

Integrative Proteogenomic Analyses Identify TOPs as Prognostic Markers and Therapeutic Targets in Sarcoma

Genchun Wang

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Kai Sun

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Li Ting Liu

Department of Obstetrics and Gynecology Department, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Jiamin Lin

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Xudong Yao

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Jiachao Guo

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Zhou Guo

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Liang Cai Hou

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

Ting Xu

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

fengjing guo (✉ guofjdoc@163.com)

HUST: Huazhong University of Science and Technology <https://orcid.org/0000-0002-5968-4836>

Keywords: bioinformatic analysis, soft-tissue sarcoma, DNA topoisomerase, immune infiltration, mutation

Posted Date: January 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-139866/v1>

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

[Read Full License](#)

Abstract

Background: Soft-tissue sarcomas (STSs) are uncommon malignant tumors with varied highly aggressive and histological presentations. They comprise approximately 1% of newly diagnosed adult malignancies each year, but the sarcoma's accurate diagnosis prediction remains a challenge. TOP1/1MT/2A/2B/3A/3B were aberrantly expressed in tumors and associated with tumor progression, but the roles of the six TOP members in sarcoma in clinical trials remains uncertain.

Method: Relationship between the mRNA levels of the TOPs and the clinicopathological parameters evaluated by Oncomine, GEPIA, cBioPortal, TIMER, Metascape, TCGAportal, Kaplan-Meier plotter, String, and Linkedomics. We focused on the potential mechanism of TOPs in the occurrence and development of SARC. We were investigating the prognostic value of TOPs expression in patients with SARC and correlations with clinicopathological features.

Results: In patients with SARC, TOPs mRNA expression level upregulated. Increased TOP family members' expression was associated with different subtypes, FNCLCC grades, Local recurrence, DNA methylation, different miRNA clusters, and poor prognosis in SARC patients. We further found that TOPs showed distinct immune infiltration patterns, TOP gene mutation rate and that TOP somatic copy number variations were negatively related to the survival of sarcoma patients. Survival data analyzed that high transcription levels of TOPs were associated with low overall survival and disease-free survival in sarcoma. Enrichment analyses for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genome (KEGG) were conducted to reveal TOPs enriched in DNA topological change, Chromosome segregation, DNA, replication signaling pathway. BUB1B, CENPF, SPC25, NUF2, KNL1SGO1, RPS20, SNRPD3, and MED9 were significantly associated with TOP mutations. These Hub gene expressions in SARC were increased and were associated with unfavorable prognosis.

Conclusion: The results show that multiple TOP family members are involved in the development and progression of sarcoma and suggest that TOPs can be used as potential targets for the diagnosis and treatment of sarcoma. TOP1/1MT/2A/2B/3A/3B may affect the prognosis of SARC through DNA topological change, Chromosome segregation, DNA, replication signaling pathway.

1. Background

Adult soft-tissue sarcomas are diverse mesenchymal malignancies that account for 1% of solid adult tumors. Many are highly aggressive, accounting for a disproportionate share of cancer mortality among young adults. DNA topoisomerase (TOP) is responsible for DNA unlinking, and these ubiquitous enzymes play critical roles in many biological processes involving DNA(1)(2). There are two types of DNA TOPs, type I and type II.

Currently, there are only a few diagnostic and prognostic markers, and the cellular origin of several sarcoma subtypes is unknown. Therefore, the accurate diagnosis and prediction of many of these tumors' clinical behavior remain a challenge. High-grade sarcomas have a high rate of local recurrence,

frequent metastases, and poor prognosis. For sarcoma patients with isolated, resectable, local recurrences, surgery is the sole potentially curative treatment. However, the oncological results remain worse than those in patients with the locally recurrent disease even after macroscopic resection. Unfortunately, these surgeries are often extensive, with significant postoperative morbidity and mortality.

Therefore, different strategies are needed to treat patients with soft-tissue sarcoma. The heterogeneity of sarcomas in molecular genesis, histology, clinical features, and treatment response regulates these rare but diverse neoplasms and is particularly challenging. Often, search results in one subtype do not translate to others. These limitations increase in the context that sarcomas are among the rarest cancers(3).

Meanwhile, the histopathologic type of sarcoma is sometimes tricky to identify, and morphological evaluation based on the microscopic examination of histologic sections remains the gold standard for sarcoma diagnosis. Several types of adjuvant methods are emerging in support of the morphological diagnosis. In recent years, scientific research on sarcoma based on genetics has made great strides. Considering the underlying pathogenesis and etiology of sarcoma will facilitate the discovery of advanced treatments and diagnostic biomarkers. Studies have shown that specific cancer genetics, such as mutations, single nucleotide polymorphisms (SNPs), translocations, deletions, and insertions, may help regulate cancer genes.

To date, 6 DNA TOPs have been detected and numbered in the order of their detection (TOP1, TOP1MT, TOP2A, TOP2B, TOP3A, and TOP3B), and each TOP enzyme has a specific function. The integrity of the DNA helix during the DNA transaction processes includes replication, transcription, and recombination. DNA TOPs have maintained a domain and are a natural tool in addressing DNA complexity by local transformation, controlling the structure of the DNA chromosome. Increasing experimental evidence indicates that TOPs are involved in breast cancer tumorigenesis. TOP1MT was overexpressed in cancerous tissue, and TOP1MT deficiency has been demonstrated to attenuate tumor growth in human and mouse models of colon cancer and liver cancer(4). Top3 α is a crucial enzyme, and null homozygous mice die in the womb. The phenotypes highlight the critical role of Top3 α in maintaining genomic stability during development(5). The expression of Top3 β in breast ductal carcinomas was a significant predictor of survival, indicating that Top3 β plays a significant role in suppressing tumorigenesis and is a significant predictor of survival in ductal breast cancer(6). TOP1-targeted and TOP2-targeted anticancer drugs are commonly used in gastrointestinal malignancies (colorectal and gastroesophageal malignancies). To our knowledge, no bioinformatic analysis has been performed to study the role of TOPs in sarcoma. However, the potential clinical and prognostic value of TOPs and the underlying mechanism of TOPs are inconsistent and even contradictory. Microarray-based comparative genomic hybridization (CGH) and the differential expression of mRNAs and miRNAs have revealed genomic changes, candidates for genes and miRNAs that can be utilized to differentiate sarcoma subtypes and to confirm disease progression, and that is therapeutic targets(7)(8)(9)(10)(11). Based on multiplatform sarcoma databases, we analyzed the expression level of the TOP family members, the relationship between TOPs and clinicopathological features and prognosis of patients with sarcoma, identified

prognostic and predictive biomarkers using genomics and proteomics approaches in the present study. TOPs were demonstrated to have therapeutic potential for sarcoma.

2. Materials And Methods

2.1 OncoPrint analysis (<https://www.oncoPrint.org/resource/login.html>)

The expression levels of TOPs in sarcoma patient tumor samples were determined from the OncoPrint Compendium of Expression Array data. OncoPrint gene expression array records were used to analyze the transcription levels of TOPs in different types of cancer. The mRNA expression of TOPs in clinical cancer samples was comparable to that in normal controls using Student's t-test to generate a P-value. The threshold values of the P-value and fold change were set as 0.01 and 1.5, respectively.

2.2 GEPIA dataset

Gene Expression Profiling Interactive Analysis (GEPIA, <https://www.gepia.cancer-pku.cn>) is an open-source database used for performing in-depth analyses of The Cancer Genome Atlas (TCGA) gene expression data and was employed to analyze the expression of TOPs and hub genes in sarcoma and normal samples. We sought to determine the correlation between TOPs and the clinicopathological features of sarcoma (cancer stage, tumor grade, race, weight, stage.) and its potential prognostic value. An analysis of disease-free survival (DFS, also known as recurrence-free survival (RFS)) based on gene expression, was carried out via GEPIA(12). ANOVA and nonparametric tests were used to determine the association between the clinicopathological parameters and the co-expression eigenvector modules were accurate.

2.3 The Kaplan-Meier Plotter

Kaplan-Meier Plotter (www.kmplot.com) was used to analyze patients' gene expression data and survival data with sarcoma. For the analysis of overall survival (OS) and DFS, the samples were divided into two groups based on the mean expression level. The relationship with biomarker values was evaluated by the log-rank test ($P < 0.05$)(13).

2.4 cBioPortal

cBioPortal is a web resource for researching, visualizing, and analyzing multidimensional cancer genome data and mapping the hub gene network(14).

