

Bioinformatics Analysis Identified Five Widely Expressed Genes Associated with Prognosis in Sarcoma

Bizhi Tu^{1,*}, Yaya Jia^{2,*}, Jun Qian¹

¹Department of Orthopedics, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, People's Republic of China; ²Department of Pediatrics, The Shanxi Medical University, Taiyuan, Shanxi, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jun Qian, Department of Orthopedics, The First Affiliated Hospital of Anhui Medical University, 218 Jixi Road, Hefei, 230022, Anhui, People's Republic of China, Tel +86-13955116514, Fax +86-551-62922025, Email qjpaper@sina.cn

Background: Human sarcomas (SARC) are a group of malignant tumors that originated from mesenchymal lineages with more than 60 subtypes. However, potential biomarkers for the diagnosis and prognosis of SARC remain to be investigated.

Methods: We obtained three GSE raw matrix files (GSE39262, GSE21122, GSE48418) that related to various subtypes of sarcoma from the public GEO database and explored the widely differential expression genes in three obtained GSE files. Then common differential expression genes (CDGEs) were identified. We analyzed the correlation between the expression of the top five interacted genes of CDGEs and genome-wide differences, prognosis, genetic mutation, functional enrichment, immune infiltration, immune checkpoint, and marker genes' expression of N⁶-methyladenosine (m⁶A) modification in SARC patients. Besides, a prognostic nomogram was constructed to predict the survival of SARC patients.

Results: Among the three GSE files, 42 CDGEs were identified, and the top five interacted genes were ASPM, CCNB2, PRC1, AURKA, and SCM2. The expression levels of the five genes were higher in the SARC group than that in the normal group. The transcriptional level of CCNB2, PRC, and SCM2 was correlated to the worse survival of SARC. The constructed nomogram that combined CNB2, PRC1, and SCM2 showed a fairly good credibility in predicting the survival of SARC (C-index: 0.711). Furthermore, the five genes were widely involved in immune infiltration, immune checkpoint, and m⁶A modification. In addition, we found a minor survival-related mutation rate (9%) of the five identified genes in SARC patients ($p < 0.05$).

Conclusion: The results suggested the five identified genes widely participated in the prognosis, immune infiltration, immune checkpoint, and m⁶A modification of SARC patients. This study provided a theoretical basis for the research about the correlation between the level of five identified genes and sarcoma, but the further mechanism needs to be verified by experiments.

Keywords: sarcoma, bioinformatics, prognostic values, differential expression gene, *ASPM*, *CCNB2*, *PRC1*, *AURKA*, *SCM2*

Introduction

Human sarcomas (SARC) encompass more than sixty subtypes, which originate from mesenchymal lineages with the multifactorial somatic genetic alteration. Newly confirmed cases of SARC (13,130 people) accounted for 1% of all malignant tumors and more than 5350 deaths in 2020 around America.¹ However, the five-year survival rate for advanced SARC patients is less than 20–30%.² Surgical resection is the primary strategy for the early treatment of sarcoma. For advanced-stage SARC patients, radiotherapy and combination chemotherapy are applied in some sensitive types of sarcomas preoperatively or postoperatively, which has acquired an initial ideal effect.³ However, approximately 50% of patients with SARC tend to recurrence and metastases after early radiotherapy, chemotherapy, and surgery treatment.⁴ The current chemotherapy regimen used in advanced sarcoma patients is usually difficult to achieve substantial

improvement, the expected survival time ranges from 12 to 18 months.⁴⁻⁶ Therefore, identifying the potential diagnostic and prognostic biomarkers of SARC is of great importance.

Different subtypes of sarcoma have various clinical processes and outcomes.⁷ For example, the biological characteristic of desmoid tumors tends to local infiltration and causes various clinical symptoms, such as chronic pain, psychological diseases, and function dysfunction,⁸ while osteosarcoma has a propensity for bone destruction and early distant metastasis.⁹ The features common to biological characteristics and molecular markers among different subtypes of sarcoma are less known.¹⁰ Recently, the use of high-throughput gene sequencing provides a better understanding of SARC at the genomic heterogeneity level.^{11,12} Utilizing gene sequence analysis, several overexpressed genes related to spindle assembly and centrosome activity had been identified.¹³ Similarly, minichromosome maintenance family proteins were found to widely participate in the tumor progress and prognosis in patients with osteosarcoma.¹⁴ However, the role of common differential expression genes (CDEGs) among diverse subtypes of SARC has not been reported. In the current study, we obtained several CDEGs from three GEO series (GSE) profiles and identified several genes that interacted the most. A better understanding of the multimolecular characteristics of these genes in SARC is expected to open novel perspectives for identifying the potential biomarkers for tumor diagnosis and prognosis and would provide guidance for future studies.

Methods

Identification of CDEGs from the GEO Database

We downloaded three sarcoma-related GSE gene expression matrix files (GSE39262, GSE21122, GSE48418) from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), and 10 subtypes of SARC were included: osteosarcoma, chondrosarcoma, rhabdomyosarcoma, fibrosarcoma, neuroblastoma, liposarcoma, Ewing sarcoma, leiomyosarcoma, synovial sarcoma, and myxofibrosarcoma. The GSE profile mentioned above was classified into sarcoma group, and control group respectively. Utilizing the limma package in R software, the differential expression of various genes in each GSE profile were analyzed. The ggplot2 and ComplexHeatmap packages were used for visualization. Genes that were differentially expressed ($|\log_{2}FC| > 1$ and $p < 0.05$) in all three databases were defined as CDEGs.

GO/KEGG Enrichment Analysis of CDEGs

Using the clusterProfile package in R software, we conducted gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis on CDEGs. And the main enrichment function and pathway of the CDEGs were retained when an adjusted p-value was less than 0.05.

Protein-Protein Interaction (PPI) Analysis of CDEGs

The PPI network of CDEGs was performed using the STRING database (<http://string-db.org>), an online network designed to explore the potential protein interaction network among genes by collecting, integrating, and scoring publicly available data. Cytoscape was used for the visualization of PPI.¹⁵ A plug-in model Molecular Complex Detection (MCODE) of Cytoscape was applied to identify the tightly connected genes. The top five genes of CDEGs with the highest connect degree were retained for further analysis.

Differential Expression and Prognostic Analysis of the Top Five Genes of CDEGs

We utilized the GEPIA (<http://GEPIA.cancer-pku.cn/index.htm>) database, a new online analysis tool that provides the RNA-seq expression data of more than 9736 tumors and 8587 normal tissue samples, to explore the association of the expression of five genes between sarcomas and normal tissues.¹⁶ Then, we analyzed the relationship between the expression of five identified genes and prognosis among SARC patients. Overall survival (OS) and disease free survival (DFS) of SARC were analyzed in the Survival Analysis module of the GEPIA database. Moreover, progress free interval (PFI) and OS were explored in the Cancer Genome Atlas Database (TCGA, <https://portal.gdc.cancer.gov/>), a comprehensive tumor database that contains expression information and survival data of 34 tumor types.¹⁷

Nomogram Model in Predicting the Prognosis of SARC

The included variables were identified in the univariate analysis module, which contains Age, Gender, Female, Male, Tumor depth, Superficial, Deep, Tumor multifocal, Radiation therapy, Tumor necrosis, Metastasis, Margin status, Race, Residual tumor (Table S1). Three of the five identified genes were also included due to the observed significant correlation in the prognosis of sarcoma including CCNB2, PRC1, and SMC2 ($p < 0.05$ for all in OS). Based on the results of the univariate analysis and relevant clinical parameters, a predictive nomogram was generated utilizing the rms and survival package in R software. To test the clinical applicability of the nomogram, calibration plots, decision curve analysis (DCA), and time-dependent ROC analysis were applied. Moreover, a risk score analysis was also performed based on the Cox regression results.

Correlation Analysis Between the Expression of Five Identified Genes and Immune Cell Infiltration

The relationship between the level of five identified genes and immune cell infiltration was investigated in the TIMER database (<https://cistrome.shinyapps.io/timer/>), an online database that provides the clinical relevance of tumor immune subsets and corrects the output results in a multivariable Cox proportional hazard model.¹⁸ In the Survival module of the TIMER database, we analyzed the relationship between the expression levels of five identified genes and the infiltration of six immune cells (neutrophils, B cells, CD8⁺ T cells, CD4⁺ T cells, dendritic cells, and macrophages).

Genetic Mutation Analysis of Five Identified Genes

Mutations of the five identified genes in tumor tissues were analyzed using the cBioPortal (<http://www.cbioportal.org>) database.¹⁹ The clinic data from 640 SARC samples containing gene copy-number alterations and prognostic information was obtained from the TCGA database. Using the cBioPortal, we also explored the correlation between mutations of the five identified genes and prognostic indicators of OS and disease-specific survival (DSS) in patients with SARC. The z-score threshold of mRNA expression (RNASeq V2 RSEM) was set as ± 1.8 .

