Identification of genes that influence poor prognosis in osteosarcoma

Yao Lu
Gannan Medical University

Yanli Li
Zhongshan Hospital, Fudan University

Xueliang Zeng
The First Affiliated Hospital of Gannan Medical University, Gannan Medical University

Bei Li
The First Affiliated Hospital of Gannan Medical University, Gannan Medical University

Qiongjun Xie
Guangzhou Health Science College

Gaolu Cao
Gannan Medical University

Panpan Huang (✉ 11618102@zju.edu.cn)
Gannan Medical University

Research Article

Abstract

Objectives Osteosarcoma (OS) is the most common primary bone cancer in children and adolescents. At present, the 5-year overall survival rate of OS patients is about 65%, and the long-term prognosis is still not ideal. The study was designed to screen genes that could contribute to the poor prognosis of OS and explore their potential pathogenic mechanisms.

Methods The gene expression profile of the GSE94805 dataset from the GEO database, containing data from 12 U2OS cell samples, including four control, four quiescent, and four senescent samples was obtained. Co-expressed differentially expressed genes (DEGs) in OS U2OS cells were selected using the GEO2R tool and Venn diagram analysis. Next, using the STRING, Cytoscap, and Molecular Complex Detection (MCODE) plug-in, the related protein-protein interaction network among upregulated genes was analyzed. Moreover, Kaplan-Meier plots were used to analyze the relationship between the identified genes and OS prognosis. Genes significantly associated with worse prognosis were evaluated using the Gene Expression Profiling Interactive Analysis.

Results Thirteen genes were confirmed to be significantly more expressed in OS than in normal tissues. Five genes (*AURKB, EXO1, KIF4A, KIF15, and MCM4*) were found to influence OS prognosis.

Conclusion We identified five core genes related to the prognosis of OS and constructed a clinical prediction model for OS. Our data may provide a reference for future research on mechanisms, clinical diagnosis, and treatment of OS.

1 Introduction

Osteosarcoma (OS) is a highly hereditary and unstable cancer, and is the most common primary malignant bone tumor in children and adolescents, accounting for approximately 20% of all bone tumors [1]. The 5-year survival rate of patients with non-metastatic OS was only 60–70% through the combined treatment of surgery and chemotherapy [2]. Moreover, OS is prone to local invasion and early metastasis, mainly affecting the lung and skeleton. Indeed, local recurrence, metastasis, and chemotherapy resistance are associated to a poor prognosis [3, 4]. Despite significant improvements in treatment strategies in recent years, treatment outcomes for patients with metastatic or recurrent osteosarcoma are still pessimistic, with a rather grim prognosis of a greatly reduced 5-year overall survival rate of only 20–30% [5]. Current surgical treatment has a serious impact on limb function and chemotherapy can induce diverse side effects; therefore, there is an urgent need of new and enhanced treatment and early diagnosis strategies.

The rapid development of gene chip technology allowed its use in the diagnosis, treatment, and prognosis evaluation of diseases. Bioinformatics, as a new interdisciplinary subject in the field of life science, can deeply mine the microarray data of gene expression profiles, identifying new cues for exploring the underlying molecular pathological mechanisms [6, 7]. In particular, studies have shown that

OS is associated to altered expression of oncogenes and tumor suppressor genes and thereby influence the cell migration, angiogenesis, cell apoptosis, and proliferation.

In this study, bioinformatics methods were used to analyze publicly available OS microarray data and identify differentially expressed genes (DEGs) that could potentially contribute for the molecular mechanisms or as potential therapeutic targets of OS. Moreover, recognition of an OS-specific gene-phenotype correlation network could aid on a comprehensive analysis to screen the disorders associated with osteosarcoma typing.

2 Methods

2.1 Microarray data information

The microarray dataset GSE94805 was obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), which was based on the GPL13252 platform (Agilent-027114 Genotypic Technology designed Custom Human Whole Genome 8 × 60 k Microarray) and included 12 samples of OS U2OS cells (four control, four quiescence, and four senescence samples).

2.2 Data processing of co-expressed DEGs

The GEO2R online tool was used to identify the DEGs in control, quiescence, and senescence OS samples of the dataset, with |log fold change (FC)| > 2 and adjusted P-value < 0.05. The co-expressed DEGs between control vs. quiescence and control vs. senescence were selected through a Venn diagram.

2.3 Enrichment analysis of co-expressed DEGs

Gene ontology (GO) analysis was performed using the differential genes selected from the David website (https://david.ncifcrf.gov/). Molecular function, biological process, and cellular component were selected to observe the functional category or cellular location of the differential genes. Using the KOBAS 3.0 database, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the differential genes was carried out to observe their role in possible signaling pathways.