2.5 TIMER

The relationship between the frequency of tumor immune infiltrates (B cells, CD4+ T cells, CD8+ T cells, dendritic cells, macrophages, and neutrophils) was analyzed using the Tumor IMMune Estimation Resource (TIMER) online platform (<http://cistrome.shinyapps.io/timer/>). This web service contains 10897 samples of various cancers included in TCGA. In short, this platform uses a computational method to estimate the infiltration of different immune cells in 23 types of tumors based on public data from TCGA and their correlation with various factors, such as gene expression levels, mutations, or prognosis. Immunohistochemistry (IHC) was used to confirm the predictions in several tumor types, including sarcoma. Correlation graphs show the partial Spearman correlation corrected for purity and its statistical significance, where purity is defined as the percentage of malignant cells in tumor tissue. TCGA contains both sequencing and pathological data on 30 different types of cancer. The sarcoma dataset (TCGA, Provisional), including data with pathology reports, was selected for further TOP analysis with cBioPortal. Genome profiles included mutations, putative copy number changes (copy number alterations [CNAs]) through genomic identification of significant cancer targets (GISTIC 2.0), mRNA expression z-scores (RNA-seq v.2 RSEM), and z-scores of protein expression (reversed-phase protein array [RPPA]). The co-expression of genes and the network calculated according to the online instructions from cBioPortal(15).

2.6 Metascape

We first identified all statistically enriched terms(16) and cumulative hypergeometric P-values and enrichment factors calculated and used for filtering. The remaining significant terms were then grouped hierarchically in a tree based on the Kappa-statistical similarities between their genetic memberships (similar to that used on the National Cancer Institute [NCI] Database for Annotation, Visualization, and Integrated Discovery [DAVID] website). A kappa score of 0.3 was used as the threshold for placing the tree in term clusters(17).

3. Results

3.1. Transcription levels of TOPs in patients with sarcoma

The development of microarray and RNA sequencing technology has made RNA research an integral part of biomedical research. Therefore, in this study, we analyzed TOP members' expression based on diverse databases and examined their correlations with clinicopathological features, prognosis, and local tumor recurrence, and explored the possible regulatory mechanism in patients with sarcoma. Six TOP family members have been found in mammalian cells. We compared the transcription levels of TOPs in 20 different types of cancer diseases with those in normal samples utilizing the OncoPrint database (Figure 1). Student's t-test compared differences in transcriptional expression. The cutoff of the P-value was 0.01, and that of fold change was 1.5.

3.2. Significant changes in TOP expression at the transcription level between sarcoma and normal tissues (Oncomine)

To date, six DNA TOPs have been identified in mammalian cells. The Oncomine database was used to compare the mRNA levels of TOP1, TOP1MT, TOP2A, TOP2B, TOP3A, and TOP3B between sarcoma and normal tissues (Table 1). It was found that the mRNA levels of TOP1, TOP2A, TOP2B, TOP3A, and TOP3B were upregulated in patients with sarcoma. Further analysis showed that TOP1 was overexpressed in sarcoma patients in the Detwiller Sarcoma dataset, with a fold change of 2.164 and a P-value of 4.5E-4. In addition, Top2A mRNA was overexpressed in fibrosarcoma (fold change=27.421), pleomorphic liposarcoma (fold change=17.927), LMS (fold change=12.112), malignant fibrous histiocytoma (fold change= 21.095), round cell liposarcoma (fold change= 13.321), and SS (fold change= 26.646). TOP2B was highly expressed in DDLPS (fold change=2.002), fibrosarcoma (fold change= 2.938), round cell liposarcoma (fold change= 2.491), and SS (fold change= 3.108). In the Barretina Sarcoma dataset, TOP2A was found to have higher expression in myxofibrosarcoma (fold change=12.264), pleomorphic liposarcoma (fold change=9.788), DDLPS (fold change=7.746), and myxoid/round cell liposarcoma (fold change=4.104). Top2b was overexpressed in myxoid/round cell liposarcoma (fold change = 2.318), DDLPS (fold change= 1.65), and pleomorphic liposarcoma (fold change = 1.505). Top3a was overexpressed in myxofibrosarcoma (fold change=1.758), pleomorphic liposarcoma (fold change=1.676), and LMS (fold change =1.757) in the Quade Uterus dataset, and high TOP2A expression was found in uterine corpus leiomyosarcoma (fold change= 3.894) compared to normal samples.

3.3. Relationship between the mRNA levels of the TOPs and the clinicopathological parameters of patients with sarcoma

Using GEPIA (<http://gepia.cancer-pku.cn/>), we compared the mRNA expression of TOP family members between sarcoma and adjacent normal tissues. The expression levels of TOP1, TOP1MT, TOP2A, TOP2B, TOP3A, and TOP3B were higher in sarcoma tissues than in normal tissues (Figure 2A, 2B, and 2C).

3.4. Relationship between TOP mRNA expression and sarcoma grade

There are differences in the expression of TOPs in distinctive grades of sarcomas. We analyzed the expression of TOPs using the classification system of the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) for sarcomas (Figure 3A). The revision of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 8th Edition (2017) built on TNM stage and tumor grade. The AJCC system follows the classification system of the sarcoma groups of the FNCLCC, a 3-level system

founded on the differentiation of tumor cells, mitotic activity, and the extent of necrosis. The panel recommends determining histological grade using the FNCLCC or AJCC/NCI system or an appropriate diagnostic-specific scoring system. The FNCLCC grading system can better reflect the patient's prognosis and predict distant metastasis and death in cancer patients. Postoperative chemotherapy can improve the metastasis-free survival and OS times in patients graded with the FNCLCC system(18). The mRNA expression level of TOPs in different FNCLCC grades correlated significantly with the individual tumor grade. In sarcoma, TOP1, TOP2A, and TOP3B differed considerably between FNCLCC grades 2 and 3, while TOP1 and TOP2A differed between grades 1 and 3. TOP mRNA expression correlated with the individual tumor grade (P-value <0.05). TOP1, TOP2A, and TOP3B can be used to identify soft-tissue sarcoma patients with different grades, and the local treatment options (low expression) can thus be discussed.

Month to distant Recurrence harms long-term survival and quality of life in several malignancies(19). Despite improvements in local tumor control through surgery, radiation therapy, and chemotherapy, distant metastases and high tumor-related mortality remain problems with current treatment strategies(20). After surgical resection, Month to distant Recurrence plays a major role in reducing patients' survival and quality of life with sarcoma. The patterns and risk factors for Month to distant Recurrence have been highlighted(21). Important prognostic factors for predicting Month to distant recurrence are patient age, tumor size, completeness of the resection, grade, tumor rupture, multiplicity, histological subtype, and previous radiation therapy(22). We divided the data of 206 patients from TCGA into two groups based on the presence or absence of TOP mutations. The mutation group contained at least one TOP gene mutation to compare the effects of TOP mutations on tumor recurrence (P-value=5.238e-3, P<0.05; q-value = 0.0489). The results suggest that the expression level of TOPs can use as a risk factor for assessing the Month to distant Recurrence of sarcoma (Figure 3B).

3.5. The relationship between the expression of TOPs in sarcoma and tumor subtypes

Soft-tissue sarcomas are malignant mesenchymal tumors, with an estimated 12,390 new sarcoma cases and 4990 deaths in the United States in 2017 and a 5-year OS rate of 64%(23). The completeness of resection, histologic subtype, and tumor grade are the main determinants of survival(24)(25). We considered the expression of TOPs in different sarcoma subtypes. A cumulative histogram of 100% multiple classifications (100% stacked bar graph) (Figure 4A) was made, and different volumes represented the percentage of gene expression in different subtypes. The chi-square test was used to assess the relationship between TOP expression and the six major sarcoma subtypes. The expression of TOPs was significantly different among the six main sarcoma subtypes. The results showed that TOP1 has the highest expression ratio in DDLPS, with a ratio of 31.82%. TOP1MT and TOP3B have the highest expression ratios in UPS, with 30% and 36.84% ratios, respectively. TOP2A and TOP3A have the highest expression ratios in STLMS, with 30% and 36.84% ratios, respectively. TOP2B has the highest expression in ULMS, with a ratio of 36.84%. There were considerable differences in the expression proportions

among these six sarcoma subtypes (P-value=8.540e-3, q-value=0.0368). The results showed that there were differences in the expression of TOPs between different subtypes. These results provide important insights into TOPs that can be used to diagnose different subtypes of sarcoma.

