R = 0.789$, $p < 0.001$) was the significant eRNA for CD8A. CEBPA was the significant TF ($R = 0.633$, $p < 0.001$) (yellow arrow) for CD8A, CD3E was the significant target gene ($R = 0.913$, $p < 0.001$) (pink hexagon) for CD8A, CD8+ T cell was the significant immune cell type ($R = 0.643$, $p < 0.001$) (green triangle) for CD8A by CIBERSORT, CD8+ T cell was the significant immune cell type ($R = 0.969$, $p < 0.001$) (violet ellipse) for CD8A by ssGSEA, and allograft rejection was the significant hallmark gene set ($R = 0.733$, $p < 0.001$) (blue rectangle) for CD8A. These factors had the most significant co-expression relationships within the regulatory network, which may play crucial roles in the tumorigenesis of OS.

Prognosis-related DEEs predict clinical outcomes of OS patients

Based on the publicly available OS ncRNA datasets, we strikingly observed that our candidate prognosis-related DEEs can robustly predict the prognosis of patients (Fig. 7D). Importantly, upregulation of CD3D ($p = 0.009$), CDK6 ($p = 0.32$), and SAMD4A ($p = 0.009$) was associated with worse prognosis of OS patients, whereas upregulation of CD8A ($p < 0.001$), CD8B ($p = 0.040$), and FAAH ($p < 0.001$) was correlated with

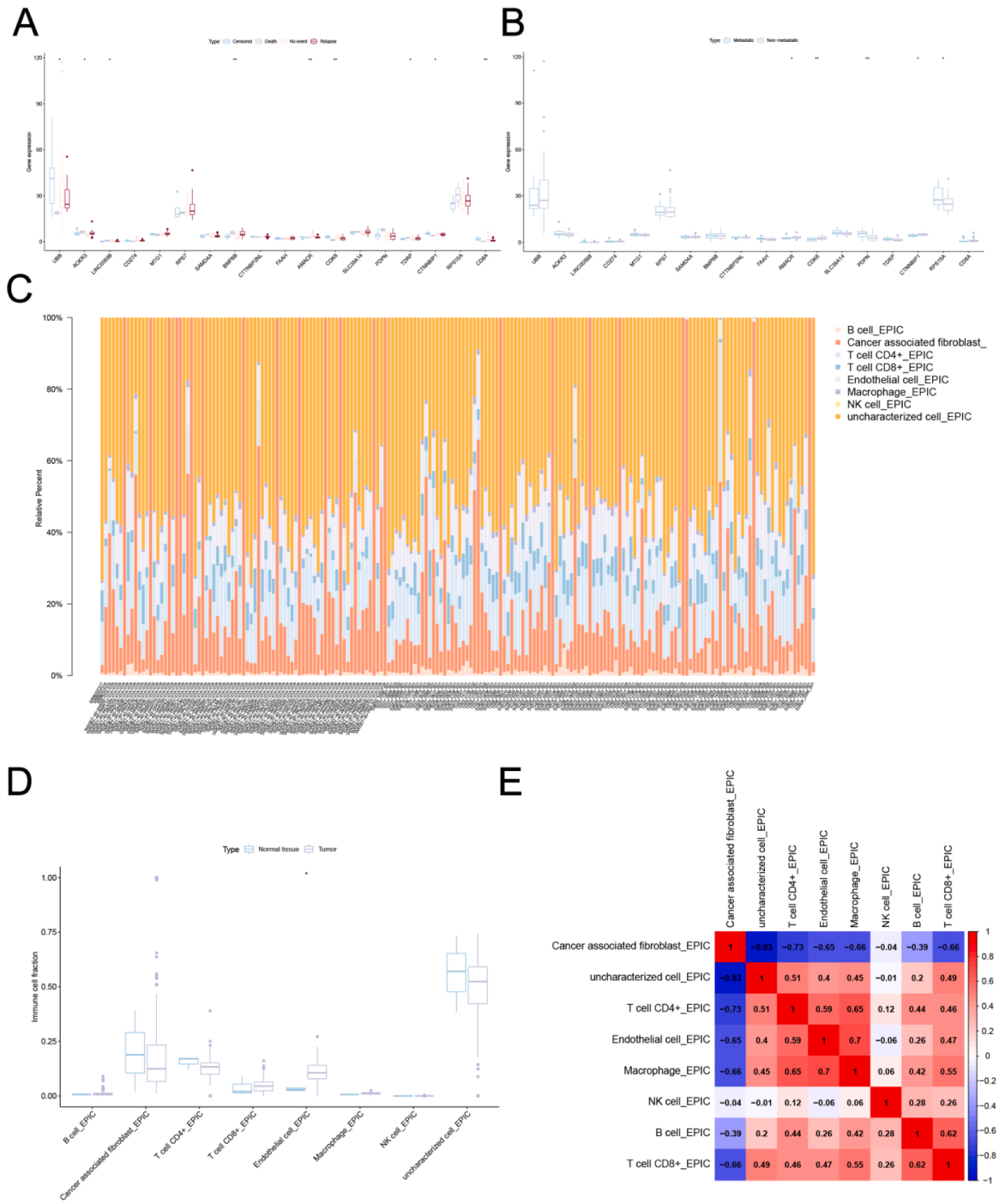


Fig. 5. Analysis for clinical correlation and co-expression. The correlation with first event (A) and disease at diagnosis (B). the barplot (C), boxplot (D) and co-analysis (E) of CIBESORT.

better prognosis of OS patients.

Identification of candidate small molecule inhibitors

To explore the potential inhibitors of prognosis-related DEEs as well as their co-factors in OS, the CMap analysis (Fig. 7E) was applied, and GW-8510 was the most significant inhibitor (mean = 0.670, $p < 0.001$).

ATAC-seq and ChIP-seq validation

Then, ATAC-seq assay was performed to depict the accessible chromatin sites for CD8A, CDK6, FAAH and SAMD4A in OS tissue, indicating the strong chromatin accessibility of these key prognosis-related DEEs (Fig. 8A). Additionally, the accessible chromatin sites for these prognosis-related DEEs in the pan-cancer level were also demonstrated based on ATAC-seq (Fig. S13). Furthermore, the UCSC Genome Browser tracks showed significant enrichment of H3K27ac on multiple loci in the prognosis-related DEEs identified in the present study (CD8A, CDK6, FAAH, and SAMD4A) (Fig. 8B).