Immune Checkpoint and N⁶-Methyladenosine (m⁶A) Modification Analysis

A standardized universal SARC dataset was downloaded in the UCSC database (<https://xenabrowser.net/>).²⁰ Correlation between 60 checkpoint inhibitor genes (24 inhibitory genes and 36 stimulatory genes) or 21 marker genes of m⁶A modification and expression of five identified genes was analyzed using the Sangerbox database (<http://sangerbox.com/>).

Results

Identification of CDEGs Between the SARC and Normal Groups

The differential expression genes of SARC and normal tissues were obtained from three GSE data files. As shown in Figure 1C, principal component analysis (PCA) manifested that the transcriptional levels of genes between SARC and normal tissues were significantly separated in GSE48418, and coincident in GSE39262 (Figure 1A) and GSE21122 (Figure 1B). In Figure 1D–F, the volcano plot presented the distribution of the differential genes in three GSE files. The red point represented upregulation genes based on the criteria of $p\text{-value} < 0.05$ and $\log\text{FC} > 1$, whereas blue points represented the downregulated genes in three GSE files ($p\text{-value} < 0.05$ and $\log\text{FC} < -1$). The clustering heatmap indicated that the differential expression genes in SARC were distinct from normal samples (Figure 2A–C). As shown in the Venn diagram, 42 CDEGs genes in the three GSE files were identified (Figure 2D).

Functional Enrichment and PPI Analysis of CDEGs in SARC

As shown in Figure 3A, the CDEGs significantly participated in the “response to acid chemical” and “collagen-containing extracellular matrix” pathways. Figure 3B shows the visualized network of the functional enrichment. Details were presented in Table S1. With respect to the PPI analysis, the top 5 interacted genes of CDEGs (ASPM, CCNB2, PRC1, AURKA, SMC2) were identified according to the degree of tight connection in Cytoscape (Figure 4A and B).

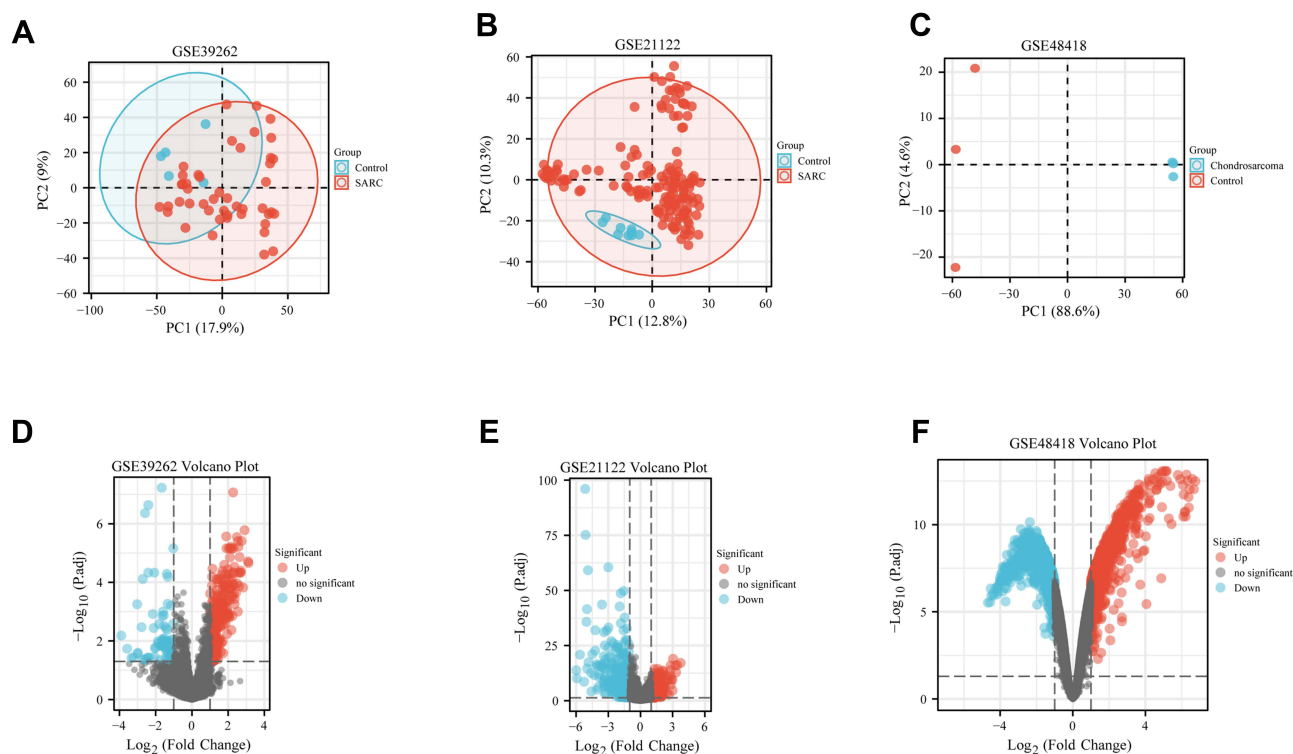


Figure 1 PCA plots of GSE files and Volcano maps of differential expression genes. (A–C) represents the PCA analysis plot of GSE39262, GSE 21122, and GSE48418. Red and blue spots represent samples from SARC and Normal (Control) groups, respectively. (D–F) represents the Volcano plot of GSE39262, GSE21122, and GSE48418. Red and green spots represent differentially expressed genes, red represent up-regulated genes and green represent down-regulated genes.

Aberrant Expressions of the Five Identified Genes and Prognostic Values in SARC

Figure 5 presented the differential expression of five identified genes between tumor and normal tissues, and we found the five genes were significantly amplified in the tumor samples ($p < 0.05$ for all). Moreover, we found the overexpression of ASPM was associated with poor DFS and PFI (Figure 6B and L), but not for OS (Figure 6A and K). A higher level of CCNB2 was related to poor OS (Figure 6C and M) of SARC. While no relationship was observed between CCNB2 expression and DFS, PFI (Figure 6D and N). As shown in Figure 6E and O, a higher level of PRC1 was associated with poor OS of SARC. Meanwhile, we found that PRC1 expression was significantly correlated to DFS and PFI (Figure 6F and P). Notably, we found AURKA was only statistically correlated with the DFS of SARC (Figure 6G–H and Q–R). Furthermore, SMC2 expression was also associated with the worse OS and PFI in SARC patients (Figure 6I–J, Figure 6S–T). Based on the univariate analysis, we found that CCNB2, PRC1, SMC2, Residual tumor, Metastasis, Margin status, and Tumor multifocal were associated with the clinical outcomes of SARC (Table S2). Combining these parameters, we constructed a prognostic nomogram, and the C-index was 0.711 (Figure 7A). The predictive ability and clinical capacity of the nomogram were corroborated by the calibration curve and DCA (Figure 7B and C). By time-dependent ROC analysis (Figure 7D), we found that the constructed nomogram has great applicability to predict 3-year and 5-year survival in SARC, with an AUC of 0.749 and 0.729. Based on the variables of the generated nomogram, the risk score for each patient was calculated and presented in Figure 7E. According to the median value of the risk score, the patients can be divided into low and high-risk groups. On the basis of the distribution of survival times, we can tell that the higher the risk score, the poorer survival outcomes in SARC patients. To investigate whether the heterogeneity of the sarcoma subtypes influences the results of our study, we obtained six sarcoma subtypes from TCGA database including dedifferentiated liposarcoma (58 cases), leiomyosarcoma (106 cases), myxofibrosarcoma (25 cases), pleomorphic sarcoma (52 cases), synovial sarcoma (10 cases), and malignant peripheral nerve sheath tumors (10 cases). We compared the expression level of the five target genes among six subtypes. As shown in Figure S1, overall expression differences of the target genes in six subtypes of SARC were relatively small.