3.6. The relationships between the expression of TOPs and DNA methylation and miRNA clusters

Because the binary definition of hypermethylation is based entirely on the DNA methylation signal and DNA methylation profiles, sarcoma subtype information could not be separated at the DNA methylation level. Thus, we used the Partition Around Medoids (PAM) clustering method. The dendrogram from divisive hierarchical clustering revealed five main subsets of sarcomas. We integrated the methylation data and the mRNA expression data to identify the DNA methylation changes in different sarcoma subtypes and function-related genes (including potential tumor suppressor candidates). Previous studies have examined DNA methylation in soft-tissue sarcoma and based on genome-level DNA methylation studies. There are different patterns of DNA methylation in pediatric embryos and alveolar rhabdomyosarcoma. We found that Ewing's sarcoma euphemistically inactivates potential target genes. We further explored the correlation between TOPs and methylation, dividing data from TCGA into two groups based on TOP mutations' presence or absence. The mutation group contained TOP gene mutations to compare the effects of TOP mutations on DNA methylation. The results suggested that there are group differences between genes in separate clusters (P-value < 0.05).

Sarcomas were divided into distinct miRNA clusters, and different miRNA clusters were associated with DFS. The results showed that there were significant differences in TOP1, TOP2A, TOP2B, TOP3A, and TOP3B expression between cluster 1 and cluster 2 by the Wilcoxon signed-rank test.

3.7. Gene mutations of the TOP family and their relationship with OS and DFS in sarcoma (cBioPortal)

We next proceeded from the genome to disease intervention, rather than judging by tumor subtype. To better understand the relationship between TOP mutations and sarcoma, we elaborated and standardized RNA-Seq V2 data in the TCGA database. RNA-Seq V2 data in RSEM corresponded to RSEM genes from TCGA. The Z-score determined a threshold of ± 2.0 missense mutations. The results show that the TOP gene mutation rate was higher in sarcoma patients. The mutation rates were 11% for TOP1, 10% for TOP1MT, 10% for TOP2A, 9% for TOP2B, 28% for TOP3A and 16% for TOP3B (Figure 5A and 5B). High expression of TOPs in sarcoma was associated with shorter OS and DFS. We believe this finding may be related to TOP gene mutations.

3.8. Mutation and copy number variation analysis

There are somatic copy number variations of TOPs in sarcomas, and we examined the variation in the somatic copy numbers of the TOP genes in sarcoma. Molecular genetic testing has been suggested to be an efficient approach to additional testing because various types of sarcoma have specific genetic aberrations, including substitutions, deletions, amplifications, and single base pair translocations. Moreover, the accumulation of somatic mutations is one of the main reasons for tumor development and contributes to the expression of neoantigens(26). We sought to determine whether mutations in TOP genes are associated with sarcoma. We used a multivariate Cox proportional hazards model to correct clinical factors, including age, sex, race, and tumor stage, to show the clinical correlation between gene expression related to sarcoma and immune subgroups. We identified somatic mutations in sarcoma using the International Cancer Genome Consortium (ICGC) dataset and the TCGA dataset. We further investigated the association of gene mutations with prognosis. Transcription signatures associated with selected mutations were identified and analyzed in terms of immune cell infiltration and outcome. Finally, we examined whether gene mutations affect immunity. The results of this study may reveal novel biomarkers and suggest potential immunotherapy for patients with sarcoma.

We used the somatic copy number alteration (SCNA) module to obtain the copy number variations. The SCNA module provides an opportunity to compare the somatic copy number of a given gene with tumor invasion level between tumor subtypes. Sus is certified by GISTIC 2.0 and includes deep deletion (-2), arm-level deletion (-1), diploid/normal (0), arm-level gain (1), and high gain (2). Graphs were used to display each immune subset's distribution in each copy number of the selected cancer types. The two-sided Wilcoxon rank-sum test was used to compare each SCNA category's penetration level with the normal level. We found that TOP1 has highly amplified somatic copy number variations in neutrophils and dendritic cells and is an immunization promoter; TOP1MT has highly amplified somatic copy number variations in CD4+ T cell and neutrophil immune subsets; TOP2B has high amplification in the B cell immune subset, arm-level gain in the CD4+ T cell immune subset, and somatic copy number variation with an arm-level deletion in dendritic cells; TOP3A has arm-level and high gain amplification, the presence or absence of arm-level variations in neutrophils, and somatic copy number variations in CD4 + T cells and neutrophils immune subsets; TOP3B has high amplification in the immune subset of B cells, CD4+ T cells, CD8+ T cells, and dendritic cells, as well as arm-level somatic copy number variations in CD4 +T cells (Figure 5C).

3.9. Immune cell infiltration of TOPs in patients with sarcoma

Immunotherapy has become a promising treatment option for the treatment of solid tumors(27), but there is no research on the immune correlation between TOPs and sarcoma development. Our report describes the outcome and recruitment of TOPs and immune cells in sarcoma for the first time. We utilized the TIMER database to initiate a comprehensive study of the correlation between TOP and immune cell infiltration. Since GBM/OV microarray data contain more samples than RNA-seq data, we used GBM/OV microarray expression values to determine whether genes were available. By studying their

relationship with immune infiltration, we observed that TOP-expressing tumors have low purity in the entire sinusoidal and basic subtypes and upper and lower cancer populations (as described in the Materials and Methods section, tumor tissue immune population). We found that the high expression of TOP1 in sarcoma was negatively correlated with neutrophils (Cor = 0.1996, P = 0.001) and B cells (Cor = 0, 1611, P = 0.012). TOP1MT expression was negatively correlated with macrophages (Cor = -0.236, P = 0.0002) and neutrophils (Cor = -0.144, P = 0.024). TOP2A expression was negatively correlated with CD4+ T cell infiltration (Cor = -0.2322, P = 0.0002) and macrophages (Cor = -0.1713, P = 0.0084). Similarly, TOP2B expression was negatively correlated with CD4+ T cell infiltration (Cor = -0.2554, P = 6.48E-05), macrophages (Cor = -0.226594498, P = 0.00046395) and dendritic cells (Cor = -0.294179571, P = 3.38E-06). TOP3A expression was negatively correlated with CD4+ T cell infiltration (Cor = -0.2522, P = 8.07E-05). TOP3B expression was negatively correlated with macrophage infiltration (Cor = -0.14316, P = 0.028219) (Figure 6). We describe the TOP and immune cell recruitment as well as changes in the tumor microenvironment in sarcoma. The results showed that the transcriptional characteristics associated with TOP mutations are related to immune cell populations' infiltration into tumors. The high expression of TOPs in CD4+ T cells and neutrophils is positively correlated with sarcoma prognosis. In contrast, the high expression of TOPs in other immune cells is negatively correlated with sarcoma prognosis. The elevated expression of TOP1, TOP1MT, TOP2A, TOP2B, TOP3A, and TOP3B in CD4+ T cells and neutrophils increased the OS rate and improved the immune response (Figure 7). TOPs are associated with inflammatory responses and immune cell infiltration, influencing the clinical outcome of SARC patients. TOP mutations are a risk factor for prognosis and have a predictive effect on sarcoma prognosis. After considering age, sex, and race (Table 2), TOP mutations were considered significantly different based on age.

We explored the clinical relevance of one or more tumor immune subsets, with the flexibility to correct multiple covariates in a multivariable Cox proportional hazards model. The covariates included clinical factors (age, sex, ethnicity, and tumor stage) and gene expression. For the Cox model outputs, survival (cancer type)~variables was the formula of the user-defined Cox regression model. This model was fitted by the function in the R package 'survival.' Coef is the regression coefficient. HR is the hazard ratio. Its lower and upper 95% confidence intervals are denoted as 95%CI_l and 95%CI_u.