Multidimensional validation

Serval online databases were utilized to validate the expression level and the prognostic value of key biomarkers identified in our study. By

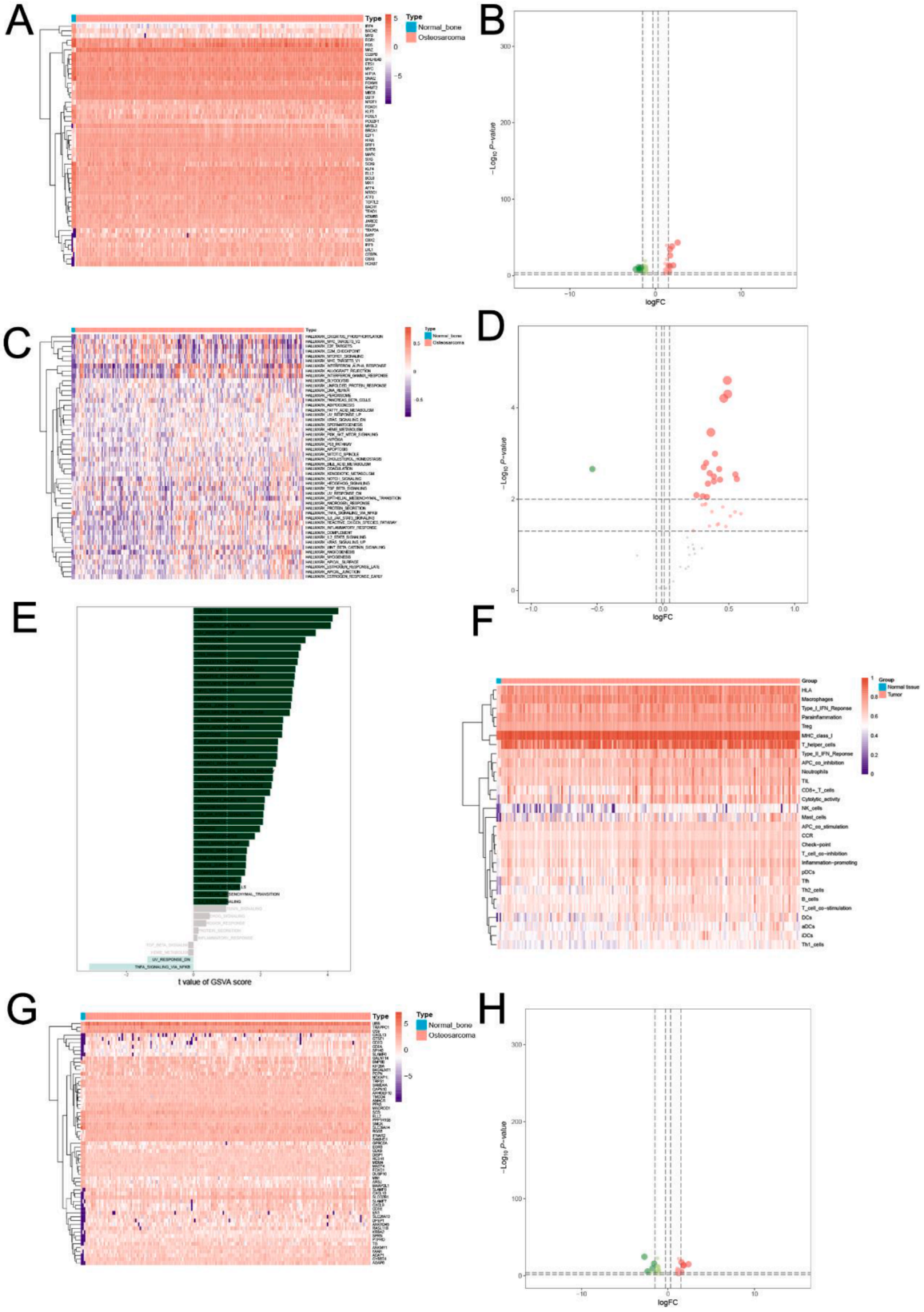


Fig. 6. The differential analysis of transcription factors, target genes, immune cells and hallmark gene sets. The heatmap plot (A) and volcano plot (B) of transcription factors. The heatmap plot (C) and volcano plot (D) of hallmark gene sets. The GSVA analysis (E) of hallmark gene sets. The heatmap plot (F) immune cells in ssGSEA. The heatmap plot (G) and volcano plot (H) of target genes.