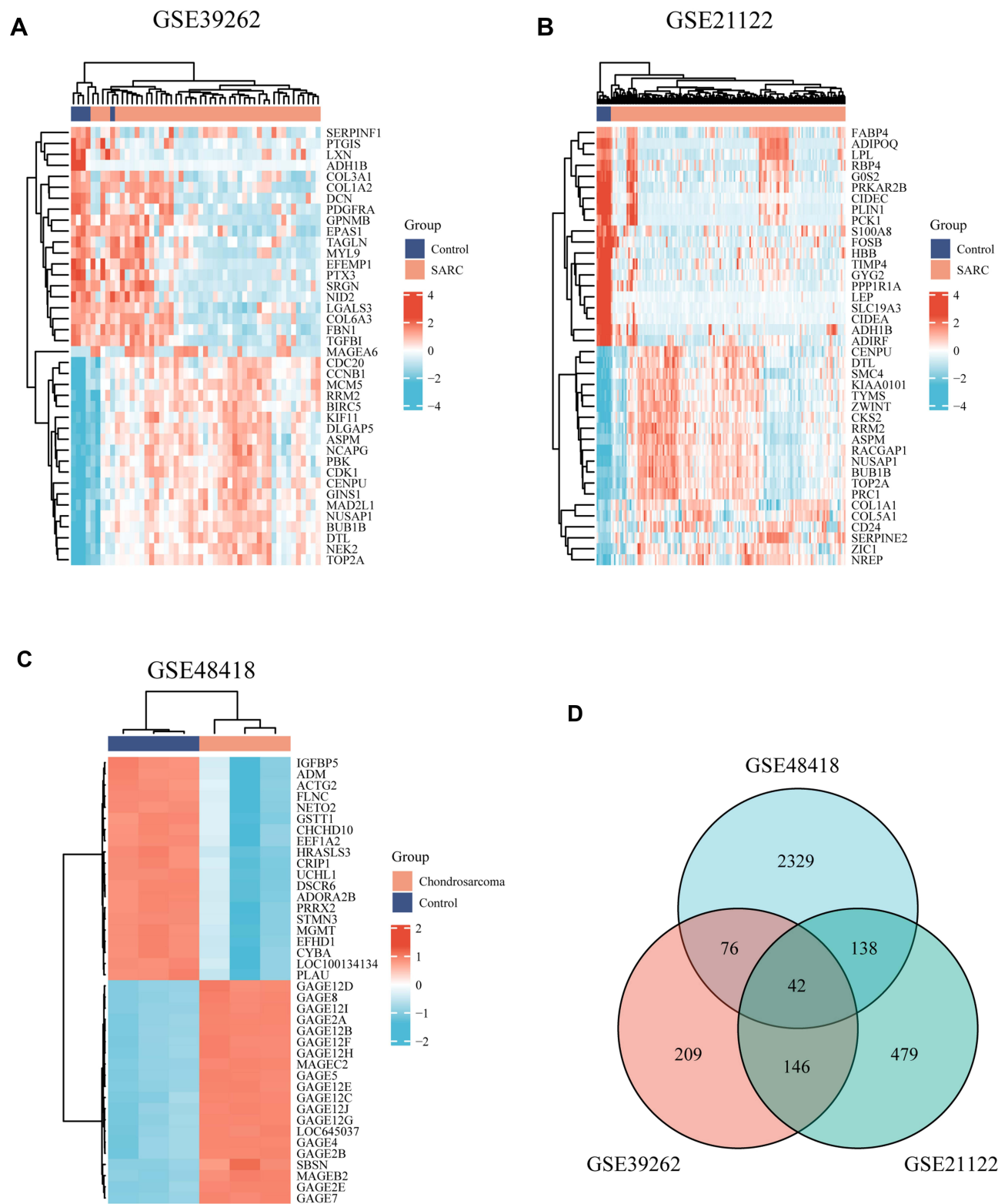


Figure 2 Heatmaps and Venn diagram of differential expression genes. Heatmap of differential expression genes in GSE39262 (A), GSE21122 (B), and GSE48418 (C). Venn diagram of differential expression genes in GSE39262, GSE21122, and GSE48418 (D).

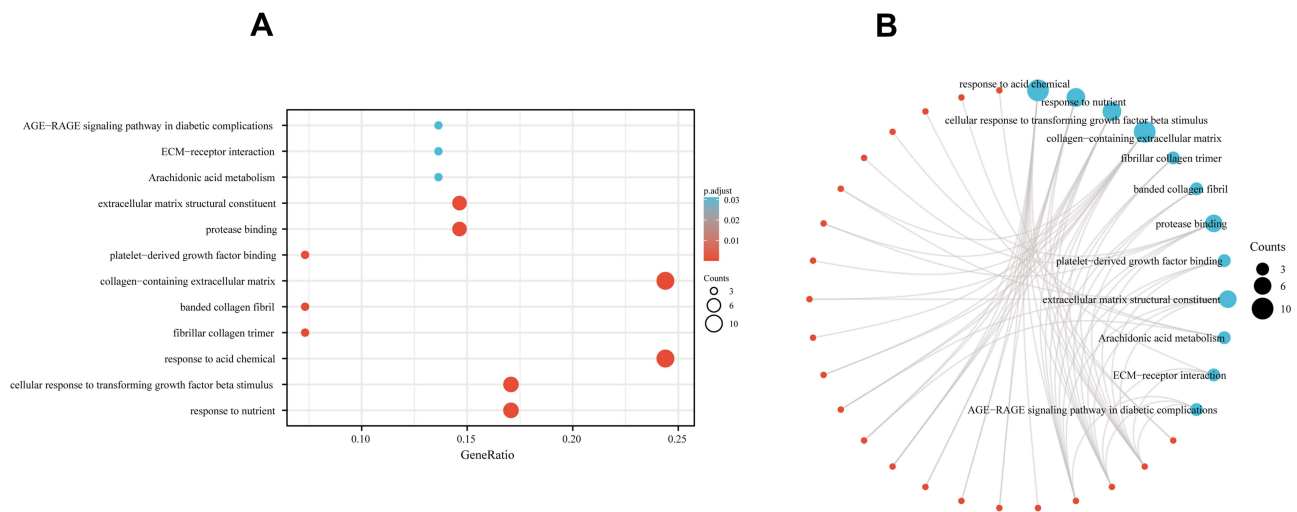


Figure 3 Enrichment analysis of the CDEGs in SARC (Metascape). Go and KEGG enrichment analysis predicted the functional roles of the target host gene (A) and provided a visual network (B).

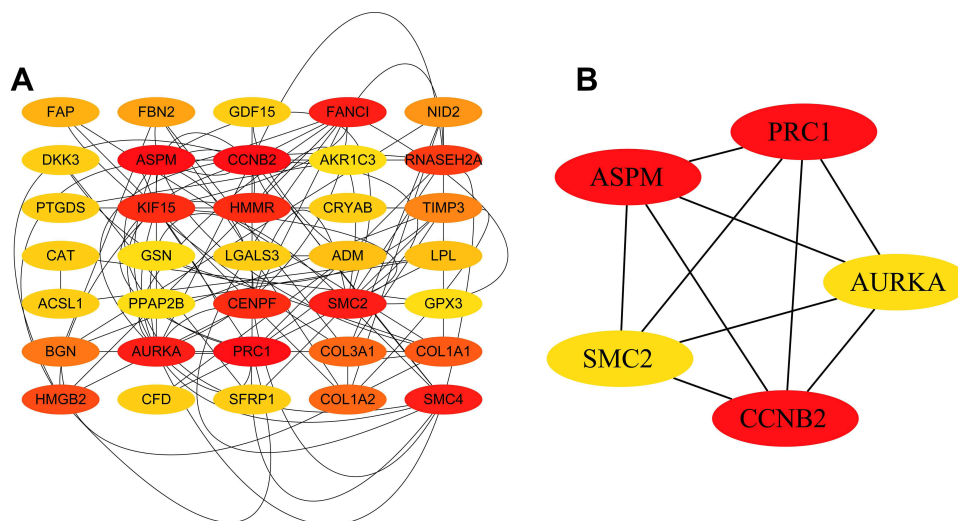


Figure 4 Protein-protein interaction networks for CDEGs (Cytoscape). (A) Protein-protein interaction networks for 42 CDEGs. (B) Protein-protein interaction networks for five identified genes.

The Correlation Between the Expression of the Five Identified Genes and Infiltration Levels of Immune Cells

The tumorigenesis and development of tumors are linked to inflammatory responses and immune cell infiltration.²¹ Thus, we conducted a comprehensive analysis of the relationship between the level of the five identified genes and immune cell infiltration (B cells, CD8⁺T cells, CD4⁺T cells, macrophage, neutrophil, and dendritic cells) in SARC. As shown in Figure 8A, there were significant correlations between the infiltration levels of B cells, CD8⁺T cells, CD4⁺T cells, and ASPM expression levels, while CD4⁺T cell infiltration was significantly correlated to CCNB2 expression level. Besides, the infiltration of CD4⁺T cells and macrophages were closely related to PRC1 and SMC2 expression, while B cells and dendritic cell infiltration were correlated to AURKA expression. Moreover, the enrichment of CD4⁺T and neutrophil cells in sarcoma tissue was accompanied by a longer life expectancy (Figure 8B). Specific information on the distribution of each immune subset at each copy number (arm-level deletion, diploid/normal, arm-level gain, and high amplification) status in SARC is presented in Figure S2.