3.10. Association of the increased mRNA expression of TOPs with the improved prognosis of patients with sarcoma

The existing evidence suggests that TOPs would improve DFS in selected patients at high risk of recurrence. We used public datasets (2015 version; <http://kmplot.com/analysis/index.php?>) to analyze the correlation between the TOP mRNA levels and survival time of patients with sarcoma. The Kaplan-Meier curve and log-rank test showed that elevated TOP mRNA levels were significantly related to the OS and DFS times of all patients with sarcoma (P<0.05). Compared to normal tissues, high TOP expression

in sarcoma tissues was negatively correlated with the OS and DFS of patients with sarcoma. The results showed that the high expression of TOP1 (hazard ratio (HR) 2.11, confidence interval (CI) 1.41-3.17; $P = 2.3e-4$), TOP1MT (HR 1.83, CI 1.2-2.79; $P = 4.1e-3$), TOP2A (HR 2.2, CI 1.39-3.47; $P = 5.2e-4$), TOP2B (HR 2.11, CI 1.36-3.27; $P = 6.4e-4$), TOP3A (HR 1.62, CI 1.07-2.46; $P = 2.2e-2$) and TOP3B (HR 1.57, CI 1.05-2.33; $P = 2.5e-2$) was associated with lower OS. The high expression of TOP2A (HR 2.41, CI 1.38-4.2; $P = 1.5e-3$), TOP2B (HR 2.03, CI 1.2-3.44; $P = 6.9e-3$) and TOP3A (HR 1.76, CI 1.06-2.93; $P = 2.7e-2$) was correlated with lower DFS. We also divided the sarcoma patients into two groups: one group was defined as the TOP mutation group, and the other group was defined as the TOP non-mutation group. The survival curve analysis of OS and DFS using cBioPortal showed that when the TOP family is regarded as a whole, the increased expression of TOPs in sarcoma is related to lower OS, and the result is of scientific and statistical significance. Simultaneously, the medium-to-high expression of the TOP family in sarcoma is not completely correlated with lower DFS, and the result is statistically significant. The result shows that each TOP member plays a different role in the prediction of DFS. (additional file shows this in more detail [see Additional file 3])

3.11 Functional enrichment analysis of TOP co-expression modules

GSEA is a genomic expression profile micro class data analysis tool used to collect genes composed of numerous genes in an entire transcript modification group to perform a simple and straightforward correlation analysis(28). The functions of TOPs and the genes significantly related to TOP expression changes were predicted by performing GO and the KEGG analyses in the database. We analyzed the enriched biological processes (BPs), cellular components (CCs), and molecular functions (MFs) of TOPs as well as the main BPs regulated by the signing of the TOPs, classified by the combination of the results. We found that GO:0006265 (DNA topological change), GO:0007059 (chromosome segregation), and GO:0006260 (DNA replication) (Figure 9A, 9B) were significantly different, and TOP alterations in sarcoma significantly regulated these changes. (additional file shows this in more detail [see Additional file 1])

3.12 Co-expressed genes of TOPs (Metascape)

The co-expressed genes of TOP1, TOP1MT, TOP2A, TOP2B, TOP3A, and TOP3B were tested by Metascape (Figure 10A and Table 3) and network topology-based analysis (Figure 10B). We set the top-ranking neighbors as 10 to identify the adjacent genes associated with TOPs with differential expression(additional file shows this in more detail [see Additional file 2]).

3.13. Predicted cellular functions and pathways of TOP genes and TOP-related neighboring genes in sarcoma

A network of TOP mutations and their 120 frequently changed neighboring genes was created. The results showed that BUB1B, CENPF, SPC25, NUF2, KNL1SGO1, RPS20, SNRPD3, and MED9 were significantly associated TOP mutations (Figure 11A and 11B). GO functional enrichment analysis predicted the three main mutation functions of TOPs and their 120 frequently changed neighboring genes, including BPs, CCs, and MFs. KEGG pathway analysis revealed the pathways of TOPs and their 120 most frequently changed neighboring genes (Figure 11C).

3.14. Protein interaction network of TOPs (Metascape)

For a given list of TOP genes, an enrichment analysis of protein-protein interactions was performed using the following databases: BioGrid, InWeb_IM, and OmniPath. The resulting network contains a subset of proteins with physical interaction with at least one additional member of the list(29)(30)(31)(32)(33)(34). GO and KEGG results were obtained and showed that in sarcoma, the neighboring proteins of TOPs are mainly expressed on the GO: 0000775 chromosomes, the centromeric region (log10P: -12.7); GO: 0005819 spindles (log10P: -12.0); and GO: 0000819 sister (log10P: -11.2). There were four protein clusters with apparent interaction relationships (Table 4)(35)(36)(additional file shows this in more detail [see Additional file 4]).

4. Discussion

There are six topoisomerases (TOP1, TOP1MT, TOP2 α , TOP2 β , TOP3 α , and TOP3 β) in human cells, which mainly operate in a wide range of DNA and RNA substrates in the nuclear and mitochondrial genomes. Previous studies have shown that their catalytic intermediate DNA topoisomerase cleavage complex (TOPcc) is a therapeutic target for numerous anticancer drugs(37). Eukaryotic DNA TOPs play a significant role in transcription, replication, genome stability, and previous studies focused on TOPs in cancer, immune disorders, and the nervous system's diseases have shown that they have therapeutic relevance(38). Although the role of TOPs in the onset and prognosis of various cancers has been partially confirmed, there is no additional bioinformatics analysis on the relationship between TOPs and the development of sarcoma. Since transcriptase sequencing and other methods can help identify molecular events, these molecular events may help in the differential diagnosis of sarcomas. In the past year, research teams have reviewed the influence of molecular detection on histological diagnosis internationally, showing that molecular marker detection can provide a more diagnostic basis. This study took the lead in studying the mRNA expression of TOPs in sarcoma. It explored the relationships between TOPs and gene mutation, immune infiltration, sarcoma typing, DNA methylation, miRNA grouping, and prognosis (OS and DFS). We hope our findings will help enrich existing knowledge, improve existing treatment designs, and improve patients' prognostic accuracy with sarcoma.

Studies have shown that the lack of TOP activity or TOP mutations is related to neurodegenerative genomic instability syndrome and autism spectrum disorder(39). Top1, as a molecular target of SN38, is overexpressed in 43% to 51% of colon cancers(40)(41). Elevated TOP1 expression was found to be

associated with a low OS rate in colon cancer. According to the literature, TOP1 inhibitors are associated with a good prognosis in sarcoma. After clarifying other cellular and genetic factors related to these drugs' susceptibility, it can be reasonably predicted that TOP inhibitors could treat at least some tumors with specific chromosomal changes in the cell cycle and DNA repair genes. These drugs are actively involved and have an antitumor effect. TOP2A is a marker related to the prognosis of STS proliferation, and its expression is related to the poor prognosis of peripheral nerve sheath malignancies and SS(42). As expected, LASTS with high expression of Top2 α has a higher recurrence rate, and the annual risk of recurrence increases as the percentage of Top2 α -positive cells increases(43). In liver cancer, the increase in TOP1 and TOP2A expression levels are associated with a worse OS rate and is closely related to the DFS rate(44). Research reports indicate that the lack of TOP1MT dramatically reduces the proliferation and regeneration of liver cells(45). Besides, TOP1MT deficiency can impair tumor growth in colon cancer and liver cancer in human and mouse models. The importance of TOP1MT in tumor development provides a theoretical basis for the development of anticancer drugs targeting TOP1MT. Research has also revealed that androgen signaling promotes the co-recruitment of androgen receptor and topoisomerase II beta (TOP2B)(46). TOP2B is associated with distinctive genome rearrangements and copy number changes in prostate cancer(47). Little is known about the expression and role of TOP1MT, TOP3A, and TOP3B in sarcoma. This study found that the expression of TOP1, TOP1MT, TOP2A, TOP2B, TOP3A, and TOP3B in sarcoma was higher than that in normal tissues.

The FNCLCC scoring system can reflect the prognosis of patients with sarcoma and the metastasis and death of cancer patients. This study showed that in sarcoma, TOP1, TOP2A, and TOP3B were significantly altered in patients with FNCLCC grades 2 and 3. There were noteworthy differences in TOP1 and TOP2A expression between FNCLCC grades 1 and 3. Regarding the relationship between the local recurrence of sarcoma and TOPs, studies have shown that the most critical factor affecting local recurrence is the resectability of recurrent disease. The patients most likely to benefit from resection are disease-free and have a longer course, no primary tumor rupture, low-grade tumors, and unidirectional recurrence. The results showed that the increase in TOP mutations caused an increase in distant tumors' recurrence rate. Therefore, it is believed that the expression level of TOPs can be used as a risk factor for predicting the recurrence of local sarcoma.

We used the chi-square test to explore the expression of TOPs in different sarcoma subtypes and determine the relationship between the TOP expression level and the six main sarcoma subtypes. We found that among the six main sarcoma subtypes, TOP1, TOP1MT, TOP2A, TOP2B, TOP3A, and TOP3B all had expression differences, and different TOP members had altered expression levels in different subtypes. We also explored the relationship between the DNA methylation of TOPs and miRNA typing in sarcoma. The results showed that TOPs were expressed differently in diverse DNA methylation clusters and miRNA clusters in sarcoma. The relative rarity of each subtype of sarcoma and the potential genetic differences between sarcomas can complicate the evaluation of the efficacy of any drug⁶¹. Some researchers believe that the treatment of sarcoma should be based on genotype rather than subtype(48). Cambodia et al. (49) combined SNP sequences with global and target exon sequencing to characterize the genome patterns of 86 liposarcomas in all major subtypes.