2.4 Protein-protein interaction (PPI) network and module analysis

The PPI network of the co-expressing DEGs was constructed using the retrieval tool STRING (https://string-db.org/). The MCODE plug-in of Cytoscape (version 3.7.1) was used to select the densely connected modules in the PPI network and the key genes were determined.

2.5 Survival analysis and RNA sequencing expression of core genes

Kaplan-Meier-plotter is a commonly used online tool for assessing the effect on survival of a great number of genes based on the European Genome-Phenome Archive, The Cancer Genome Atlas (TCGA), and GEO databases (Affymetrix microarrays only). The log rank P-value and hazard ratio (HR) with 95% confidence intervals were computed and are shown in the plots. To validate the identified DEGs, the Gene Expression Profiling Interactive Analysis (http://gepia.cancer-pku.cn) was performed using the data of

RNA sequencing expression based on thousands of samples from the Genotype-Tissue Expression and TCGA projects.

2.6 Evaluation of prognostic indicators

The prognostic indicators, including overall survival, progression-free interval (PFI), disease-free interval (DFI), and disease-specific survival (DSS) were evaluated using the Sangerbox software (http://www.sangerbox.com). Survival analysis was evaluated using the Kaplan-Meier method and log rank test (P < 0.05). Every median core gene expression was chosen as the cut-off value for the human cancer dichotomy, thus dividing respective patients into high- and low-risk groups.

2.7 Statistical analysis

Statistical analysis was performed using Graph Pad Prism 7 (GraphPad Software, La Jolla, CA, USA) and IBM SPSS Statistics for Windows version 23.0 (IBM Corp., Armonk, NY, USA). Data are expressed as mean ± standard deviation. The F test was used to analyze the homogeneity of variance among quantitative data groups, the Student's t-test and non-parametric test were used to compare the data between groups, and P-value < 0.05 was selected as the threshold of significance.

3 Results

3.1 Identification of DEGs

According to the GSE94805 dataset, 614 and 1,105 upregulated DEGs of normal vs. quiescent OS and normal vs. senescent OS, respectively, were identified. Overall, a total of 360 upregulated DEGs shared by the two DEG sets were selected for further analysis (Fig. 1).

3.2 Network and module analysis of co-expressed DEGs

The PPI network corresponding to the 360 intersecting DEGs was analyzed using Cytoscape (Fig. 1A). Additional analysis further revealed a total of 113 upregulated core genes among the network (Fig. 1B).

3.3 Analysis of core genes by the Kaplan-Meier plotter and GEPIA

Survival outcome related to the 113 core genes identified was evaluated by Kaplan-Meier plots. Sixteen genes had significantly worse survival, whereas 97 had no significant impact on survival (P < 0.05, Fig. 3A). GEPIA was then used to determine the expression profile of the 16 genes, revealed that 13 of these genes were highly expressed in the quiescent and senescent groups compared to the control group (P < 0.05, Fig. 3B).

3.4 Enrichment analysis of co-expressed DEGs

Based on the PPI network analysis, enrichment analysis of co-expressed DEGs was performed by GO term and KEGG pathway analysis. The GO results indicated that (1) for biological processes, 13 core genes were particularly enriched in mitotic sister chromatid segregation, G2/M transition of mitotic cell cycle, spindle checkpoint, mitotic spindle organization, regulation of cytokinesis, mitotic spindle assembly, and negative regulation of apoptotic process; (2) for cell component, 13 core genes were

significantly enriched in the chromosome passenger complex, spindle midzone, spindle microtubule, nucleoplasm, midbody, kinetochore, condensed nuclear chromosome, centromeric region, spindle pole centrosome, chromocenter, chromosome, and centromeric region; (3) for molecular function, 13 core genes were enriched in the ATP binding, histone serine kinase activity, and protein serine/threonine/tyrosine kinase activity (Table 1).