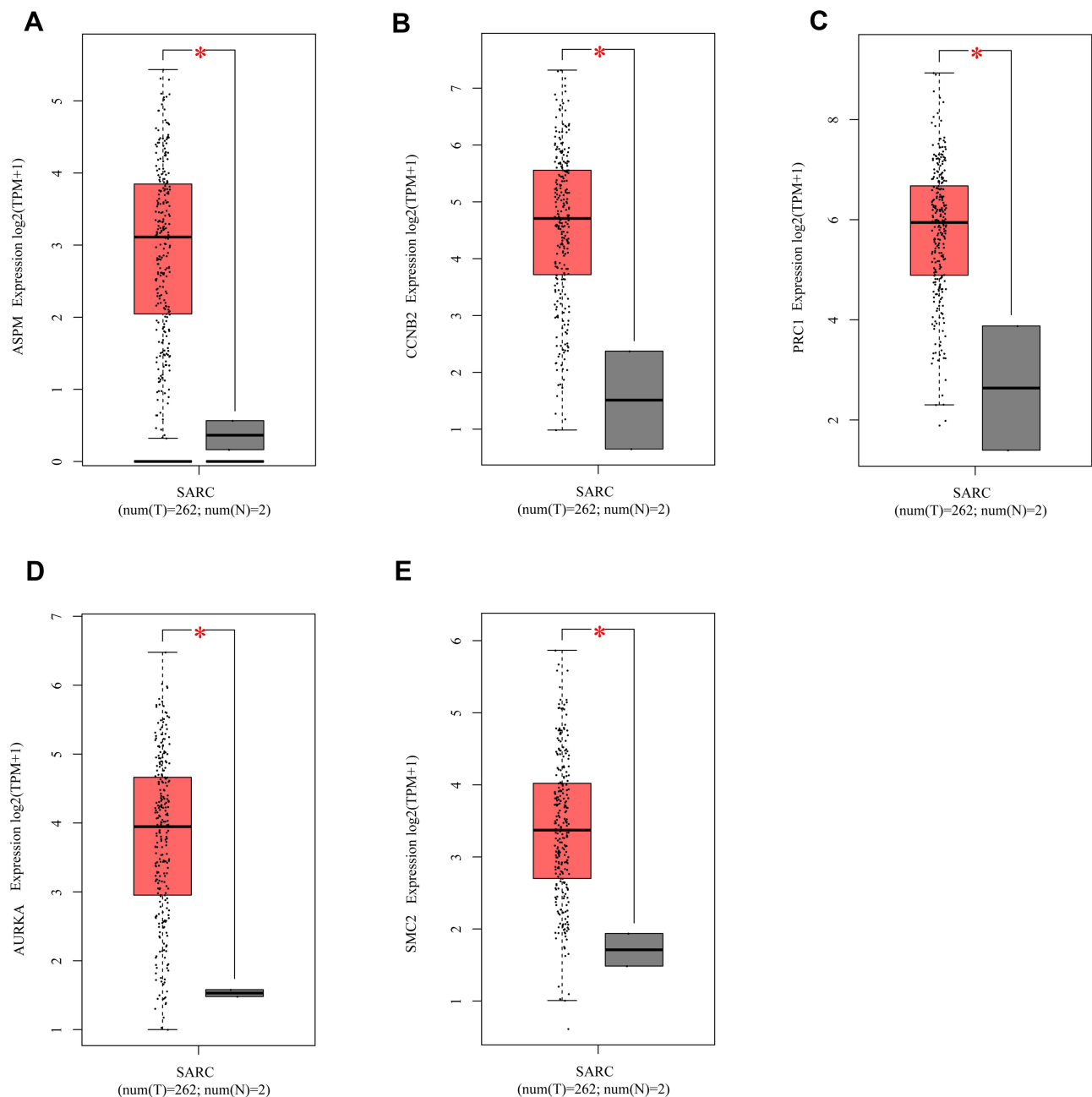


Figure 5 Distinct expression of five identified genes in SARC tissues and adjacent normal tissues (GEPIA). The aberrant expression of ASPM (A), CCNB2 (B), PRC1 (C), AURKA (D), and SMC2 (E) in SARC tissues. *P<0.05.

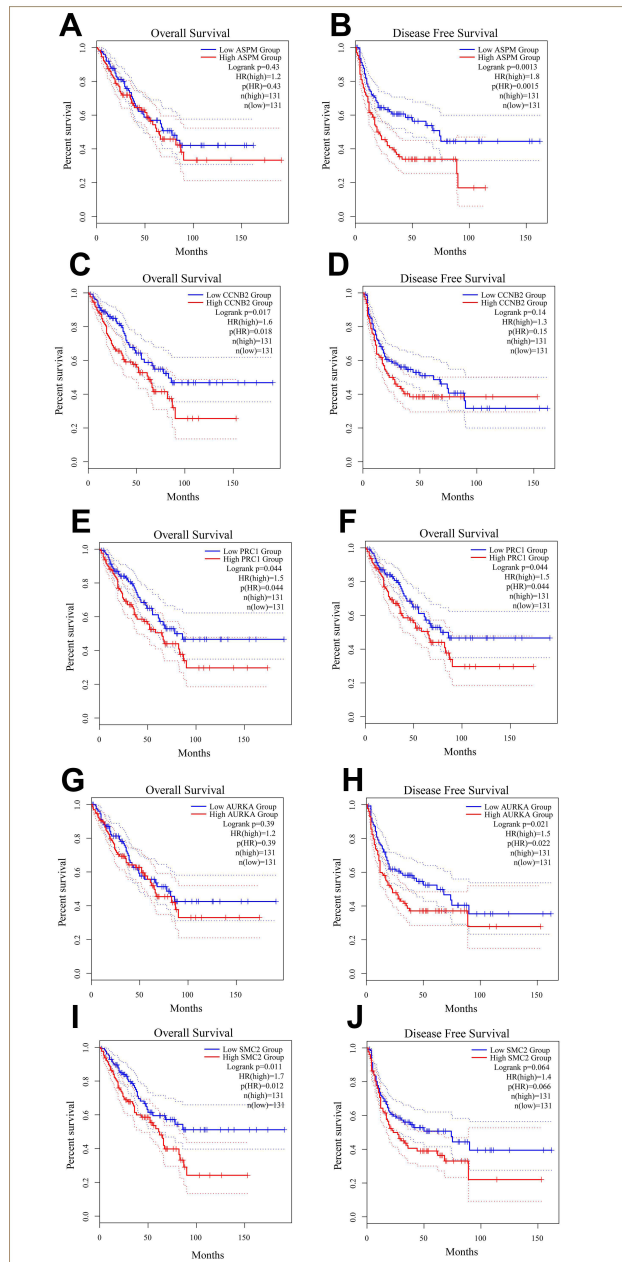
Correlations Between the Five Identified Gene Mutations in SARC and OS, DSS

As shown in Figure 9A, among 640 samples, the mutation rate of ASPM, CCNB2, PRC1, AURKA, and SMC2 was 2.2%, 0.9%, 2%, 2.3%, and 2.2%, respectively. Of interest, we observed lower OS and DSS in the altered group (Figure 9B and C). Moreover, the altered and unaltered groups differed in some clinic parameters such as tumor mutational burden (TMB), Fraction Genome Altered, and MSI sensor Score (Figure S3).

Results of Immune Checkpoint and m⁶A Marker Gene Analysis

The immune checkpoint inhibitor has been successfully applied in the treatment of several tumors.^{22,23} m⁶A modification is the most common nucleotide modification in human internal, which is considered involved in gene expression