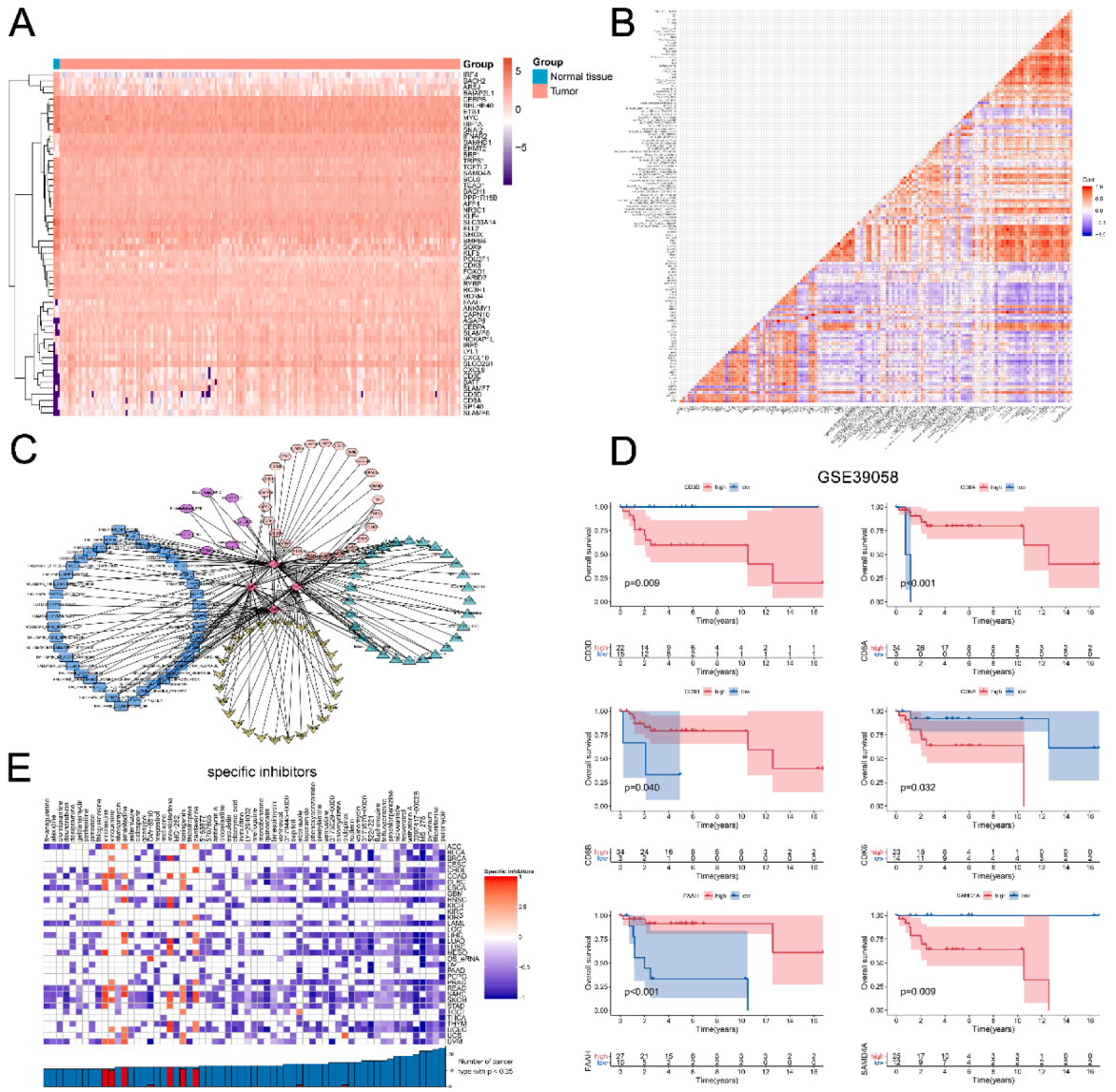


Fig. 7. Construction of regulatory network for OS tumorigenesis. The heatmap plot (A) of key genes and gene sets in the network. The heatmap plot (B) of prognosis-related DEEs, DETFs, DETGs, immune cells, and gene sets in the network by Pearson correlation analysis. The six different dimension regulatory network for OS tumorigenesis (C), and blue rectangles represent hallmark gene sets, violet ellipses represent immune cells in ssGSEA, pink hexagons represent target genes, green triangles represent immune cells in CIBESORT, yellow arrows represent transcription factors, and red diamonds represent eRNAs. Evaluation of the clinical relevance of prognosis-related DEEs (D). The cMAP analysis (E) for osteosarcoma.

utilizing GeneCards database (www.genecards.org/), we identified that the HLA-DMA and HLA-DMB were the top two marker genes of allograft rejection (hallmark gene set). Besides, the CellmMarker was applied to determine the cell markers of CD8+ T cells, and CD8A, CD3E, and CD3D were selected. The results were illustrated in Figs. S1–S11. The correlations among CEBPA, CD8A, CD3E, CD8B, CD3D, HLA-DMA, and HLA-DMB were shown in Table S1, the expression level was summarized in Table S2, and their prognostic value was summarized in Table S3. CEBPA ($p = 0.029$, figure S1H; $p = 0.026$, figure S4A; $p < 0.001$, figure S5F; $p < 0.001$, figure S6A), CD8A ($p = 0.026$, figure S4B), CD3E ($p = 0.037$, figure S1J; $p = 0.012$, figure S4C), CD8B ($p = 0.030$, figure S1K; $p = 0.002$, figure S4D), CD3D ($p < 0.001$, figure S4E), HLA-DMA ($p = 0.031$, figure S1M; $p = 0.029$, figure S4F; $p = 0.033$, figure S6F), HLA-DMB ($p = 0.016$, figure S1N; $p = 0.033$, figure S6G), and integrated genes ($p < 0.001$, figure S5M) had significant prognostic value. We strikingly identified that these key eRNAs can robustly predict the prognosis of OS patients (Fig. 4B).

Moreover, the ChIP-seq validation were performed based on the Cistrome database, and peaks of binding domain between CEBPA (chr6:52003572-52426257) and CD8A (chr2:86,784,611-86,808,396) were showed in the Fig. S12. And targeted drugs for CD8A were predicted based on eRic database and summarized in Table S4.

Single-cell RNA-seq transcriptome analysis

Unsupervised clustering clearly identified 7 cell clusters based on tumor samples from 6 OS patients, including B cell, cancer associated fibroblast (CAF), endothelial cell, myeloid cell, NK/T cell, OS cell, and osteoclast (Fig. 9A). The average cell number and proportions of 7 cell clusters diverse well among the 6 OS patients (Fig. 9B). Furthermore, the top highly expressed gene markers for each cell cluster were determined by comparing the expression profile of the target cluster with the rest of cells based on non-parametric Wilcoxon Rank Sum test ($p < 0.01$ and FC > 2). Fig. 9D illustrated the up-regulated or down-regulated genes in the 7 cell clusters. The dot plots showed the proportion of cells expressing LYZ (cancer marker) [47], CD3D (T cell marker) [48], ALPL (bone