It should be noted that the Food and Drug Administration (FDA) has approved treatments for metastatic liposarcoma, and this study focused on three different subtypes (dedifferentiated, mucinous and pleomorphic liposarcoma) of liposarcoma. The prognosis of these subtypes is the same. There are still challenges in developing targeted therapies for sarcomas because drugs target secondary mutations rather than triggers, and sarcoma genome instability can promote drug resistance. In this sense, even though many sarcomas are similar to other cancers, at least in some sarcoma subtypes, the sarcoma gene's triggering mutation is more prominent, and the genetics are more straightforward than other tumors. In the future, sarcoma diagnostic strategies should pay more attention to genomic changes rather than sarcoma subtypes. To this end, the study described the somatic mutations of TOPs in sarcomas, explored the clinical correlation between sarcoma-related gene mutations and immune subgroups, and determined that TOP gene mutations are associated with higher inflammation and immune cell infiltration, thus affecting the poor prognosis of SARC patients. These results suggest that TOP mutations are a negative risk factor for prognosis.

Moreover, the functions and pathways of TOPs and their 120 frequently changed adjacent gene mutations in sarcoma patients were explored. The results showed that DNA repair and mitochondrial fusion genes were closely linked to TOP mutations. Through multivariate analysis, we have reason to believe that the abnormal expression of TOP1, TOP1MT, TOP2A, TOP2B, TOP3A and TOP3B in sarcoma is an independent prognostic factor that affects OS and DFS times of sarcoma patients. The article's shortcoming is the lack of sufficient clinical samples and unable to perform immunohistochemical detection of TOPs on clinical samples and tumor formation experiments in animals.

5. Conclusions

Through Oncomine, TCGA, and other databases, we found that TOPs are highly expressed in sarcoma tissues at the transcriptome level compared to normal tissues. As an international standard for determining the histological grade of sarcoma, the FNCLCC system can better predict the distant metastasis and prognosis of cancer patients. We studied the relationship between TOPs and FNCLCC grades of sarcoma and found that TOP1/TOP2A/TOP2B/TOP3B can be used to predict different sarcoma grades. Local recurrence harms the long-term survival of patients with malignant tumors. The study results suggest that TOPs can be used as a risk factor for assessing sarcoma's local recurrence. There are differences in the expression of TOPs in different sarcoma subtypes, DNA methylation clusters, and miRNA clusters. We also studied the relationships between TOP gene mutations and immune infiltration, copy number variations, and survival curves. The results showed that in sarcoma, there are differences in the immune infiltration and somatic copy numbers of different TOPs, which indicates different prognoses. The above studies predict that TOPs can be used as a therapeutic target for the treatment of sarcoma. Therefore, we further performed enrichment analysis to understand the enriched BPs, CCs, and MFs of TOPs and discussed TOP-like co-expressed genes and the protein interaction network of these genes.

Abbreviations

TOPs, DNA topoisomerase; STS, Soft-tissue sarcoma; DNA Topoisomerase I; TOP1MT, DNA Topoisomerase I Mitochondrial; TOP2 α , DNA Topoisomerase II Alpha; TOP2 β , DNA Topoisomerase II Beta; TOP3 α , DNA Topoisomerase III Alpha; TOP3 β , DNA Topoisomerase III Beta; FNCLCC, Fédération Nationale des Centres de Lutte Contre le Cancer; SNP, single nucleotide polymorphisms;DDLPS, dedifferentiated liposarcoma; LMS, leiomyosarcoma; UPS, undifferentiated pleomorphic sarcoma; MFS, mucin fibroblast sarcoma malignant peripheral nerve sheath tumor MPNST; SS, synovial sarcoma; ULMS, uterine leiomyosarcoma; PD, pharmacologic biomarker; TCGA, The Cancer Genome Atlas; GO, Gene Ontology; GEPIA, Gene Expression Profiling Interactive Analysis; DFS, disease-free survival; OS, overall survival; TIMER, Tumor Immune Estimation Resource; IHC, Immunohistochemistry; CNAs, copy number alterations; RPPA, reversed-phase protein array; KEGG, Kyoto Encyclopedia of Genes and Genomes; DAVID, Database for Annotation, Visualization and Integrated Discovery; PAM, Partition Around Medoids; ICGC, International Cancer Genome Consortium; SCNA, somatic copy number alteration; BP, biological processes, CC, cellular components; MF, molecular functions; NTA, Network Traffic Analysis;

Declarations

Acknowledgement:

We thank "Springer Nature Author Services" for help in editing grammar, spelling, and other writing errors.

Funding information

This research was funded by the National Natural Science Foundation of China (Grant Nos. 81874020);

Author information

Affiliations

Department of Orthopedics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China.

Genchun wang, Kai Sun, Jiamin Lin, Xudong Yao, Jiachao Guo, Zhou Guo, Liang Cai Hou, jingTing Xu, and Fengjing Guo

Department of Obstetrics and Gynecology Department, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China.

Li Ting Liu

Contributions

GW composed the manuscript and proposed the experimental design. GW, KS, and LL contributed to data interpretation and bioinformatic analysis. JL, XY, JG, ZG, LH, JH contributed with data collection. The final manuscript was reviewed and approved by all authors before submission.

ethics approval and consent to participate

not applicable

Consent for publication

Each author approved the manuscript before submission for publication.

conflict of interest

The authors declare no conflict of interest

Availability of data and material

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

References

1. Delgado JL, Hsieh CM, Chan NL, Hiasa H. Topoisomerases as anticancer targets. *Biochem J* 2018; 475(2): 373-98
2. Brown PO, Cozzarelli NR. A sign inversion mechanism for enzymatic supercoiling of DNA. *Science (New York, NY)* 1979; 206(4422): 1081-3
3. Dancsok AR, Asleh-Aburaya K, Nielsen TO. Advances in sarcoma diagnostics and treatment. *Oncotarget* 2017; 8(4): 7068-93.
4. Baechler SA, Factor VM, Dalla Rosa I et al. The mitochondrial type IB topoisomerase drives mitochondrial translation and carcinogenesis. *Nature Communications* 2019; 10.
5. Kwan KY, Wang JC. Mice lacking DNA topoisomerase IIIbeta develop to maturity but show a reduced mean lifespan. *Proc Natl Acad Sci U S A.* 2001;98(10):5717-5721.

6. Oliveira-Costa JP, Zanetti J, Oliveira LR, Soares FA, Ramalho LZ, Silva Ramalho F, Garcia SB, Ribeiro-Silva A. Significance of topoisomerase III β expression in breast ductal carcinomas: strong associations with disease-specific survival and metastasis. *Hum Pathol.* 2010 Nov;41(11):1624-30.
7. Adamowicz M, Radlwimmer B, Rieker RJ, et al. Frequent amplifications and abundant expression of TRIO, NKD2, and IRX2 in soft tissue sarcomas. *Genes Chromosomes & Cancer* 2006; 45(9): 829-38
8. Francis P, Namlos HM, Muller C, et al. Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential. *BMC Genomics* 2007; 8: 73
9. Renner M, Czwan E, Hartmann W, et al. MicroRNA profiling of primary high-grade soft tissue sarcomas. *Genes Chromosomes & Cancer* 2012; 51(11): 982-96.
10. Subramanian S, Lui WO, Lee CH, et al. MicroRNA expression signature of human sarcomas. *Oncogene* 2008; 27(14): 2015-26.
11. Ugras S, Brill E, Jacobsen A, et al. Small RNA Sequencing and Functional Characterization Reveals MicroRNA-143 Tumor Suppressor Activity in Liposarcoma. *Cancer Research* 2011; 71(17): 5659-69.
12. 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.
13. Frid MG, McKeon BA, Thurman JM, et al. Immunoglobulin-driven Complement Activation Regulates Proinflammatory Remodeling in Pulmonary Hypertension. *Am J Respir Crit Care Med.* 2020;201(2):224-239. doi:10.1164/rccm.201903-0591OC
14. Cai P, Zheng H, She J, et al. Molecular Mechanism of Aflatoxin-Induced Hepatocellular Carcinoma Derived from a Bioinformatics Analysis. *Toxins (Basel)* 2020; 12(3).
15. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; 490(7418): 61-70.
16. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019 Apr 3;10(1):1523.
17. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019; 10(1): 1523.
18. Neuville A, Chibon F, Coindre JM. Grading of soft tissue sarcomas: from histological to molecular assessment. *Pathology* 2014; 46(2): 113-20.