GEPIA database



TCGA database

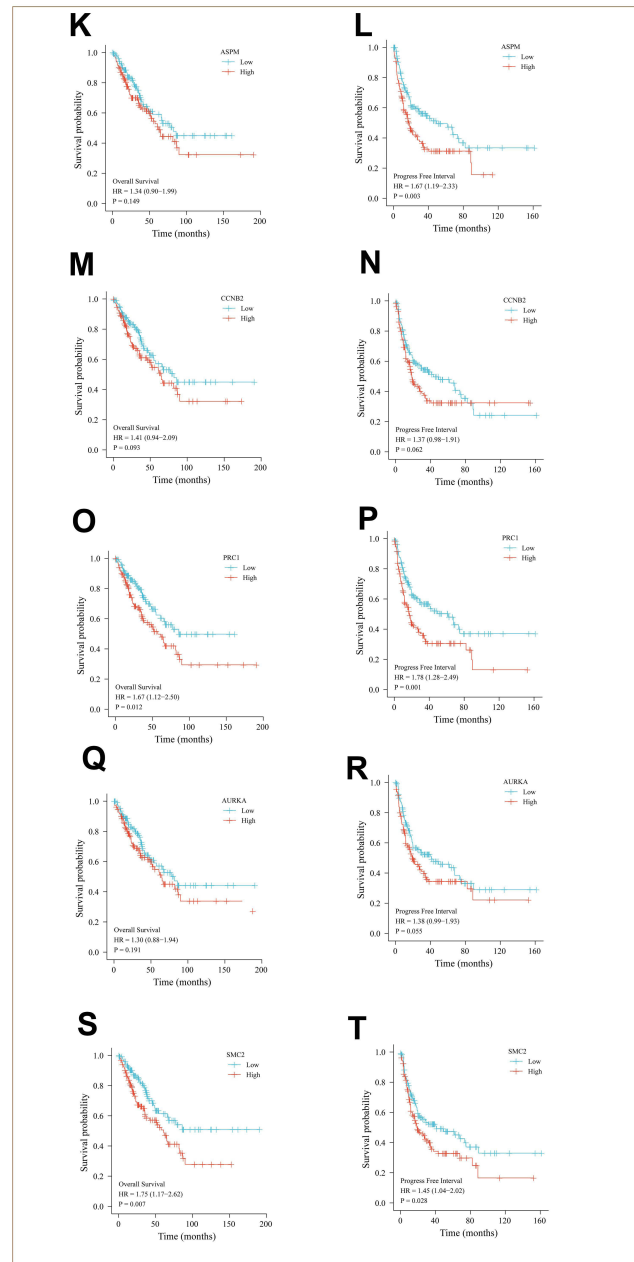


Figure 6 Prognostic role of the five identified genes in SARC patients (GEPIA, TCGA). Effect of ASPM expression on OS (A and K), DFS (B) and PFI (L) in SARC patients. Effect of CCNB2 expression on OS (C and M), DFS (D) and PFI (N) in SARC patients. Effect of PRC1 expression on OS (E and O), DFS (F) and PFI (P) in SARC patients. Effect of AURKA expression on OS (G and Q), DFS (H) and PFI (R) in SARC patients. Effect of SMC2 expression on OS (I and S), DFS (J) and PFI (T) in SARC patients.

regulation and diseases.^{24,25} In this study, we explored the relationship between 60 immune checkpoint genes, m⁶A modification, and the expression of five identified genes. As presented in Figure 10, the ASPM was significantly correlated with 42 of 60 immune checkpoints and 18 of 21 marker genes. And the CCNB2 was obviously associated with 18 of 60 immune checkpoints and 12 of 21 marker genes. For the expression level of PRC1, we observed a significant correlation with 44 of 60 immune checkpoints and 18 of 21 marker genes. As for AURKA, we found that 45 of 60 immune checkpoints and 19 of 21 marker genes were associated with the level of AURKA. Then, the expression level of SMC2 was correlated with 31 of 60 immune checkpoints and 19 of 21 marker genes.

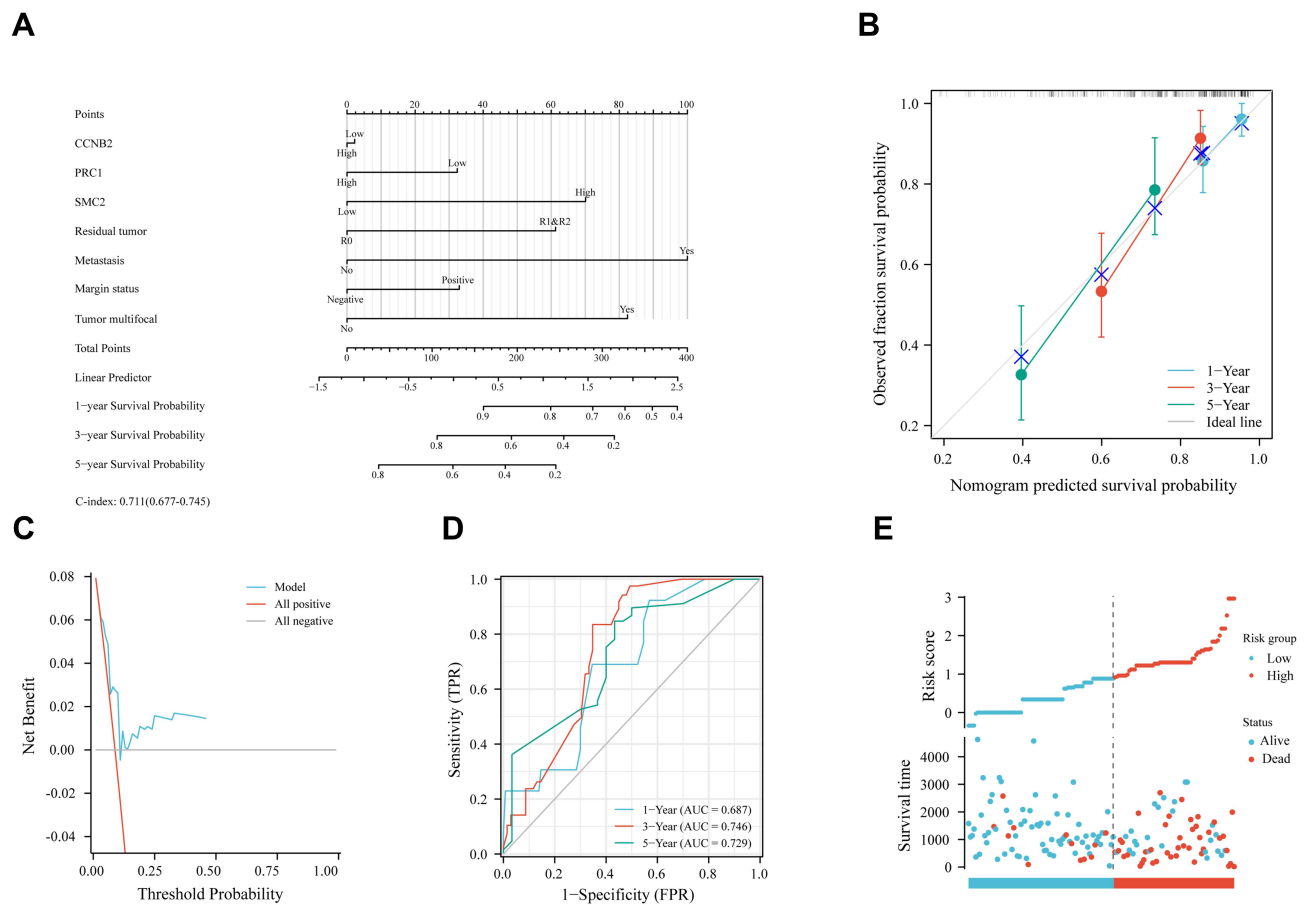


Figure 7 The logistic regression-based predicted nomogram for the probability of survival time among SARC individuals (A). Calibration evaluation of nomogram predicted survival probability (B). Decision curve analysis (DCA) of the nomograms for OS of SARC (C). Time-dependent ROC analysis of the nomograms to predict 1-year, 3-year, and 5-years of OS (D). Analysis of Risk score of a risk-score model based on the Cox regression results of the OS of SARC (E).

Discussion

Many studies have clarified the molecular mechanism and prognostic markers of various subtypes of sarcoma from the perspective of gene expression.^{10–12,14,26,27} However, the prognostic value of the differential expressed genes in the integrated subtypes of sarcoma remains to be reported. In this study, we obtained three GSE series profiles from the GEO database, and differentially expressed genes were selected. Then, 42 CDEGs that were differentially expressed were identified. Based on the analysis of the PPI network, the top five genes (ASPM, CCNB2, PRC1, AURKA and SMC2) were further selected. After that, the correlation between the expression of the five genes and the prognosis, genes mutation, immune infiltration, and methylation of SARC were explored.

Previous studies had reported the suppressive effect in cancer stemness and tumorigenicity when downregulation the expression of ASPM in prostate cancer.²⁸ Consistently, we found the lower expression of ASPM was correlated with better OS and PFI in SARC. It is well acknowledged that the expression of CCNB2 is closely related to the cell cycle pathway. Based on previous research, CCNB2 is correlated with the plasmalemma and is indispensable for the meiosis of mouse oocytes.²⁹ Song et al have reported that the CCNB2 could act as a hub gene related to the progression and prognosis of hepatocellular carcinoma,³⁰ and lower expression of CCNB2 restricts the proliferation and metastasis of tumor cells.³¹ In this study, CCNB2 was overexpressed in SARC than in normal samples and such overexpression was associated with a worse prognosis of SARC. Hernandez-Ortega et al indicated that PRC1 is the first substrate of phosphoregulation in the complex of cyclin-dependent kinase/ cyclin, which decreased the cellular proliferative capacity after downregulating the level of PRC1.^{32,33} AURKA belongs to the family of serine/threonine kinases, which widely participate in the mechanism of occurrence and development of various tumors.³⁴ The high level of AURKB is