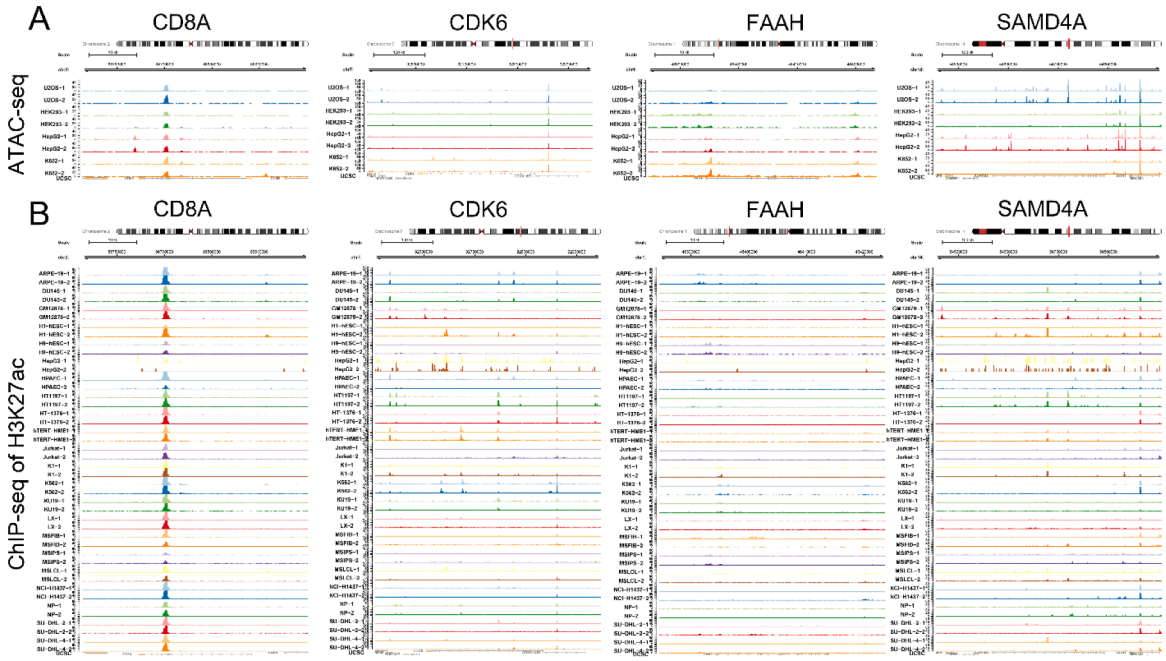


Fig. 8. The ChIP–seq validation for CD8A, CDK6, FAAH, and SAMD4A in OS tissue (A). The ATAC-seq validation for CD8A, CDK6, FAAH, and SAMD4A in OS tissue (B).

remodeling indicator or marker of osteoblast activity) [49], and ACP5 (progression and metastasis-related oncogene) [50] and their scaled relative expression level in 7 cell clusters (Fig. 10A). Specifically, LYZ, ALPL, and ACP5 was highly expressed in OS cell clusters, indicating highly malignant behaviors and osteogenic activities of these tumor cells. As demonstrated in the top variable genes, specifically OS cell clusters expressed high levels of stemness-related gene markers (CD24, CD44, and MKI67) [51–53] (Fig. 10B). The expression profiling of prognosis-related DEEs in distinct cell subclusters were presented (Fig. 10B). As demonstrated in the feature plots, specifically NK/T cell clusters expressed high levels of CD3D, CD3E, CD8A, CD8B, while CAF cluster expressed high levels of SAMD4A (Fig. 10B). Additionally, myeloid cell clusters exhibited high expression of HLA-DMA and HLA-DMB, which were the representative genes involved in our candidate DEE-associated pathway (allograft rejection) (Fig. 10B). All these findings showed that our candidate prognosis-related DEEs and their cofactors were extensively expressed in the immune microenvironment of OS, which were potential targets for the treatment of OS. Cell cycle distribution of the 7 cell clusters was illustrated in the UMAP plot (Fig. 10C). OS cells were mainly in G2 phase and S phase. The ligand-receptor plot showed ligand-receptor interactions across these cell clusters (Fig. 10D). It showed that OS cells and CAF occupied the cellular communication center and mostly sent out signals to other cell clusters and even to themselves, whereas endothelial cells and osteoclasts turned more role on receptors. 6 genes (ligands: COL1A1, COL1A2, SPP1; receptors: ITGA1, ITGA5, and CD36) showed the most significant ligand-receptor function on cellular communication, which were potential targets for tumor treatment.

Discussion

OS is a rare malignant disease that occurs mainly in adolescents, and the pain, the pathological fracture decreases the life quality [1]. What' worse, the prognosis gets poor as the consequence of distant metastasis [1]. However, the clear pathogenic mechanism remains unclear, more signatures and targets are required to early diagnose, precise therapy target and predict the prognosis. eRNAs are generated during the transcription process of active enhancer [10]. Importantly, expression patterns of eRNAs are specific to cancer types in human cancers [54]. In prostate cancer, multiple eRNAs was validated to be differentially expressed [55]. In breast cancer cell lines, transcription of eRNAs induced by estrogen was significantly upregulated [56]. Nevertheless, a recent study indicated that expression of eRNAs was significantly inhibited in throat cancer [10]. Upregulation of oncogenes or oncogenic pathways was related to the aberrant generation of eRNAs in various human cancers, and eRNAs may play a broad role in the pathophysiology of OS.

Based on edge R algorithm, 3981 DEGs, 468 DEEs, 51 DETFs, and 27 differentially expressed hallmark gene sets were identified. A total of 468 DEEs were screened by univariate Cox regression analysis, and 72 prognosis-related DEEs were identified. After the LASSO regression, 18 key prognosis-related DEEs were integrated into multivariate Cox regression. In addition, the Kaplan-Meier curve ($p < 0.001$) and the ROC curve (AUC = 0.896) were utilized to access the accuracy of risk score and detect the discrimination of the multivariate model. Additionally, the Pearson analysis was performed to plot the six different dimension regulatory network, which included 4 prognosis-related DEEs, 27 DETFs, 25 DETGs, 8 immune cells by CIBERSORT, 28 immune gene sets by ssGSEA, and 42 hallmark pathways by GSVA. CD8A, CDK6, FAAH and SAMD4A were the most significant prognosis-related DEEs, which occupied the communication center of the regulatory network. CD8B ($R = 0.789$, $p < 0.001$), CEBPA ($R = 0.633$, $p < 0.001$), CD3E ($R = 0.913$, $p < 0.001$), CD8+ T cell ($R = 0.643$, $p < 0.001$; $R = 0.969$, $p < 0.001$) and allograft rejection ($R = 0.733$, $p < 0.001$) were the significant eRNA, TF, target gene, immune cell, and hallmark gene sets for CD8A, respectively. Based on the CMap analysis, GW-8510 was the most significant inhibitor for OS (mean = 0.670, $p < 0.001$). Then, ATAC-seq was performed to depict the accessible chromatin sites for CD8A, CDK6, FAAH, and SAMD4A, which demonstrated the strong chromatin accessibility of these key prognosis-related DEEs. Further, external validation based on multidimensional online databases was performed to verify our key findings, which showed the identified biomarkers had important prognostic value and could be used as potential reference markers. Eventually, mapping the single-cell transcriptomic landscape of OS, we explored the intercellular and intracellular communications in detail. It showed that our candidate prognosis-related DEEs played a critical role