Table 1
Gene ontology analysis of 13 key candidate genes in osteosarcoma

Category	Term	Count	%	*p*-value	FDR
GOTERM_BP_DIRECT	GO:0000070 ~ mitotic sister chromatid segregation	3	7.8	3.396E-4	0.037
GOTERM_BP_DIRECT	GO:0000086 ~ G2/M transition of mitotic cell cycle	3	7.8	7.163E-4	0.039
GOTERM_BP_DIRECT	GO:0031577 ~ spindle checkpoint	2	5.2	0.005	0.166
GOTERM_BP_DIRECT	GO:0007052 ~ mitotic spindle organization	2	5.2	0.033	0.603
GOTERM_BP_DIRECT	GO:0032465 ~ regulation of cytokinesis	2	5.2	0.033	0.604
GOTERM_BP_DIRECT	GO:0090307 ~ mitotic spindle assembly	2	5.2	0.038	0.604
GOTERM_BP_DIRECT	GO:0043066 ~ negative regulation of apoptotic process	3	7.8	0.039	0.604
GOTERM_CC_DIRECT	GO:0032133 ~ chromosome passenger complex	4	10.5	2.152E-8	9.90E-7
GOTERM_CC_DIRECT	GO:0051233 ~ spindle midzone	4	10.5	1.194E-6	2.75E-5
GOTERM_CC_DIRECT	GO:0005876 ~ spindle microtubule	4	10.5	1.043E-5	1.60E-4
GOTERM_CC_DIRECT	GO:0005654 ~ nucleoplasm	9	23.5	1.364E-4	0.002
GOTERM_CC_DIRECT	GO:0030496 ~ midbody	4	10.5	1.658E-4	0.002
GOTERM_CC_DIRECT	GO:0000776 ~ kinetochore	3	7.8	0.002	0.019
GOTERM_CC_DIRECT	GO:0000780 ~ condensed nuclear chromosome, centromeric region	2	5.2	0.011	0.070
GOTERM_CC_DIRECT	GO:0031616 ~ spindle pole centrosome	2	5.2	0.012	0.070
GOTERM_CC_DIRECT	GO:0010369 ~ chromocenter	2	5.2	0.015	0.076
GOTERM_CC_DIRECT	GO:0000775 ~ chromosome, centromeric region	2	5.2	0.024	0.112
GOTERM_MF_DIRECT	GO:0005524 ~ ATP binding	8	20.9	2.816E-4	0.011
GOTERM_MF_DIRECT	GO:0035174 ~ histone serine kinase activity	2	5.2	0.006	0.126

Category	Term	Count	%	*p*-value	FDR
GOTERM_MF_DIRECT	GO:0004712 ~ protein serine/threonine/tyrosine kinase activity	2	5.2	0.019	0.249

The results from KEGG analysis showed that the 13 upregulated core genes were enriched in multiple pathways (Table 2 and Fig. 4A), including cell cycle, progesterone-mediated oocyte maturation, mismatch repair, cellular senescence, DNA replication, viral carcinogenesis, oocyte meiosis, and FxO signaling pathways. According to the biological pathway for key genes, 13 upregulated key genes were enriched in PLK1 signaling events, mitotic prometaphase, aurora kinases, FOXM1 transcription factor network, cell cycle, and aurora B signaling (Fig. 4B)

Table 2
KEGG pathway analysis of 13 key candidate genes in osteosarcoma

Pathway ID	Name	Count	%	FDR	Genes
hsa04110	Cell cycle	5	4.1	8.09e-06	RBL1,SKP2,CCNA2,MCM4,PLK1
hsa04914	Progesterone-mediated oocyte maturation	3	3.2	0.0023	CCNA2,PLK1,AURKA
hsa03430	Mismatch repair	2	8.7	0.0031	RFC5,EXO1
hsa04218	Cellular senescence	3	1.9	0.0049	RBL1,CCNA2,MYBL2
hsa03030	DNA replication	2	5.6	0.0049	RFC5,MCM4
hsa05203	Viral carcinogenesis	3	1.6	0.0052	SKP2,CCNA2,RBL1
hsa04114	Oocyte meiosis	2	1.7	0.0287	AURKA,PLK1
hsa04068	FoxO signaling pathway	2	1.5	0.0312	SKP2,PLK1

3.5 Five core genes reduce the prognosis of OS

To evaluate the prognostic value of the identified five core genes in OS, the survival outcomes were assessed. As shown in Fig. 5, high expression of the five core genes (*AURKB*, *EXO1*, *KIF4A*, *KIF15*, and *MCM4*) had worse OS prognosis.

4 Discussion

Osteosarcoma is the most common primary bone tumor in children and adolescents, and patients with advanced osteosarcoma with signs of metastasis have a poor prognosis [8]. In order to identify new

valuable prognostic biomarkers for OS, bioinformatics methods based on the GSE94805 dataset were used in this study.

Herein, the gene expression profiles of U2OS cells in the quiescence and senescence phases were analyzed, revealing 360 overlapping upregulated DEGs. However, these genes preliminarily screened were not necessarily related to the occurrence and development of OS. To further refine the study targets, String, Cytoscap, and MCODE plug-ins were used to conduct enrichment analysis, narrowing the investigation list to 113 key candidate genes. Next, 13 significantly upregulated genes were selected by comparing cancer and para-cancerous samples from OS patients. Furthermore, analysis with the Sangbox software showed that the expression of five closely related core genes – *AURKB*, *EXO1*, *KIF4A*, *KIF15*, and *MCM4* – had significant impact on several survival and disease-progression prognostic indicators.