19. Bramwell V, Rouesse J, Steward W, Santoro A, Schraffordt-Koops H, Buesa J, Ruka W, Priario J, Wagener T, Burgers M, et al. Adjuvant CYVADIC chemotherapy for adult soft tissue sarcoma—reduced local recurrence but no improvement in survival: a study of the European Organization for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group. *J Clin Oncol*. 1994 Jun;12(6):1137-49.
20. Mendenhall WM, Indelicato DJ, Scarborough MT, et al. The management of adult soft tissue sarcomas. *American journal of clinical oncology* 2009; 32(4): 436-42.
21. Mahvi DA, Liu R, Grinstaff MW, Colson YL, Raut CP. Local Cancer Recurrence: The Realities, Challenges, and Opportunities for New Therapies. *CA Cancer J Clin* 2018; 68(6): 488-505.
22. Gronchi A, Strauss DC, Miceli R, et al. Variability in Patterns of Recurrence After Resection of Primary Retroperitoneal Sarcoma (RPS): A Report on 1007 Patients From the Multi-institutional Collaborative RPS Working Group. *Annals of surgery* 2016; 263(5): 1002-9.
23. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin*. 2017 Jan;67(1):7-30.
24. Bonvalot S, Rivoire M, Castaing M, et al. Primary retroperitoneal sarcomas: a multivariate analysis of surgical factors associated with local control. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2009; 27(1): 31-7.
25. Singer S, Antonescu CR, Riedel E, Brennan MF. Histologic subtype and margin of resection predict pattern of recurrence and survival for retroperitoneal liposarcoma. *Annals of surgery* 2003; 238(3): 358-70; discussion 70-1.
26. Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol*. 2013 Apr;13(4):227-42.
27. Luke JJ, Flaherty KT, Ribas A, Long GV. Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol*. 2017;14(8):463-482.
28. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health*. 1999;20:145-57.
29. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019 Apr 3;10(1):1523.
30. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990 Jul;9(7):811-8.
31. Roberts C. Modelling patterns of agreement for nominal scales. *Stat Med*. 2008 Mar 15;27(6):810-30.

32. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowitz G, Workman CT, Rigina O, Rapacki K, Stærfeldt HH, Brunak S, Jensen TS, Lage K. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods*. 2017 Jan;14(1):61-64.
33. Zhou YY, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* 2019; 10.
34. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003 Jan 13;4:2.
35. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker C. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003 Nov;13(11):2498-504.
36. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*. 2006 Jan 1;34(Database issue):D535-9.
37. Pommier Y, Sun Y, Huang SN, Nitiss JL. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol*. 2016 Nov;17(11):703-721.
38. Nitiss JL. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer*. 2009 May;9(5):338-50.
39. Katyal S, Lee Y, Nitiss KC, Downing SM, Li Y, Shimada M, Zhao J, Russell HR, Petrini JH, Nitiss JL, McKinnon PJ. Aberrant topoisomerase-1 DNA lesions are pathogenic in neurodegenerative genome instability syndromes. *Nat Neurosci*. 2014 Jun;17(6):813-21.
40. Boonsong A, Curran S, McKay JA, Cassidy J, Murray GI, McLeod HL. Topoisomerase I protein expression in primary colorectal cancer and lymph node metastases. *Hum Pathol*. 2002 Nov;33(11):1114-9.
41. Paradiso A, Xu J, Mangia A, Chiriatti A, Simone G, Zito A, Montemurro S, Giuliani F, Maiello E, Colucci G. Topoisomerase-I, thymidylate synthase primary tumour expression and clinical efficacy of 5-FU/CPT-11 chemotherapy in advanced colorectal cancer patients. *Int J Cancer*. 2004 Aug 20;111(2):252-8.
42. Paik S, Bryant J, Tan-Chiu E, Yothers G, Park C, Wickerham DL, Wolmark N. HER2 and choice of adjuvant chemotherapy for invasive breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-15. *J Natl Cancer Inst*. 2000 Dec 20;92(24):1991-8.
43. Skotheim RI, Kallioniemi A, Bjerkhagen B, Mertens F, Brekke HR, Monni O, Mousses S, Mandahl N, Soeter G, Nesland JM, Smeland S, Kallioniemi OP, Lothe RA. Topoisomerase-II alpha is upregulated in malignant peripheral nerve sheath tumors and associated with clinical outcome. *J Clin Oncol*. 2003 Dec 15;21(24):4586-91.

44. Liu LM, Xiong DD, Lin P, Yang H, Dang YW, Chen G. DNA topoisomerase 1 and 2A function as oncogenes in liver cancer and may be direct targets of nitidine chloride. *Int J Oncol*. 2018 Nov;53(5):1897-1912.
45. Khiati S, Baechler SA, Factor VM, Zhang H, Huang SY, Dalla Rosa I, Sourbier C, Neckers L, Thorgeirsson SS, Pommier Y. Lack of mitochondrial topoisomerase I (TOP1mt) impairs liver regeneration. *Proc Natl Acad Sci U S A*. 2015 Sep 8;112(36):11282-7.
46. Baechler SA, Factor VM, Dalla Rosa I, Ravji A, Becker D, Khiati S, Miller Jenkins LM, Lang M, Sourbier C, Michaels SA, Neckers LM, Zhang HL, Spinazzola A, Huang SN, Marquardt JU, Pommier The mitochondrial type IB topoisomerase drives mitochondrial translation and carcinogenesis. *Nat Commun*. 2019 Jan 8;10(1):83.
47. Haffner MC, Aryee MJ, Toubaji A, Esopi DM, Albadine R, Gurel B, Isaacs WB, Bova GS, Liu W, Xu J, Meeker AK, Netto G, De Marzo AM, Nelson WG, Yegnasubramanian S. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nat Genet*. 2010 Aug;42(8):668-75.
48. Lim J, Poulin NM, Nielsen TO. New Strategies in Sarcoma: Linking Genomic and Immunotherapy Approaches to Molecular Subtype. *Clin Cancer Res*. 2015 Nov 1;21(21):4753-9. doi: 10.1158/1078-0432.CCR-15-0831. Epub 2015 Sep 1. PMID: 26330427.
49. Kanojia D, Nagata Y, Garg M, et al. Genomic landscape of liposarcoma. *Oncotarget*. 2015;6(40):42429-42444.

Tables

Table 1

Significant changes in TOP transcription levels between different types of sarcoma and normal tissues.
*P<0.05.

	Type of Sarcoma versus Normal Tissue	Fold Change	P-value	t-test	Source and/or Reference
TOP1	Sarcoma vs. Normal	2.164	0.000457	3.895	Detwiller Sarcoma Statistics
TOP1MT	NA	NA	NA	NA	NA
TOP2A	Myxofibrosarcoma vs. Normal	12.264	3.47E-24	24.318	Barretina Sarcoma Statistics
	Pleomorphic Liposarcoma vs. Normal	9.788	3.33E-16	18.002	Barretina Sarcoma Statistics
	Dedifferentiated Liposarcoma vs. Normal	7.746	4.15E-19	14.12	Barretina Sarcoma Statistics
	Leiomyosarcoma vs. Normal	12.112	1.92E-12	11.088	Barretina Sarcoma Statistics
	Myxoid/Round Cell Liposarcoma vs. Normal	4.104	2.79E-11	11.745	Barretina Sarcoma Statistics
	Fibrosarcoma vs. Normal	27.421	1.14E-08	8.906	Detwiller Sarcoma Statistics
	Pleomorphic Liposarcoma vs. Normal	17.927	5.88E-08	9.384	Detwiller Sarcoma Statistics
	Leiomyosarcoma vs. Normal	48.841	3.46E-08	8.485	Detwiller Sarcoma Statistics
	Malignant Fibrous Histiocytoma vs. Normal	21.095	1.79E-08	8.258	Detwiller Sarcoma Statistics
	Round Cell Liposarcoma vs. Normal	13.321	1.91E-07	8.171	Detwiller Sarcoma Statistics
	Synovial Sarcoma vs. Normal	26.646	2.55 E-06	6.884	Detwiller Sarcoma Statistics
	Uterine Corpus Leiomyosarcoma vs. Normal	3.894	7 E-03	2.898	Quade Uterus Statistics

TOP2B	Dedifferentiated Liposarcoma vs. Normal	2.002	1.02 E-04	4.767	Detwiller Sarcoma Statistics
	Fibrosarcoma vs. Normal	2.938	2.24 E-06	6.429	Detwiller Sarcoma Statistics
	Round Cell Liposarcoma vs. Normal	2.491	5.58 E-05	5.581	Detwiller Sarcoma Statistics
	Synovial Sarcoma vs. Normal	3.108	1.22 E-04	6.033	Detwiller Sarcoma Statistics
	Myxoid/Round Cell Liposarcoma vs. Normal	2.318	1.25E-10	10.793	Barretina Sarcoma Statistics
	Dedifferentiated Liposarcoma vs. Normal	1.65	4.71E-07	7.352	Barretina Sarcoma Statistics
	Pleomorphic Liposarcoma vs. Normal	1.505	1.41 E-05	5.151	Barretina Sarcoma Statistics
TOP3A	Myxofibrosarcoma vs. Normal	1.758	7.09E-11	10.087	Barretina Sarcoma Statistics
	Pleomorphic Liposarcoma vs. Normal	1.676	3.19E-08	7.126	Barretina Sarcoma Statistics
	Leiomyosarcoma vs. Normal	1.757	1.75E-09	8.012	Barretina Sarcoma Statistics
TOP3B	NA	NA	NA	NA	NA
NA, not available; TCGA, The Cancer Genome Atlas.					