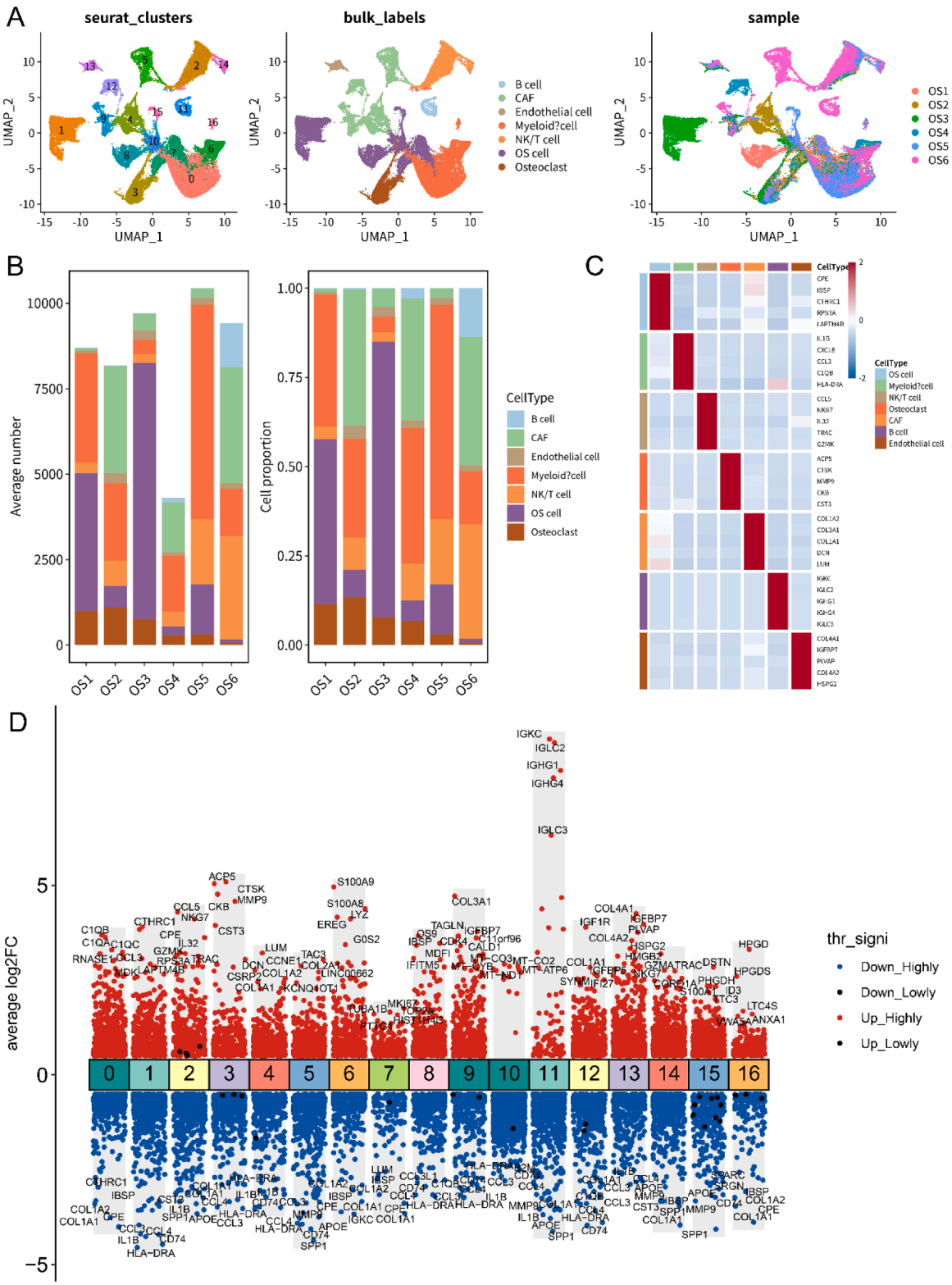


Fig. 9. Single-cell transcriptomic analysis reveals cellular heterogeneity in OS tissue. (A) The uniform manifold approximation and projection for dimension reduction (UMAP) plot of the 17 cell clusters and 7 identified main cell types in tumor tissue collected from 6 OS patients. (B) Average number and cell proportion of 7 cell types in tumor tissue collected from 6 OS patients. (C) Gene co-expression of top 5 genes among the 7 cell types, where the colors from red to blue represented changes from high expression to low expression. (D) Significantly up- or down-regulated genes in the 17 cell clusters.

in the immune microenvironment of OS, which could be used as important reference markers for future research. It indicated highly malignant features and osteogenic activities of OS cells, as well as the extensive cellular communications between OS cells and other cell types, plenty of which were well worth further researches.

Tumor microenvironment plays an important role in the tumor genesis of OS, and T cells and macrophages actually showed the most abundant infiltration in OS [57]. Activated osteoclast cells destroy the