Aurora kinases (AURKs) are highly conserved serine/threonine kinases, including AURKA, AURKB, and AURKC, which are key regulators involved in mitosis [9]. Studies have shown that AURKB is highly expressed in a variety of human cancers, such as lung [10], gastric [11] and liver [12] cancer, and is associated with tumor invasion and metastasis, and poor patient prognosis [13, 14]. Overexpression of AURKB can lead to errors during mitosis, such as damage to the junction between kinetochores and microtubules, interfering with chromosome concentration and spindle assembly, and hindering centromere localization, which can cause abnormal chromosome segregation and cytokinesis, ultimately leading to malignant cell transformation [15]. AURKB inhibitors have no effect on non-proliferative cells and have little effect on normal cells [16]. Therefore, AURKB holds potential as a clinical therapeutic target for tumors.

Exonuclease 1 (EXO1), a member of the Rad2/XPG nuclease family, plays important roles in the regulation of cell cycle checkpoint, replication fork maintenance, and DNA repair after replication [17]. Previous studies revealed that EXO1 is associated with various cancers, including Lynch syndrome, ovarian, gastric, lung, breast, and pancreatic cancers [18]. Overexpression of EXO1 disturbs the relative stability of the genome. Dai et al. [19] reported that EXO1 knockdown can reduce the proliferation of hepatocellular carcinoma cells *in vitro*, inhibit the survival of cancer clone cells, and is associated with poor prognosis of patients with hepatocellular carcinoma. To date, studies on EXO1 and OS are still lacking, and the role of EXO1 in OS is unclear. The present study suggests that high expression of EXO1 in OS is closely related to disease prognosis, and that the upregulation of EXO1 in OS might predict a poor clinical outcome.

The kinesin superfamily (KIFs) is a group of proteins that share highly conserved motor regions, including 1−14 subtypes. Most KIFs have ATP-dependent activity and catalyze microtubule-dependent tailing reactions. KIFs play important roles in the occurrence and development of tumors. KIF4A, a member of the kinesin-4 family, is located in the nucleus during the interphase, which is mainly involved in the biological processes of chromatin separation, spindle formation, and cytoplasmic separation during mitosis. KIF4A is abnormally expressed in lung and ovarian cancers, and oral squamous cell carcinoma

[20]. KIF15, belonging to the kinesin-12 family, is an N-rectified directional motor that plays a crucial role in the formation of bipolar spindles, and maintains the overall integrity of the spindle by shaping and stabilizing the kineto-fiber [21, 22]. Studies have shown that these proteins participate in several cellular processes, including proliferation, apoptosis, differentiation, and development [23].

MCM4 is a member of the minichromosome maintenance protein (MCM) with a highly conserved sequence. Commonly, MCM4 is active in the form of an MCM4/MCM6/MCM7 spherical trimer and MCM2−7cyclic hexamer [24]. MCM2−7 proteins are present in proliferating cells. Cancers arising in different anatomic sites are also associated with MCM2, MCM4, and MCM6 overexpression [25, 26]. The MCM4 subunit is a target site for replication testing and is involved in DNA replication, assembly of related proteins, and regulation of the MCM4/MCM6/MCM7 helicase activity [27].

5 Conclusion

Taken together, by combining gene expression data and OS clinical information, this study identified five core genes related to the poor prognosis of OS. These data may provide a reference for future research on mechanisms, clinical diagnosis, and treatment of OS.

Declarations

Authors' contributions

PH and YL designed the study. YL and XZ performed the research. YL, BL and QX analyzed the data. YL and XZ wrote the main manuscript text. GC and PH revised the manuscript. All authors reviewed and approved the final manuscript.

Data availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests The authors declare that they have no competing interests.

References

1. Namlos HM, Kresse SH, Muller CR, et al. Global gene expression profiling of human osteosarcomas reveals metastasis-associated chemokine pattern. Sarcoma. 2012;2012:639038.
2. Bacci G, Ferrari S, Bertoni F, et al. Long-term outcome for patients with nonmetastatic osteosarcoma of the extremity treated at the istituto ortopedico rizzoli according to the istituto ortopedico rizzoli/osteosarcoma-2 protocol: an updated report. J Clin Oncol. 2000;18(24):4016-4027.
3. Simpson S, Dunning MD, de Brot S, Grau-Roma L, Mongan NP, Rutland CS. Comparative review of human and canine osteosarcoma: morphology, epidemiology, prognosis, treatment and genetics. Acta Vet Scand. 2017;59(1):71.

4. Zhang Y, Yang J, Zhao N, et al. Progress in the chemotherapeutic treatment of osteosarcoma. Oncol Lett. 2018;16(5):6228-6237.

5. Mialou V, Philip T, Kalifa C, et al. Metastatic osteosarcoma at diagnosis: prognostic factors and long-term outcome--the French pediatric experience. Cancer. 2005;104(5):1100-1109.

6. Yang X, Zhu S, Li L, et al. Identification of differentially expressed genes and signaling pathways in ovarian cancer by integrated bioinformatics analysis. Onco Targets Ther. 2018;11:1457-1474.

7. Liu M, Xu Z, Du Z, et al. The Identification of Key Genes and Pathways in Glioma by Bioinformatics Analysis. J Immunol Res. 2017;2017:1278081.

8. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991-995.

9. Huang D, Huang Y, Huang Z, Weng J, Zhang S, Gu W. Relation of AURKB over-expression to low survival rate in BCRA and reversine-modulated aurora B kinase in breast cancer cell lines. Cancer Cell Int. 2019;19:166.

10. Al-Khafaji AS, Davies MP, Risk JM, et al. Aurora B expression modulates paclitaxel response in non-small cell lung cancer. Br J Cancer. 2017;116(5):592-599.

11. Enjoji M, Iida S, Sugita H, et al. BubR1 and AURKB overexpression are associated with a favorable prognosis in gastric cancer. Mol Med Rep. 2009;2(4):589-596.

12. Lin ZZ, Jeng YM, Hu FC, et al. Significance of Aurora B overexpression in hepatocellular carcinoma. Aurora B Overexpression in HCC. BMC Cancer. 2010;10:461.

13. Zhang Y, Jiang C, Li H, et al. Elevated Aurora B expression contributes to chemoresistance and poor prognosis in breast cancer. Int J Clin Exp Pathol. 2015;8(1):751-757.

14. Chen YJ, Chen CM, Twu NF, et al. Overexpression of Aurora B is associated with poor prognosis in epithelial ovarian cancer patients. Virchows Arch. 2009;455(5):431-440.

15. Santaguida S, Musacchio A. The life and miracles of kinetochores. EMBO J. 2009;28(17):2511-2531.

16. Xie F, Zhu H, Zhang H, et al. In vitro and in vivo characterization of a benzofuran derivative, a potential anticancer agent, as a novel Aurora B kinase inhibitor. Eur J Med Chem. 2015;89:310-319.

17. Cotta-Ramusino C, Fachinetti D, Lucca C, et al. Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. Mol Cell. 2005;17(1):153-159.

18. Keijzers G, Bakula D, Petr MA, et al. Human Exonuclease 1 (EXO1) Regulatory Functions in DNA Replication with Putative Roles in Cancer. Int J Mol Sci. 2018;20(1).

19. Dai Y, Tang Z, Yang Z, et al. EXO1 overexpression is associated with poor prognosis of hepatocellular carcinoma patients. Cell Cycle. 2018;17(19-20):2386-2397.

20. Sheng L, Hao SL, Yang WX, Sun Y. The multiple functions of kinesin-4 family motor protein KIF4 and its clinical potential. Gene. 2018;678:90-99.

21. Tanenbaum ME, Macurek L, Janssen A, Geers EF, Alvarez-Fernandez M, Medema RH. Kif15 cooperates with eg5 to promote bipolar spindle assembly. Curr Biol. 2009;19(20):1703-1711.

22. Drechsler H, McAinsh AD. Kinesin-12 motors cooperate to suppress microtubule catastrophes and drive the formation of parallel microtubule bundles. Proc Natl Acad Sci U S A. 2016;113(12):E1635-1644.

23. Zou JX, Duan Z, Wang J, et al. Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. Mol Cancer Res. 2014;12(4):539-549.

24. Cannone G, Visentin S, Palud A, Henneke G, Spagnolo L. Structure of an octameric form of the minichromosome maintenance protein from the archaeon Pyrococcus abyssi. Sci Rep. 2017;7:42019.

25. Das M, Prasad SB, Yadav SS, et al. Over expression of minichromosome maintenance genes is clinically correlated to cervical carcinogenesis. PLoS One. 2013;8(7):e69607.

26. Kwok HF, Zhang SD, McCrudden CM, et al. Prognostic significance of minichromosome maintenance proteins in breast cancer. Am J Cancer Res. 2015;5(1):52-71.