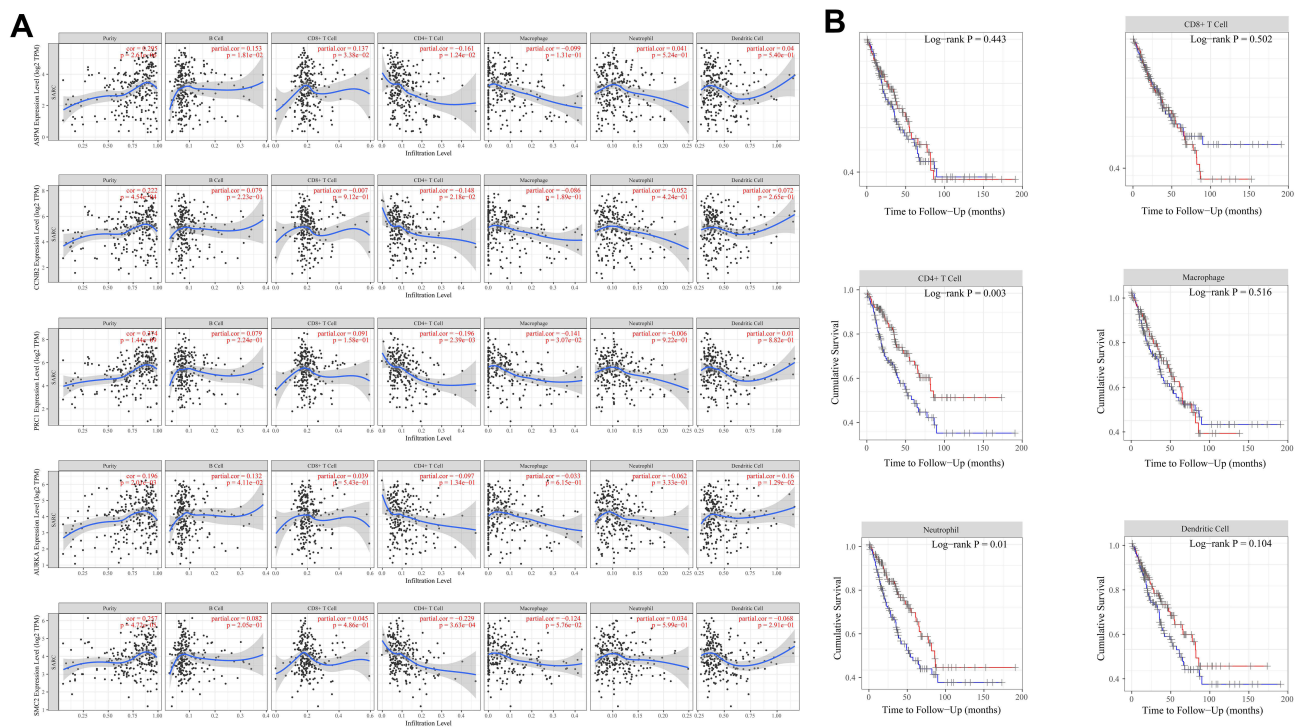


Figure 8 Interactions of COL4As expression and immune system. (A) Correlation of the expression of five identified genes and the level of six immune cells infiltration. (B) The correlation between immune cell infiltration and survival time in SARC patients.

associated with worse survival in SARC patients.³⁴ However, based on GIPIA and TCGA databases, we found the overexpression of AURKA was only related to poor DFS. The impact of the AURKA level on the OS of SARC needs to be further verified. Concerning SCM2, Zhong et al found that it plays an unfavorable role in pancreatic cancer through gene expression sequence analysis.³⁵ The prognostic role of the SMC gene family in sarcoma had also been explored, and the higher expression level of the SMC2 gene was found to be related to the worse prognosis of SARC.²⁶ These results were consistent with the finding of the current study. Sarcomas are a heterogeneous group of tumors that consist of a wide variety of histological subtypes. In terms of disease localization, recurrence rate, aggressiveness, and genetic complexity, the clinic results among sarcoma subtypes also vary.^{36,37} Besides, it's obviously challenging to perform a comprehensive analysis on a single subtype of sarcoma due to their rarity. By comparison of the differential expression of target genes in six sarcoma subtypes, we found the magnitude of target gene level within subtypes was small, which may indicate that the difference of the target genes was not as pronounced in different subtypes of sarcoma. Although age is related to the vast majority of tumors,³⁸ univariate/multivariate analysis of overall survival in SARC patients shows that age was not statistically associated with the prognosis of SARC. No doubt, more *in vivo/vitro* follow-up studies and experiments are needed to verify the clinical values of the five genes in SARC.

Over the past few years, increasing attention has been paid to the relationship between the extracellular matrix and local invasion and distant metastasis. Tumor metastasis and drug resistance are related to changes in the extracellular matrix.^{39,40} The decreased expression of types I and IV collagen $\alpha 1$ gene is related to the drug resistance of chondrosarcoma cells.^{39,41} Noticeably, among the 42 DEGs, TIMP3 belongs to the TIMP gene family, which plays an inhibitory role in matrix metalloproteinases, a group of peptidases involved in the degradation of the extracellular matrix. Interestingly, another gene of this family, TIMP1, has been recently correlated with soft sarcoma response to chemotherapy and cell invasiveness, exhibited a positive prognostic role in soft tissue sarcoma DFS.³⁹ Therefore, the abnormalities in signaling and structural components of the extracellular matrix may play a vital role in sarcoma development.³⁹ In this study, we also found that the 42 CDEGs were most significantly enriched in signaling pathways related to “collagen-containing extracellular matrix” pathways, which suggested a vital function of extracellular matrix in SARC.

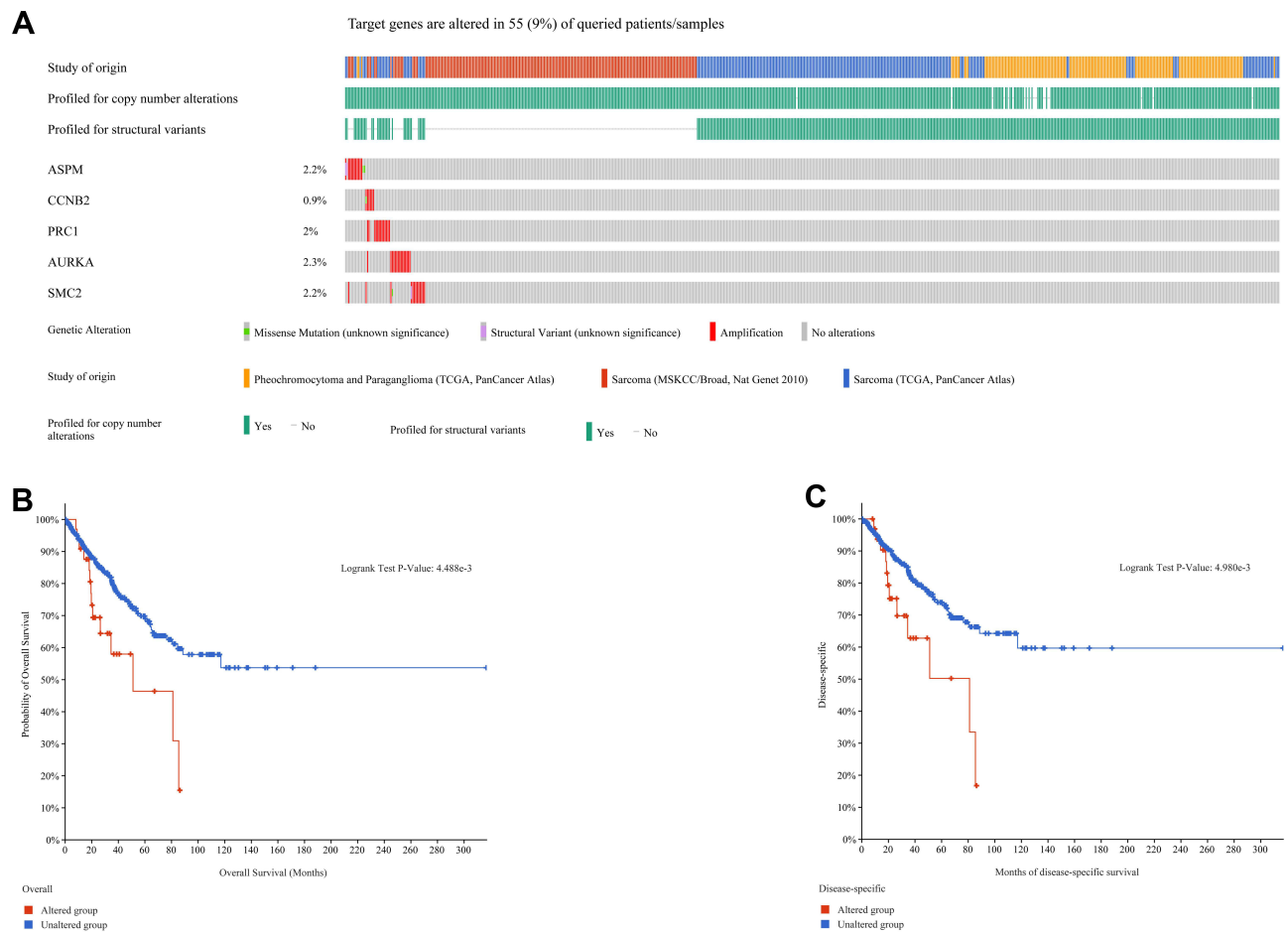


Figure 9 Genetic mutations in the five identified genes and their association with OS and DSS of SARC patients (cBioPortal). A minor mutation rate (9%) of the target genes was observed in SARC patients (A). mRNA expression z-scores of target genes were related to all samples. A significant correlation was noticed between the genetic alterations of the five identified genes and DSS, OS in SARC (B and C).