Table 2
Cox proportional hazards model. *P<0.05.

	COEF	HR	95%CI_l	95%CI_u	p-value	sig
Age	0.022	1.022	1.006	1.039	0.006	**
Gender male	0.091	1.095	0.693	1.732	0.697	
Race black	0.229	1.258	0.144	10.969	0.836	
Race white	-0.075	0.927	0.121	7.077	0.942	
Purity	0.713	0.713	0.632	6.589	0.233	
TOP1	0.1	1.105	0.767	1.591	0.591	
TOP1MT	0.227	1.255	0.951	1.657	0.109	
TOP2A	0.075	1.078	0.879	1.322	0.472	
TOP2B	0.37	1.448	0.888	2.361	0.138	
TOP3A	-0.043	0.957	0.702	1.306	0.784	
TOP3B	-0.104	0.901	0.569	1.425	0.655	

Table 3
Functional enrichment analysis of TOP co-expression modules. *P-value < 0.05.

Network	Annotation
Input ID	GO:0006265 DNA topological change -17.7 GO:0071103 DNA conformation change -9.4; R-HSA-4615885 SUMOylation of DNA replication proteins -7.2
Input ID_MCODE_ALL	GO:0006265 DNA topological change -14.1; R-HSA-4615885 SUMOylation of DNA replication proteins -7.6; GO:0071103 DNA conformation change -7.5
Input ID_SUB1_MCODE_1	GO:0006265 DNA topological change -14.1; R-HSA-4615885 SUMOylation of DNA replication proteins -7.6; GO:0071103 DNA conformation change -7.5
Network construction method: network expansion; set number of top-ranking neighbors: 10; *P-value < 0.05	

Table 4
 Protein interaction network of TOP genes. *P-value < 0.05.

Network	Annotation
Input ID	GO:0000775 chromosome, centromeric region -12.7;GO:0005819 spindle -12.0;GO:0000819 sister chromatid segregation -11.2
Input ID_MCODE_ALL	GO:0000777 condensed chromosome kinetochore -9.6;GO:0000779 condensed chromosome, centromeric region -9.3;GO:0000776 kinetochore -9.0
Input ID_SUB1_MCODE_1	GO:0000777 condensed chromosome kinetochore -14.2;GO:0000779 condensed chromosome, centromeric region -13.9;GO:0000776 kinetochore -13.5
Input ID_SUB1_MCODE_2	GO:0006413 translational initiation -12.6;GO:0006412 translation -9.2;GO:0043043 peptide biosynthetic process -9.1
Input ID_SUB1_MCODE_3	GO:0000398 mRNA splicing, via spliceosome -9.2;GO:0000377 RNA splicing, via transesterification reactions with bulged adenosine as nucleophile -9.2; GO:0000375 RNA splicing, via transesterification reactions -9.2
Input ID_SUB1_MCODE_4	GO:0003712 transcription coregulator activity -4.8

Figures

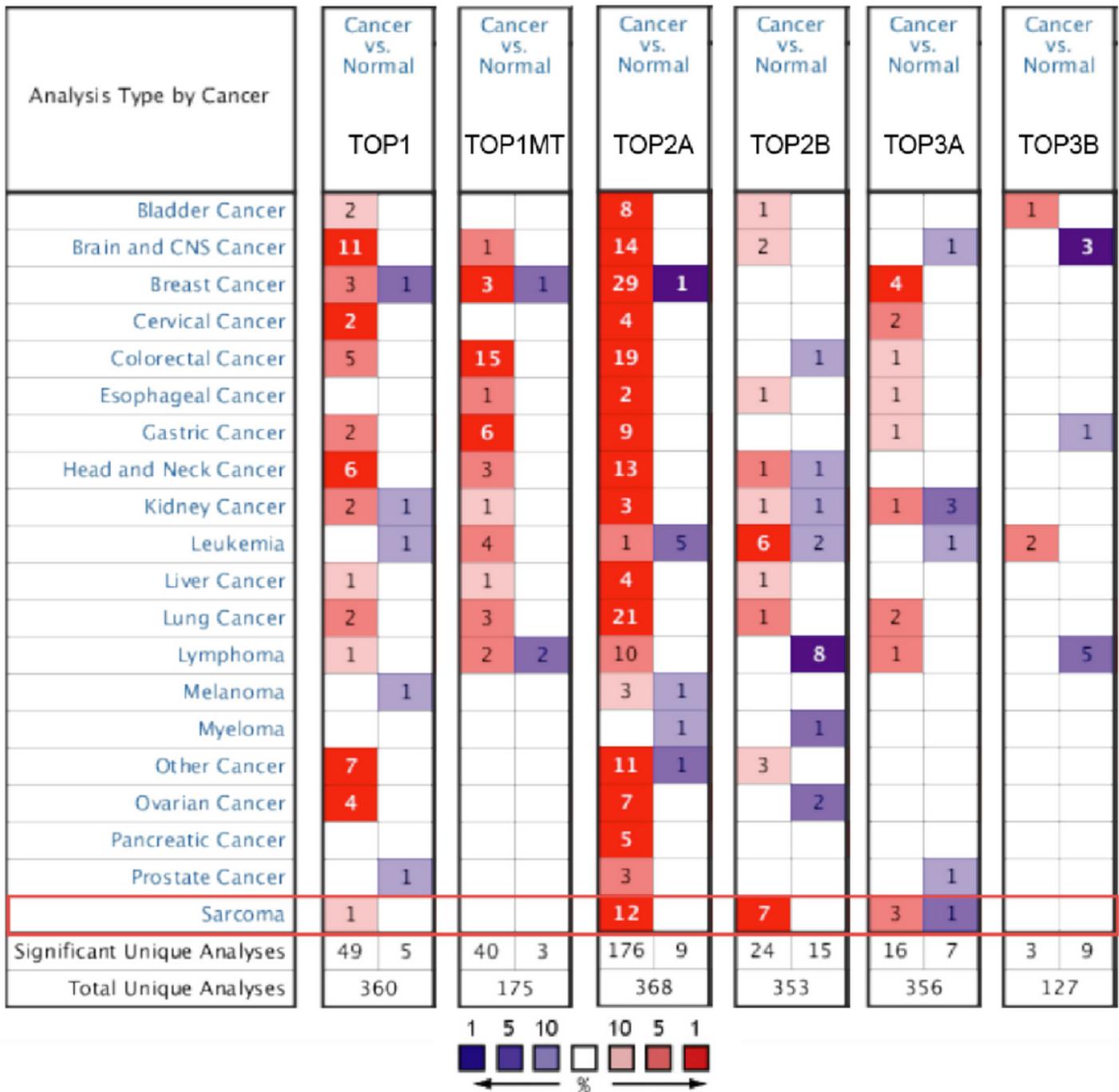


Figure 1

Transcriptional expression of TOPs in 20 different types of cancer diseases (Oncomine database). The figure shows the current research status of TOP family members in different tumors. *P<0.01, fold change=1.5. Note: red represents upregulation of the target gene, while blue represents downregulation of the target gene.