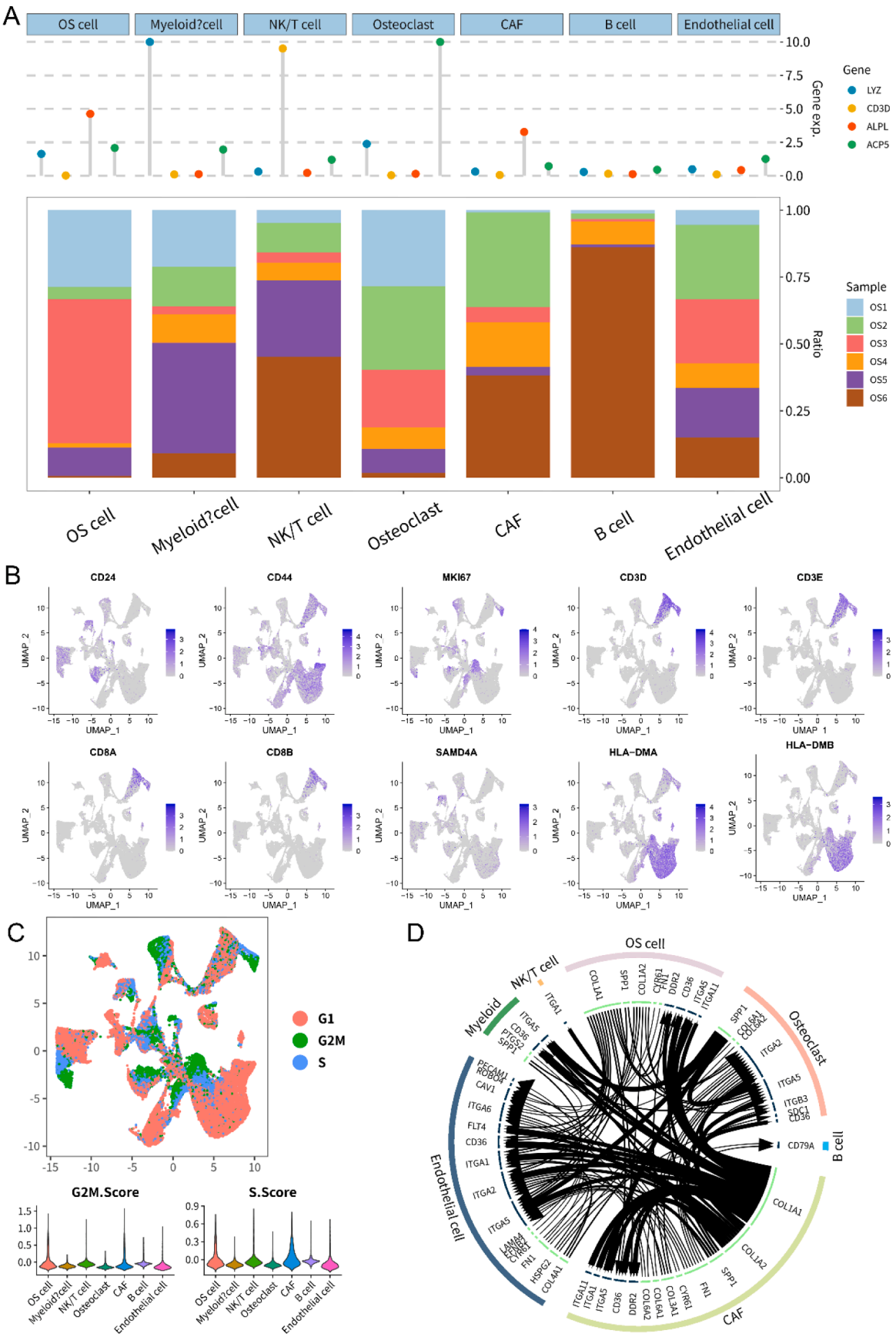


Fig. 10. Subpopulation analysis and intercellular communication analysis.
(A) The proportion of classical marker genes in all cell types. (B) Feature plots of stemness-related biomarkers and our candidate prognosis-related DEEs in all cell types. (C) The cell cycle distribution and the cell cycle score in 7 cell types. (D) The ligand-receptor pairs among 7 cell types.

normal cells and promotes the genesis of OS in early stage, and then they disturb the immune response mediated by T cells [57]. Tumor cells have an impact on the immune infiltrating environment by recruiting and interacting with immune cells, such as PD-L1 on OS cells that can bind with the PD-1/B7-1 in T cells, and then decreases T cell proliferation and increases apoptosis [57]. The ratio of CD8+ T cells in the relapse and metastasis OS was lower than primary OS in the single cell RNA-seq [58].

CEBPA encodes a transcription factor named CCAAT enhancer binding protein alpha, it is not only critical in the differentiation of myeloid lineage, but also can regulate the lineage commitment and differentiation of osteoclast cells [59]. Mesenchymal cells can differentiate into osteoblasts and adipocytes, the abnormal differentiation of which relates to the OS, and CEBPA plays a role in the adipogenesis [1, 60]. Some studies showed that lncRNA CEBPA-AS1 can regulate NCOR2, and then affected the Notch signal pathway and cell apoptosis [61].

CD8A and CD3E encode the cell markers of CD8+ T cells, and they were detected in the tumor microenvironment [62]. CD8+ T cells mediate the anti-tumor effect, and in animal experiments, the decrease of CD8+ T cells was related to the poor survival in OS [63]. Then we proposed that the disfunction of CD8A and CD3E may affect the immune environment for tumorigenesis, which was associated with clinical characteristics, the immune response, and prognosis of OS patients.

Although there was no study reporting the interaction between CEBPA and CD8A, as well as CD8A and CD3E, we proposed that CEBPA may promote the expression of CD8A, and CD8A might target on CD3E as an eRNA, thereby promoting the transcription of CD3E.

Because the allograft rejection aggregates a variety of genes related to allograft rejection, it refers to an immune cell-mediated biological process. HLA-DMA and HLA-DMB are related to the allograft rejection [64]. The tumor microenvironment of OS is related to immune infiltration. Besides, tumor cells can be the antigen presenting cells to present major histocompatibility complex (MHC) class II-restricted endogenous antigens, and leads to abnormal activated T cells [65]. Besides, HLA-DMA and HLA-DMB were the main components of the MHC II, HLA-DMB was positively related to the CD8+ T cells and may improve the prognosis in patients with ovarian cancer [66]. However, HLA-DMB was defined as an independent prognosis factor for OS [67]. And HLA-DMB was identified as a latent susceptibility gene for Kaposi's sarcoma [68].

We proposed that tumor cells recruited CD8+ T cells and presented the normal components of the normal cells to CD8+ T cells, and then depressed the activation of CD8+ T cells to escape the immune system.

GW-8510 is an inhibitor of Cdk2 (cyclin kinase 2) [69], which regulates the cell cycle. And it can be applied to the therapy of non-small cell lung cancer and colorectal cancer [70,71].

In summary, the four prognosis-related DEEs were identified to have independent prognostic significance for tumorigenesis of OS. CEBPA may regulate CD8A, thereby activating the transcription of CD3E, which could affect the allograft rejection based on tumor infiltrating CD8+ T cells. However, some limitations existed in our analysis, such as the bias of data selection, lack of sample size and sufficient demographic information to perform the survival analysis, and the exploration for mechanism is required. Therefore, a series of multidimensional validation was utilized. And in the future, experiments *in vivo* and *in vitro* to validate the indirect and direct mechanism like gain/loss function, Co-IP will be applied.

Conclusion

Based on the regulatory network for tumorigenesis of OS-related eRNAs, key eRNAs and their co-factors were identified to construct a prognostic model for OS patients. Our findings provided bioinformatics information in exploring the molecular mechanisms of the tumorigenesis of OS. We speculated that CEBPA (DETF) may regulate CD8A (DEE), thereby promoting the transcription of CD3E, which may affect allograft rejection based on CD8+ T cells.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This study was supported in part by the National Natural Science Foundation of China (Grant No. 81772856); Youth Fund of Shanghai Municipal Health Planning Commission (No.2017YQ054; 2017Y0117); Interdisciplinary Program of Shanghai Jiao Tong University (No. YG2017MS26); Shanghai Talent Development Fund (No.2018094); Henan medical science and technology research project (No. 201602031); Key project of provincial and ministerial co-construction of Henan Medical Science and Technology (No. SBGJ202002031). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

Data for analysis were from Target database, Treehouse database TCGA database, MET500 database, GEO database including samples or series with ID of GSM2104126, GSM1315482, GSM2104231, GSM1290087, GSM1131244, GSM1187164, GSM1290086, GSE162454, GSE39058, GSE139099, GSE139190, and GSE134744.