The occurrence and development of tumors are always accompanied by genes mutation.^{7,42} In the current study, a low survival-related mutation rate of five identified genes was observed, suggesting that the mutation of the five identified genes was harmful and shortened the survival time of SARC patients. Therefore, genetic assessment of five identified genes mutations should be considered in measuring clinical outcomes of SARC. However, the correlation between the clinical parameters of SARC and the mutation of COL4As needs to be verified by further lab results.

Immune cell infiltration not only plays an essential role in tumors but also acts as the reference for both immunotherapy and clinical outcome.^{43,44} In this study, the significant correlation between the expression level of all five identified genes and immune infiltration indicated a tight interaction between gene expression and immune cell infiltration in SARC. Meanwhile, immune checkpoint analysis has been a hot spot in the past few years due to ineffective immunotherapy in some tumors, especially in late-stage patients.⁴⁵ It has been proved that immune checkpoints are involved in several types of soft tissue sarcoma.⁴⁶ Previously, many studies have shown that RNA methylation (m⁶A) participated in the initiation, progression, and metastasis of many tumors.^{24,25} And M⁶A regulators are found as an independent indicator in the prognosis of some tumor patients such as colorectal cancer⁴⁷ and gastric cancer.⁴⁸ In the current study, we found that five identified gene expression levels were significantly associated with multiple immune checkpoint genes and m⁶A regulators, which indicated that five identified genes widely participated in the clinic progress of SARC through checkpoint blockade and modification of m⁶A.

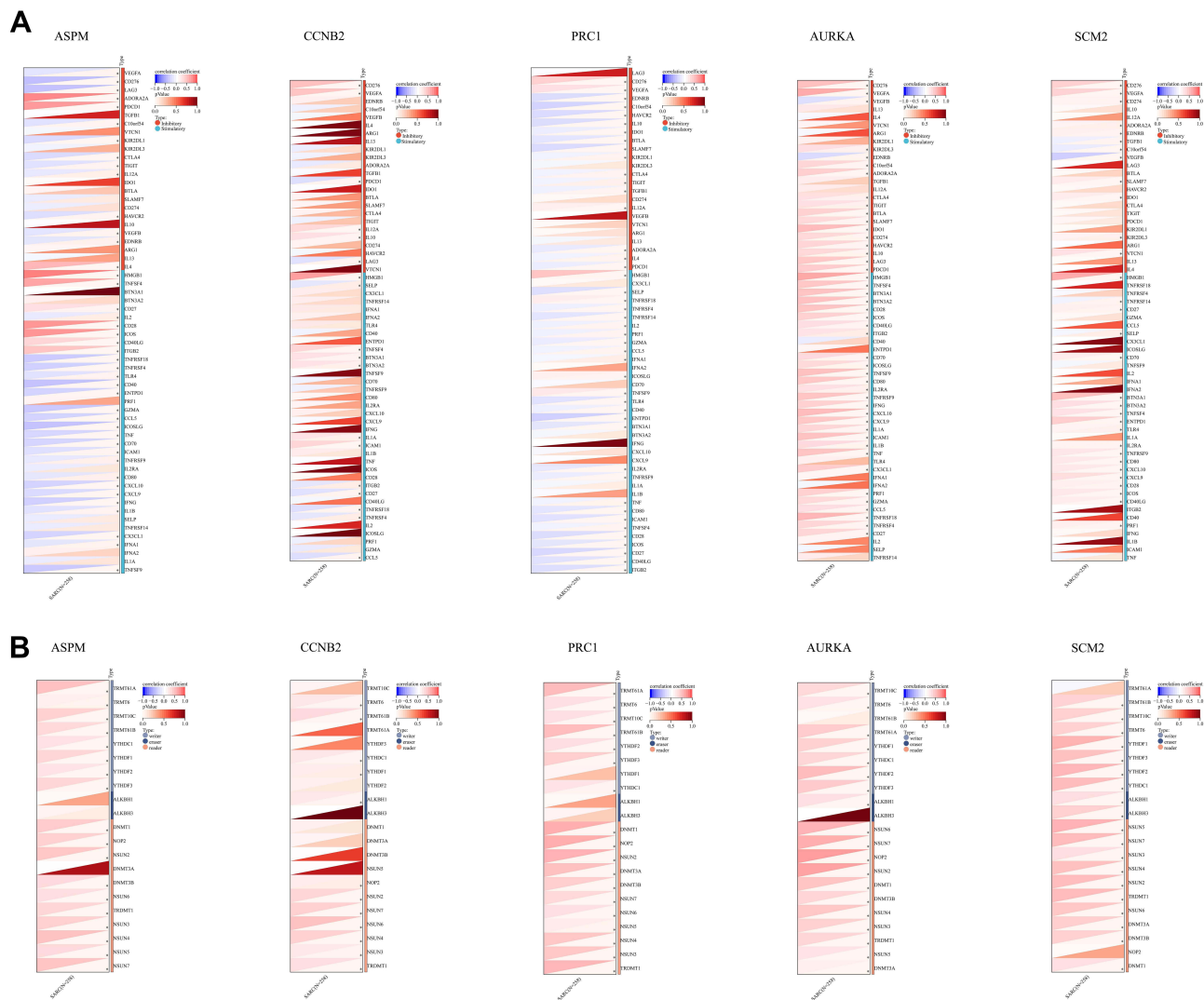


Figure 10 Correlation of the expression of five identified genes and 60 immune checkpoints (A), 21 marker genes of m⁶A modification (B). *P<0.05.

Limitations

As the first study to investigate the multimolecular characteristics and prognostic role of the five identified genes in SARC, several limitations should be recognized. First, we did not explore the role of target genes in specific sarcoma subtypes of SARC. Second, we used multiple online databases to perform the current comprehensive analysis, the data heterogeneity would be inevitably existing, which could reduce the reliability of our results. Further follow-up clinical studies and experiments in vitro/vivo are required to verify the clinical values of the five identified genes in each subtype of SARC.

Conclusion

We comprehensively analyzed the relationship between the expression level of the five identified genes and the prognosis, mutation, immune infiltration, methylation, and checkpoints of SARC patients. We found the expression level of CCNB2, PRC1, and SMC2 was significantly correlated with the survival of SARC. The five identified genes are widely involved in immune infiltration and impact the prognosis of SARC. Meanwhile, we found the five identified genes were significantly associated with m⁶A modification and immune checkpoints. This study provides a theoretical basis for the research about the correlation between the level of five identified genes and sarcoma, but further experiments are needed to explore the underlying mechanisms.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding authors upon reasonable request.

Ethics Approval and Consent to Participate

Ethical approval for our study was granted by The Committee on Medical Ethics of The First Affiliated Hospital of Anhui Medical University (Reference number: Quick-PJ2021-16-30). Since all the data used in the current study was available online, and no individual patient was involved, it could be confirmed we have obtained all the written informed consent.

Acknowledgments

We thank all the people who offer help for this study. Co-first author: Bizhi Tu and Yaya Jia contributed equally to this paper.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by grants from National Natural Science Foundation of China (81471273) and Natural Science Research Key Project in College of Anhui Province (KJ2022ZD23).