27. Sheu YJ, Kinney JB, Stillman B. Concerted activities of Mcm4, Sld3, and Dbf4 in control of origin activation and DNA replication fork progression. Genome Res. 2016;26(3):315-330.
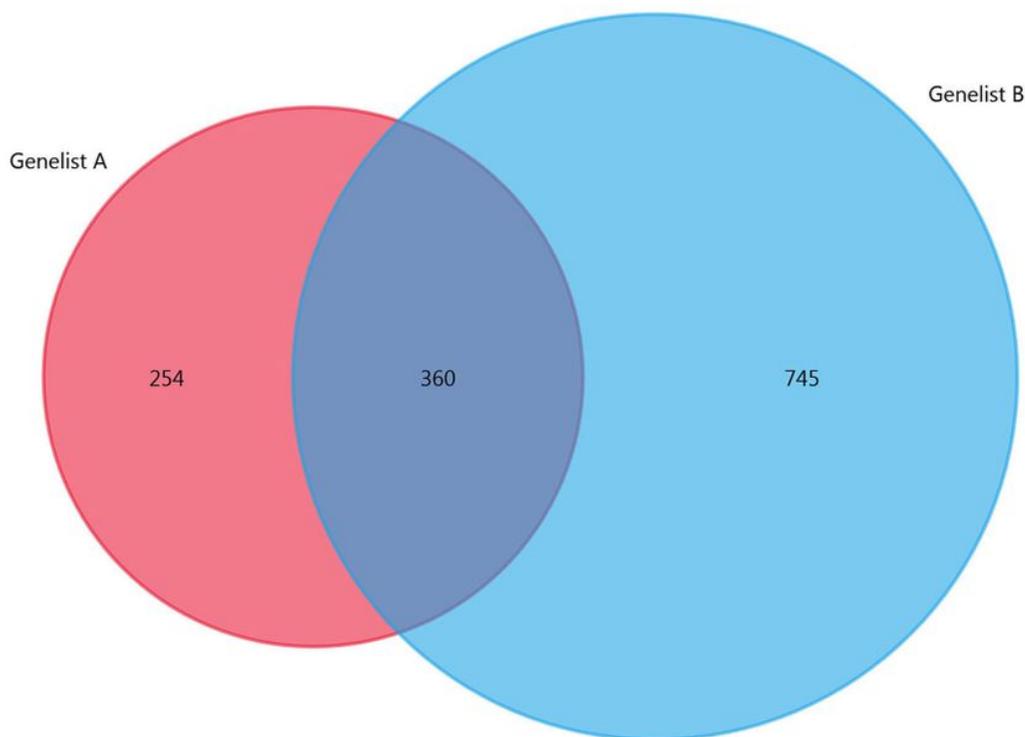
Figures

Figure 1

Venn diagram of differential genes. Genelist A, indicates upregulated DEGs of normal vs quiescent OS, Genelist B, indicates upregulated DEGs of normal vs senescent OS.
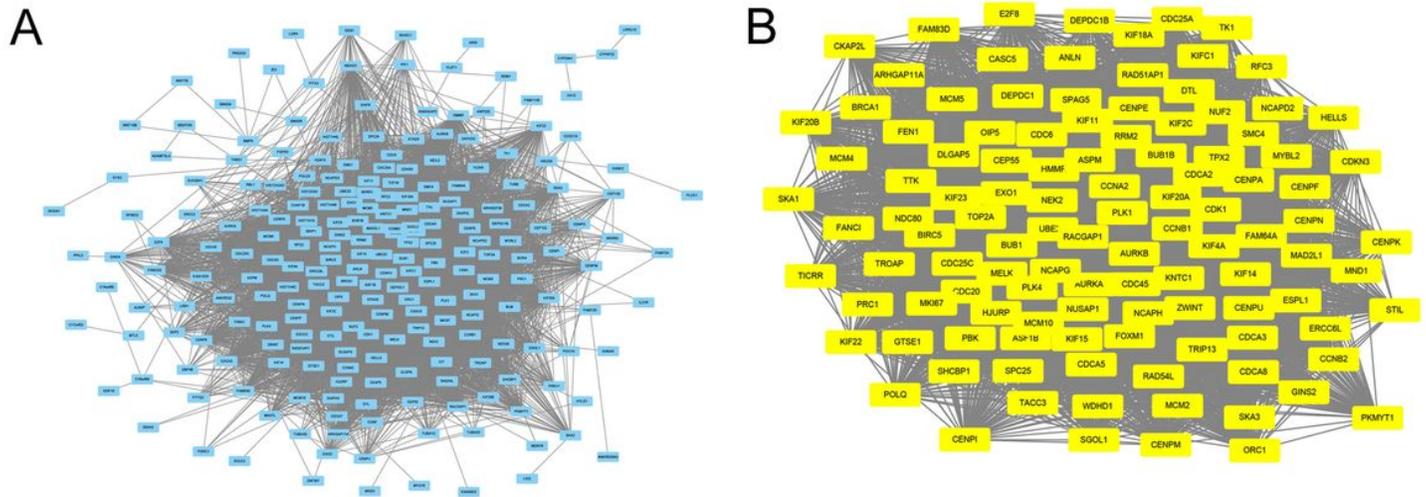