CRediT authorship contribution statement

Penghui Yan: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Zhenyu Li:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Shuyuan Xian:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Siqiao Wang:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Qing Fu:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Jiwen Zhu:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Xi Yue:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Xinkun Zhang:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Shaofeng Chen:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Wei Zhang:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Jianyu Lu:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Huabin Yin:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Runzhi Huang:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Zongqiang Huang:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We thank the TCGA team of the National Cancer Institute for using their data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2022.101499.

References

[1] J. Ritter, S.S. Bielack, Osteosarcoma, Ann. Oncol. 21 (7) (2010) vii320–vii325. **Suppl**.
[2] R.L. Siegel, et al., Cancer statistics, 2021, CA Cancer J. Clin. 71 (1) (2021) 7–33.
[3] L.M. Kelley, et al., Pathological fracture and prognosis of high-grade osteosarcoma of the extremities: an analysis of 2847 Consecutive Cooperative Osteosarcoma Study Group (COSS) patients, J. Clin. Oncol. 38 (8) (2020) 823–833.
[4] M. Kansara, et al., Translational biology of osteosarcoma, Nat. Rev. Cancer 14 (11) (2014) 722–735.
[5] I. Fernandes, et al., Osteosarcoma pathogenesis leads the way to new target treatments, Int. J. Mol. Sci. 22 (2) (2021) 813, https://doi.org/10.3390/ijms22020813.
[6] T.K. Kim, et al., Widespread transcription at neuronal activity-regulated enhancers, Nature 465 (7295) (2010) 182–187.
[7] V. Sartorelli, S.M. Lauberth, Enhancer RNAs are an important regulatory layer of the epigenome, Nat. Struct. Mol. Biol. 27 (6) (2020) 521–528.
[8] W.B. Li, D. Notani, M.G. Rosenfeld, Enhancers as non-coding RNA transcription units: recent insights and future perspectives, Nat. Rev. Genet. 17 (4) (2016) 207–223.
[9] R.Z. Huang, et al., Identification of key eRNAs for spinal cord injury by integrated multinomial bioinformatics analysis, Front. Cell Dev. Biol. 9 (2021), 728242, https://doi.org/10.3389/fcell.2021.728242.
[10] Z. Zhang, et al., Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer, Nat. Commun. 10 (1) (2019) 4562.
[11] G.K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments, Stat. Appl. Genet. Mol. Biol. 3 (2004) Article3.
[12] G. Yu, L.G. Wang, Q.Y. He, ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization, Bioinformatics 31 (14) (2015) 2382–2383.
[13] R. Zheng, et al., Cistrome data browser: expanded datasets and new tools for gene regulatory analysis, Nucleic Acids Res. 47 (D1) (2019) D729–D735.
[14] A. Liberzon, et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection, Cell Syst. 1 (6) (2015) 417–425.
[15] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (10) (2010), R106, https://doi.org/10.1186/gb-2010-11-10-r106.
[16] K.S. Hung, et al., Functional enrichment analysis based on long noncoding RNA associations, BMC Syst. Biol. 12 (4) (2018) 45. Suppl.
[17] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (1) (2010) 1–22.
[18] Y. Wu, et al., A risk score model with five long non-coding RNAs for predicting prognosis in gastric cancer: an integrated analysis combining TCGA and GEO datasets, PeerJ 9 (2021) e10556.
[19] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.
[20] P. Charoentong, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, Cell Rep. 18 (1) (2017) 248–262.
[21] D.A. Barbie, et al., Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1, Nature 462 (7269) (2009) 108–112.
[22] D. Hnisz, et al., Super-enhancers in the control of cell identity and disease, Cell 155 (4) (2013) 934–947.
[23] K.Y. Ngiam, I.W. Khor, Big data and machine learning algorithms for health-care delivery, Lancet Oncol. 20 (5) (2019) e262–e273.
[24] G.P. Diller, et al., Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients, Eur. Heart J. 40 (13) (2019) 1069–1077.
[25] A. Meyer, et al., Machine learning for real-time prediction of complications in critical care: a retrospective study, Lancet Respir. Med. 6 (12) (2018) 905–914.
[26] A.D. Kelly, et al., MicroRNA paraffin-based studies in osteosarcoma reveal reproducible independent prognostic profiles at 14q32, Genome Med. 5 (1) (2013) 2.
[27] J. Lamb, et al., The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease, Science 313 (5795) (2006) 1929–1935.
[28] F. Grubert, et al., Landscape of cohesin-mediated chromatin loops in the human genome, Nature 583 (7818) (2020) 737–743.
[29] Z. Tang, et al., GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, Nucleic Acids. Res. 45 (W1) (2017) W98–W102.
[30] D.R. Rhodes, et al., ONCOMINE: a cancer microarray database and integrated data-mining platform, Neoplasia 6 (1) (2004) 1–6.
[31] D.S. Chandrashekar, et al., UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, Neoplasia 19 (8) (2017) 649–658.
[32] S.V. Vasaikar, et al., LinkedOmics: analyzing multi-omics data within and across 32 cancer types, Nucleic Acids Res. 46 (D1) (2018) D956–D963.
[33] R. Aguirre-Gamboa, et al., SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis, PLoS One 8 (9) (2013) e74250.
[34] E. Cerami, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (5) (2012) 401–404.
[35] G. Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans, Science 348 (6235) (2015) 648–660.
[36] M. Goldman, et al., The UCSC cancer genomics browser: update 2015, Nucleic Acids. Res. 43 (2015) D812–D817. Database issue.
[37] M. Ghandi, et al., Next-generation characterization of the cancer cell line encyclopedia, Nature 569 (7757) (2019) 503–508, https://doi.org/10.1038/s41586-019-1186-3.
[38] M. Uhlen, et al., Proteomics. Tissue-based map of the human proteome, Science 347 (6220) (2015), 1260419.
[39] B. Snel, et al., STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene, Nucleic Acids. Res. 28 (18) (2000) 3442–3444.
[40] M.B. Meyer, et al., Epigenetic plasticity drives adipogenic and osteogenic differentiation of marrow-derived mesenchymal stem cells, J. Biol. Chem. 291 (34) (2016) 17829–17847.
[41] A. Crotti, et al., Mutant huntingtin promotes autonomous microglia activation via myeloid lineage-determining factors, Nat. Neurosci. 17 (4) (2014) 513–521.
[42] C. van Oevelen, et al., C/EBPalpha activates pre-existing and de novo macrophage enhancers during induced Pre-B cell transdifferentiation and myelopoiesis, Stem Cell Rep. 5 (2) (2015) 232–247.
[43] S. Heinz, et al., Effect of natural genetic variation on enhancer selection and function, Nature 503 (7477) (2013) 487–492.
[44] M.S. Hasemann, et al., C/EBPalpha is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors, PLos Genet. 10 (1) (2014), e1004079.
[45] Y. Liu, et al., Single-cell transcriptomics reveals the complexity of the tumor microenvironment of treatment-naive osteosarcoma, Front. Oncol. 11 (2021), 709210.
[46] Yuanxin Wang, R.W., et al, iTALK: an R package to characterize and illustrate intercellular communication. bioRxiv, 2019.
[47] S.K. Sarvestani, et al., Cancer-predicting transcriptomic and epigenetic signatures revealed for ulcerative colitis in patient-derived epithelial organoids, Oncotarget 9 (47) (2018) 28717–28730.
[48] M.J. Shi, et al., High CD3D/CD4 ratio predicts better survival in muscle-invasive bladder cancer, Cancer Manag. Res. 11 (2019) 2987–2995.
[49] C. Wennberg, et al., Functional characterization of osteoblasts and osteoclasts from alkaline phosphatase knockout mice, J. Bone Miner. Res. 15 (10) (2000) 1879–1888.
[50] M. Kawamura, et al., Clinical significance of tartrate-resistant acid phosphatase type-5 expression in human gastric cancer, Anticancer Res. 34 (7) (2014) 3425–3429.
[51] M. Al-Hajj, et al., Prospective identification of tumorigenic breast cancer cells, Proc. Natl. Acad. Sci. U. S. A. 100 (7) (2003) 3983–3988.
[52] C. Li, et al., Identification of pancreatic cancer stem cells, Cancer Res. 67 (3) (2007) 1030–1037.
[53] T. Scholzen, J. Gerdes, The Ki-67 protein: from the known and the unknown, J. Cell. Physiol. 182 (3) (2000) 311–322.
[54] J.H. Lee, F. Xiong, W. Li, Enhancer RNAs in cancer: regulation, mechanisms and therapeutic potential, RNA Biol. 17 (11) (2020) 1550–1559.
[55] J. Zhao, et al., Alterations of androgen receptor-regulated enhancer RNAs (eRNAs) contribute to enzalutamide resistance in castration-resistant prostate cancer, Oncotarget 7 (25) (2016) 38551–38565.
[56] F. Crudele, et al., The network of non-coding RNAs and their molecular targets in breast cancer, Mol. Cancer 19 (1) (2020) 61.
[57] M.F. Heymann, F. Lezot, D. Heymann, The contribution of immune infiltrates and the local microenvironment in the pathogenesis of osteosarcoma, Cell. Immunol. 343 (2019), 103711.
[58] Y. Zhou, et al., Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma, Nat. Commun. 11 (1) (2020) 6322.
[59] W. Chen, et al., C/EBPalpha regulates osteoclast lineage commitment, Proc. Natl. Acad. Sci. U. S. A. 110 (18) (2013) 7294–7299.
[60] L. Wang, et al., Obesity-associated MiR-342-3p promotes adipogenesis of mesenchymal stem cells by suppressing CtBP2 and releasing C/EBPalpha from CtBP2 binding, Cell. Physiol. Biochem. 35 (6) (2015) 2285–2298.
[61] P. Xia, et al., lncRNA CEBPA-AS1 overexpression inhibits proliferation and migration and stimulates apoptosis of OS cells via notch signaling, Mol. Ther. Nucleic Acids 19 (2020) 1470–1481.
[62] S. Chen, et al., Combination of 4-1BB agonist and PD-1 antagonist promotes antitumor effector/memory CD8 T cells in a poorly immunogenic tumor model, Cancer Immunol. Res. 3 (2) (2015) 149–160.
[63] B.J. Biller, et al., Decreased ratio of CD8+ T cells to regulatory T cells associated with decreased survival in dogs with osteosarcoma, J. Vet. Intern. Med. 24 (5) (2010) 1118–1123.
[64] M. Rist, et al., Cross-recognition of HLA DR4 alloantigen by virus-specific CD8+ T cells: a new paradigm for self-/nonself-recognition, Blood 114 (11) (2009) 2244–2253.