Disclosure

The authors declare that they have no conflicts of interest in this work.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30. doi:10.3322/caac.21590
2. Ferguson JL, Turner SP. Bone cancer: diagnosis and treatment principles. *Am Fam Physician*. 2018;98(4):205–213.
3. Judson I, Verweij J, Gelderblom H, et al. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled Phase 3 trial. *Lancet Oncol*. 2014;15(4):415–423. doi:10.1016/S1470-2045(14)70063-4
4. Karavasilis V, Seddon BM, Ashley S, Al-Muderis O, Fisher C, Judson I. Significant clinical benefit of first-line palliative chemotherapy in advanced soft-tissue sarcoma: retrospective analysis and identification of prognostic factors in 488 patients. *Cancer*. 2008;112(7):1585–1591. doi:10.1002/cncr.23332
5. Maki RG. Gemcitabine and docetaxel in metastatic sarcoma: past, present, and future. *Oncologist*. 2007;12(8):999–1006. doi:10.1634/theoncologist.12-8-999
6. Ayodele O, Razak ARA. Immunotherapy in soft-tissue sarcoma. *Curr Oncol*. 2020;27(Suppl 1):17–23. doi:10.3747/co.27.5407
7. Bourcier K, Le Cesne A, Tselikas L, et al. Basic knowledge in soft tissue sarcoma. *Cardiovasc Intervent Radiol*. 2019;42(9):1255–1261. doi:10.1007/s00270-019-02259-w
8. Howard JH, Pollock RE. Intra-abdominal and abdominal wall desmoid fibromatosis. *Oncol Ther*. 2016;4(1):57–72. doi:10.1007/s40487-016-0017-z
9. Berhe S, Danzer E, Meyers P, Behr G, LaQuaglia MP, Price AP. Unusual abdominal metastases in osteosarcoma. *J Pediatr Surg Case Rep*. 2018;28:13–16. doi:10.1016/j.epsc.2017.09.022
10. Abeshouse A, Adebamowo C, Adebamowo SN; Cancer Genome Atlas Research Network. Electronic address edsc, cancer genome atlas research N. comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell*. 2017;171(4):950–65 e28. doi:10.1016/j.cell.2017.10.014
11. Long Z, Wu T, Tian Q, Carlson LA, Wang W, Wu G. Expression and prognosis analyses of BUB1, BUB1B and BUB3 in human sarcoma. *Aging*. 2021;13(9):12395–12409. doi:10.18632/aging.202944
12. Liu Q, Bao Q, Xu Y, et al. MCM4 is a novel biomarker associated with genomic instability, BRCAness phenotype, and therapeutic potentials in soft-tissue sarcoma. *Front Cell Dev Biol*. 2021;9:666376. doi:10.3389/fcell.2021.666376
13. Shan W, Akinfenwa PY, Savannah KB, et al. A small-molecule inhibitor targeting the mitotic spindle checkpoint impairs the growth of uterine leiomyosarcoma. *Clin Cancer Res*. 2012;18(12):3352–3365. doi:10.1158/1078-0432.CCR-11-3058
14. Zhou J, Wang MY, Zhou Z, Wang WC, Duan J, Wu G. Expression and prognostic value of MCM family genes in osteosarcoma. *Front Mol Biosci*. 2021;8. doi:10.3389/fmolb.2021.668402

15. Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics*. 2014;47:8–13. doi:10.1002/0471250953.bi0813s47
16. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98–W102. doi:10.1093/nar/gkx247
17. Tomczak K, Czerwinska P, Wiznerowicz M. The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol*. 2015;19(1A):A68–A77.
18. Li TW, Fan JY, Wang BB, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. 2017;77(21):E108–E110. doi:10.1158/0008-5472.CAN-17-0307
19. Gao JJ, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269). doi:10.1126/scisignal.2004088
20. Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity*. 2018;48(4):812–30 e14. doi:10.1016/j.immuni.2018.03.023
21. Grivnennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell*. 2010;140(6):883–899. doi:10.1016/j.cell.2010.01.025
22. Jerby-Arnon L, Shah P, Cuoco MS, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell*. 2018;175(4):984–97 e24. doi:10.1016/j.cell.2018.09.006
23. Abril-Rodriguez G, Ribas A. SnapShot: immune checkpoint inhibitors. *Cancer Cell*. 2017;31(6). doi:10.1016/j.ccell.2017.05.010
24. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev*. 2015;29(13):1343–1355. doi:10.1101/gad.262766.115
25. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. *Science*. 2018;361(6409):1346–1349. doi:10.1126/science.aau1646
26. Zhou J, Wu G, Tong Z, et al. Prognostic relevance of SMC family gene expression in human sarcoma. *Aging*. 2020;13(1):1473–1487. doi:10.18632/aging.202455
27. Zhu Z, Jin Z, Deng Y, et al. Co-expression network analysis identifies four hub genes associated with prognosis in soft tissue sarcoma. *Front Genet*. 2019;10:37. doi:10.3389/fgene.2019.00037
28. Pai VC, Hsu CC, Chan TS, et al. ASPM promotes prostate cancer stemness and progression by augmenting Wnt-Dvl-3-beta-catenin signaling (vol 38, pg 1340, 2018). *Oncogene*. 2019;38(8):1354. doi:10.1038/s41388-018-0561-0
29. Daldello EM, Luong XG, Yang CR, Kuhn J, Conti M. Cyclin B2 is required for progression through meiosis in mouse oocytes. *Development*. 2019;146(8). doi:10.1242/dev.172734
30. Song XD, Du R, Gui H, et al. Identification of potential hub genes related to the progression and prognosis of hepatocellular carcinoma through integrated bioinformatics analysis. *Oncol Rep*. 2020;43(1):133–146. doi:10.3892/or.2019.7400
31. Lei CY, Wang W, Zhu YT, Fang WY, Tan WL. The decrease of cyclin B2 expression inhibits invasion and metastasis of bladder cancer. *Urol Oncol-Semin Ori*. 2016;34(5):237–e1.
32. Li T, Wang XX, Jing LJ, Li Y. MiR-1-3p inhibits lung adenocarcinoma cell tumorigenesis via targeting protein regulator of cytokinesis 1. *Front Oncol*. 2019;9:120.
33. Hernandez-Ortega S, Sanchez-Botet A, Quandt E, et al. Phosphoregulation of the oncogenic protein regulator of cytokinesis 1 (PRC1) by the atypical CDK16/CCNY complex. *Exp Mol Med*. 2019;51(4):1–17. doi:10.1038/s12276-019-0242-2
34. Du RJ, Huang CT, Liu KD, Li X, Dong ZG. Targeting AURKA in Cancer: molecular mechanisms and opportunities for Cancer therapy. *Mol Cancer*. 2021;20(1). doi:10.1186/s12943-020-01305-3
35. Zhong J, Jermusyk A, Wu L, et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer. *J Natl Cancer Inst*. 2020;112(10):1003–1012. doi:10.1093/jnci/djz246
36. Agulnik M, Costa RLB, Milhem M, et al. A Phase II study of tivozanib in patients with metastatic and nonresectable soft-tissue sarcomas. *Ann Oncol*. 2017;28(1):121–127. doi:10.1093/annonc/mdw444
37. Davis LE, Bolejack V, Ryan CW, et al. Randomized double-blind Phase II study of regorafenib in patients with metastatic osteosarcoma. *J Clin Oncol*. 2019;37(16):1424–1431. doi:10.1200/JCO.18.02374
38. Hashimoto K, Nishimura S, Hara Y, et al. Clinical outcomes of patients with primary malignant bone and soft tissue tumor aged 65 years or older. *Exp Ther Med*. 2019;17(1):888–894. doi:10.3892/etm.2018.7013
39. De Vita A, Recine F, Miserocchi G, et al. The potential role of the extracellular matrix in the activity of trabectedin in UPS and L-sarcoma: evidences from a patient-derived primary culture case series in tridimensional and zebrafish models. *J Exp Clin Cancer Res*. 2021;40(1):165. doi:10.1186/s13046-021-01963-1
40. Cui J, Dean D, Hornicek FJ, Chen Z, Duan Z. The role of extracellular matrix in osteosarcoma progression and metastasis. *J Exp Clin Cancer Res*. 2020;39(1):178. doi:10.1186/s13046-020-01685-w
41. D'Incalci M, Galmarini CM. A review of trabectedin (ET-743): a unique mechanism of action. *Mol Cancer Ther*. 2010;9(8):2157–2163. doi:10.1158/1535-7163.MCT-10-0263
42. Asthagiri AR, Parry DM, Butman JA, et al. Neurofibromatosis type 2. *Lancet*. 2009;373(9679):1974–1986. doi:10.1016/S0140-6736(09)60259-2
43. Zhou CK, Li CM, Peng SQ, Zhou LC, Li H. Comprehensive analysis of the relationships between tumor mutation burden with immune infiltrates in cervical cell carcinoma. *Front Mol Biosci*. 2020;7. doi:10.3389/fmolb.2020.582911
44. Hu J, Yu J, Gan J, et al. Notch1/2/3/4 are prognostic biomarker and correlated with immune infiltrates in gastric cancer. *Aging*. 2020;12(3):2595–2609. doi:10.18632/aging.102764
45. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248–262. doi:10.1016/j.celrep.2016.12.019
46. Hashimoto K, Nishimura S, Ito T, Akagi M. Characterization of PD-1/PD-L1 immune checkpoint expression in soft tissue sarcomas. *Eur J Histochem*. 2021;65(3). doi:10.4081/ejh.2021.3203
47. Ji L, Chen S, Gu L, Zhang X. Exploration of potential roles of m6A regulators in colorectal cancer prognosis. *Front Oncol*. 2020;10:768. doi:10.3389/fonc.2020.00768
48. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol Cancer*. 2020;19(1):53. doi:10.1186/s12943-020-01170-0

International Journal of General Medicine

Dovepress

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>