Figure 2

Screening of key candidate genes. (A) 360 upregulated DEGs in the PPI network complex. The nodes represent proteins; the edges represent the interaction. (B) Module analysis via Cytoscape software.
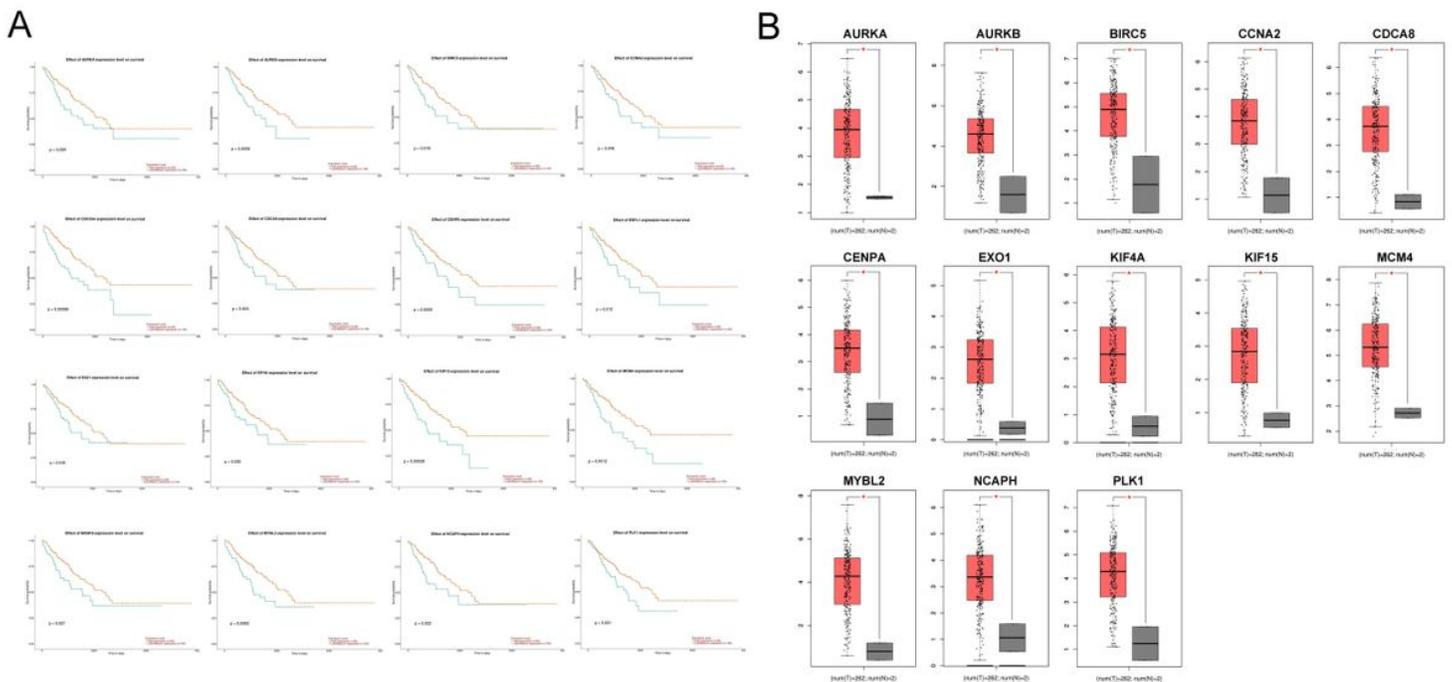

Figure 3

The prognosis analysis of key genes. (A) The prognostic information of the 113 key candidate genes, among which 16 genes had a significantly worse survival rate (P < 0.05). (B) Significantly expressed genes related with poor prognosis were analyzed between OS and normal samples using the GEPIA

software. Thirteen genes had significant expression in OS cancer samples compared to normal samples (*P < 0.05). Red and grey colors represent tumor and normal tissues, respectively.
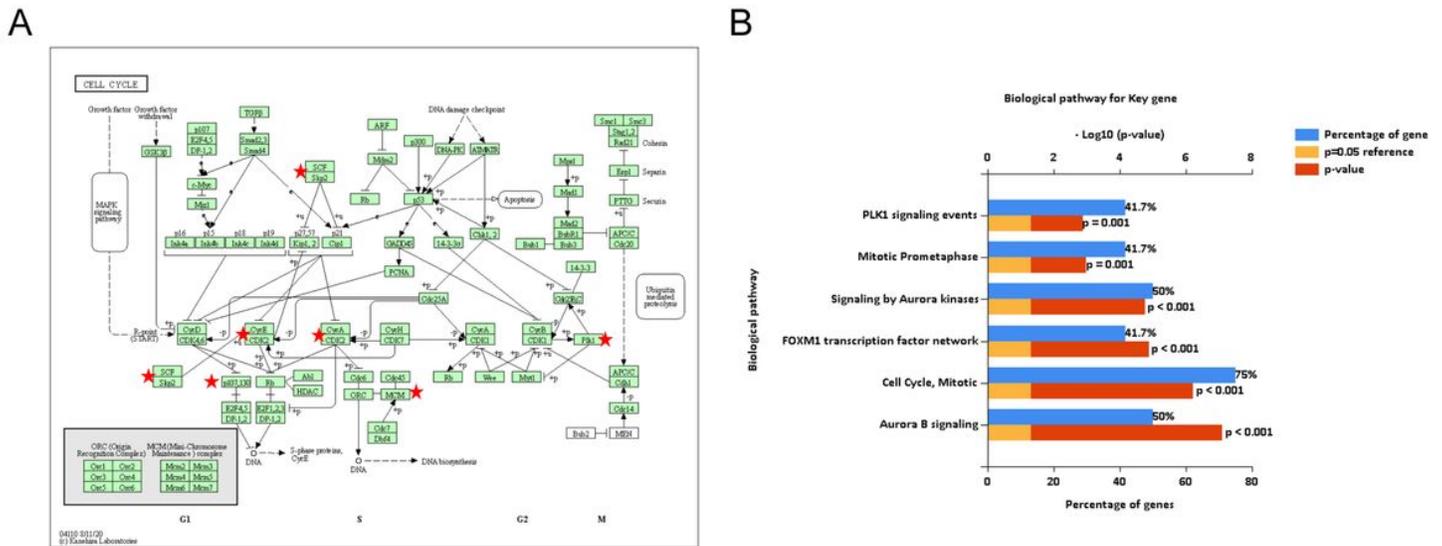