[65] L. Qi, J.M. Rojas, S. Ostrand-Rosenberg, Tumor cells present MHC class II-restricted nuclear and mitochondrial antigens and are the predominant antigen presenting cells *in vivo*, J. Immunol. 165 (10) (2000) 5451–5461.
[66] M.J. Callahan, et al., Increased HLA-DMB expression in the tumor epithelium is associated with increased CTL infiltration and improved prognosis in advanced-stage serous ovarian cancer, Clin. Cancer Res. 14 (23) (2008) 7667–7673.
[67] X. Guan, Z. Guan, C. Song, Expression profile analysis identifies key genes as prognostic markers for metastasis of osteosarcoma, Cancer Cell Int. 20 (2020) 104.
[68] B. Aissani, et al., SNP screening of central MHC-identified HLA-DMB as a candidate susceptibility gene for HIV-related Kaposi's sarcoma, Genes Immun. 15 (6) (2014) 424–429.
[69] K. Johnson, et al., Inhibition of neuronal apoptosis by the cyclin-dependent kinase inhibitor GW8510: identification of 3′ substituted indolones as a scaffold for the development of neuroprotective drugs, J. Neurochem. 93 (3) (2005) 538–548.
[70] Y.Y. Hsieh, et al., Repositioning of a cyclin-dependent kinase inhibitor GW8510 as a ribonucleotide reductase M2 inhibitor to treat human colorectal cancer, Cell Death Discov. 2 (2016) 16027.
[71] F. Dong, et al., Downregulation of XIAP and induction of apoptosis by the synthetic cyclin-dependent kinase inhibitor GW8510 in non-small cell lung cancer cells, Cancer Biol. Ther. 5 (2) (2006) 165–170.