Figure 4

Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the 13 key genes. (A) Analysis of 13 selected genes by KEGG pathway enrichment. Seven genes were significantly enriched in the cell cycle. (B) Biological pathways of the 13 key genes evaluated by KEGG enrichment.
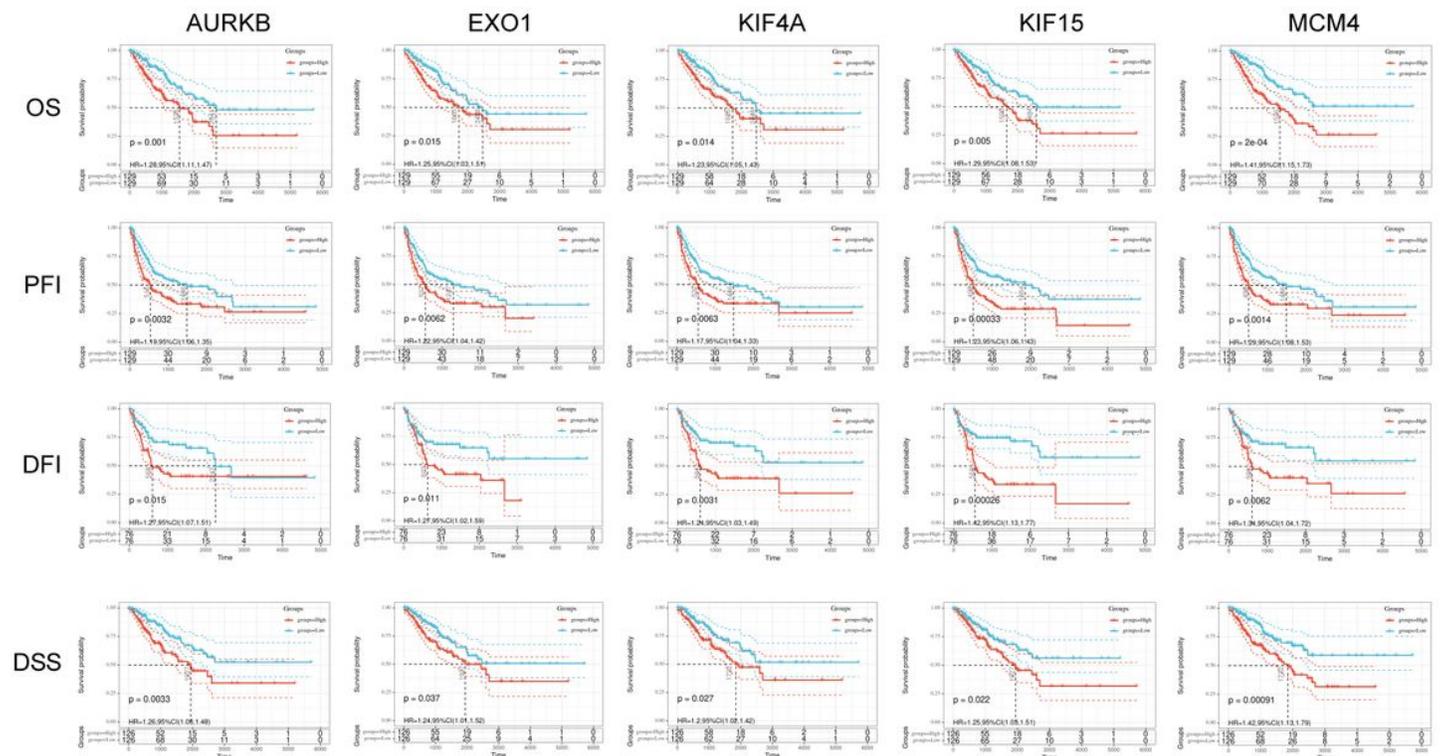

Figure 5

Kaplan-Meier survival curves according to the expression of the five core genes in OS.