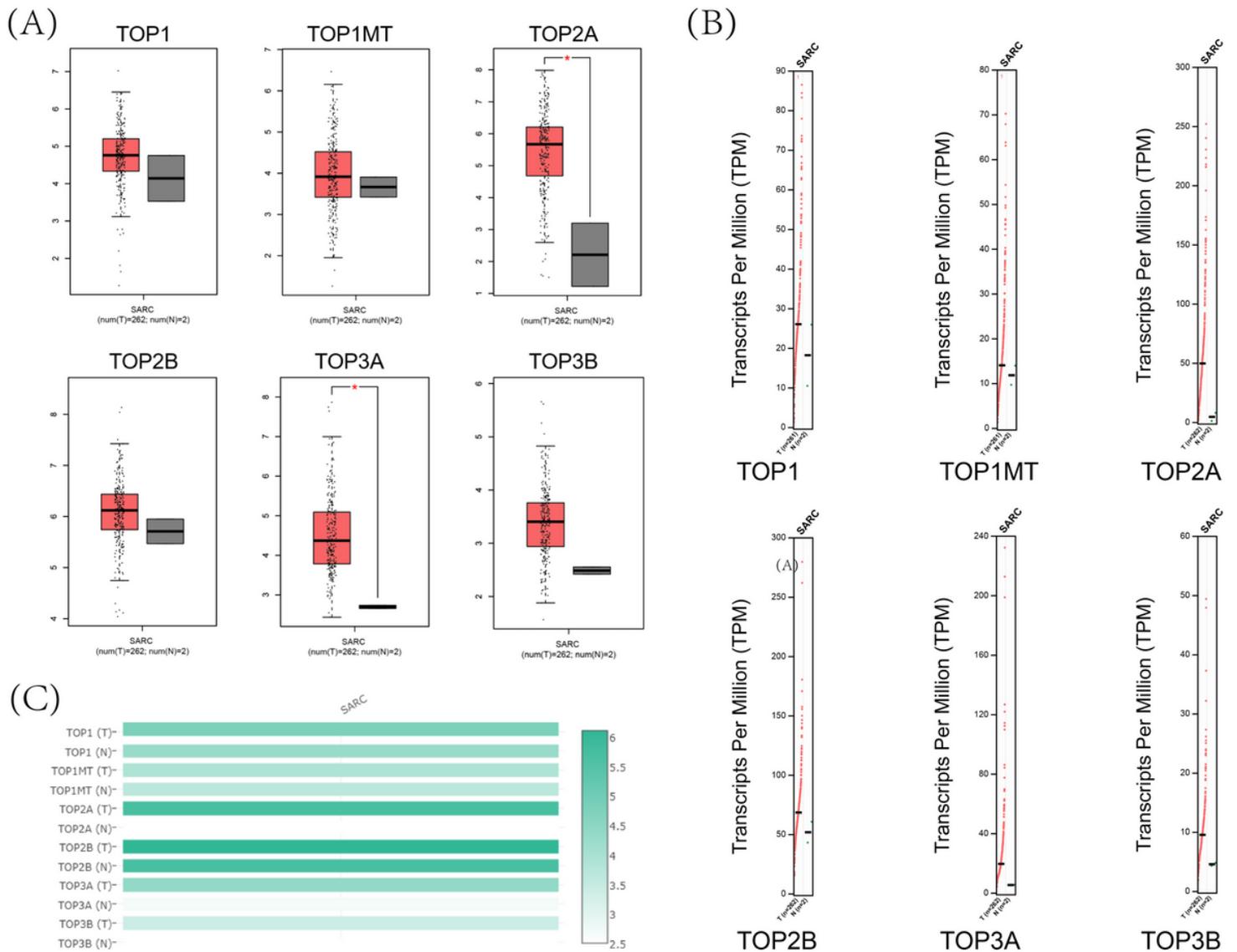


Figure 2

(A), (B) Differences in the expression of TOPs between sarcoma and normal tissues. The expression of TOPs in sarcoma (box plot; scatter diagram); methods: ANOVA; $|\log_2FC|$ cutoff: 1; q-value cutoff: 0.05. (C) The expression of TOPs in sarcoma (GEPIA). Multiple gene comparison: The expression of TOP1/1MT/2A/2B/3A/3B in sarcoma (GEPIA). * $P < 0.001$. TOP1(T)-/TOP1(N)-=4.8/4.3; TOP1MT(T)-/TOP1MT(N)-=3.9/3.7; TOP2A(T)-/TOP2A(N)-=5.7/2.5; TOP2B(T)-/TOP2B(N)-=6.1/5.7; TOP3A(T)-/TOP3A(N)-=4.4/2.7; TOP3B(T)-/TOP3B(N)-=3.4/2.5. Expression matrix diagrams based on a specific list of genes. The color density in each block represents the mean expression level of TOPs in sarcoma tissue normalized by the maximum mean expression level in all blocks. Different genes in the same tumors or normal tissues can be compared on a graph. $\log_2(TPM + 1)$ was used for the logarithmic scale. The TCGA and GTEx data were normalized. Note: TOPs(T)-: TOPs in sarcoma; TOPs(N)-: TOPs in normal tissue.

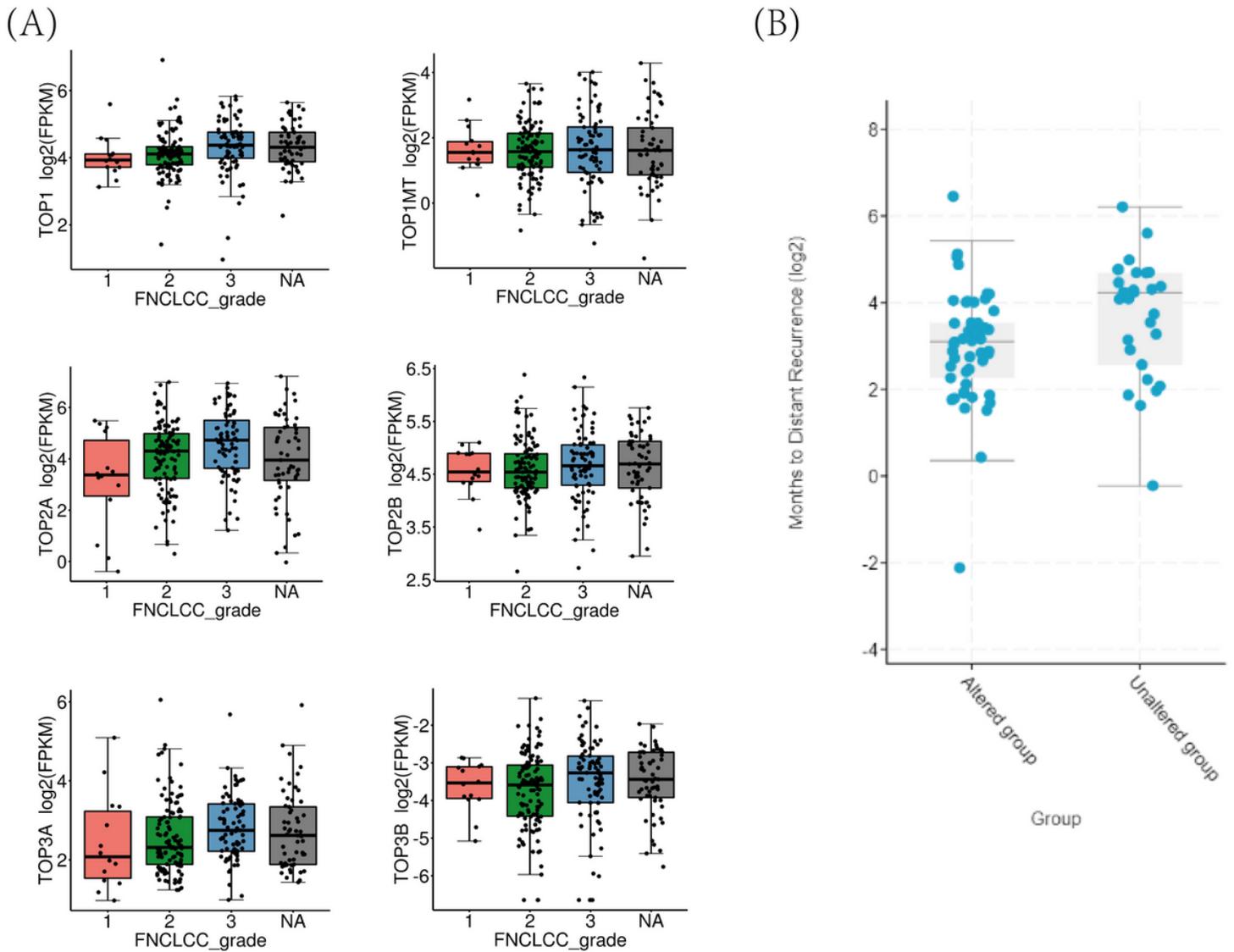


Figure 3

(A) mRNA expression levels of TOPs in different FNCLCC grades correlated significantly with the sarcoma individual tumor grade. Testing method: Wilcoxon signed-rank test; *P-value < 0.05. (B) Correlation between TOP mutations and distant tumor recurrence. For the statistical test method, continuous variables with a nonnormal distribution or ordered classification variables were used. *P-value < 0.05; *q-value < 0.05.

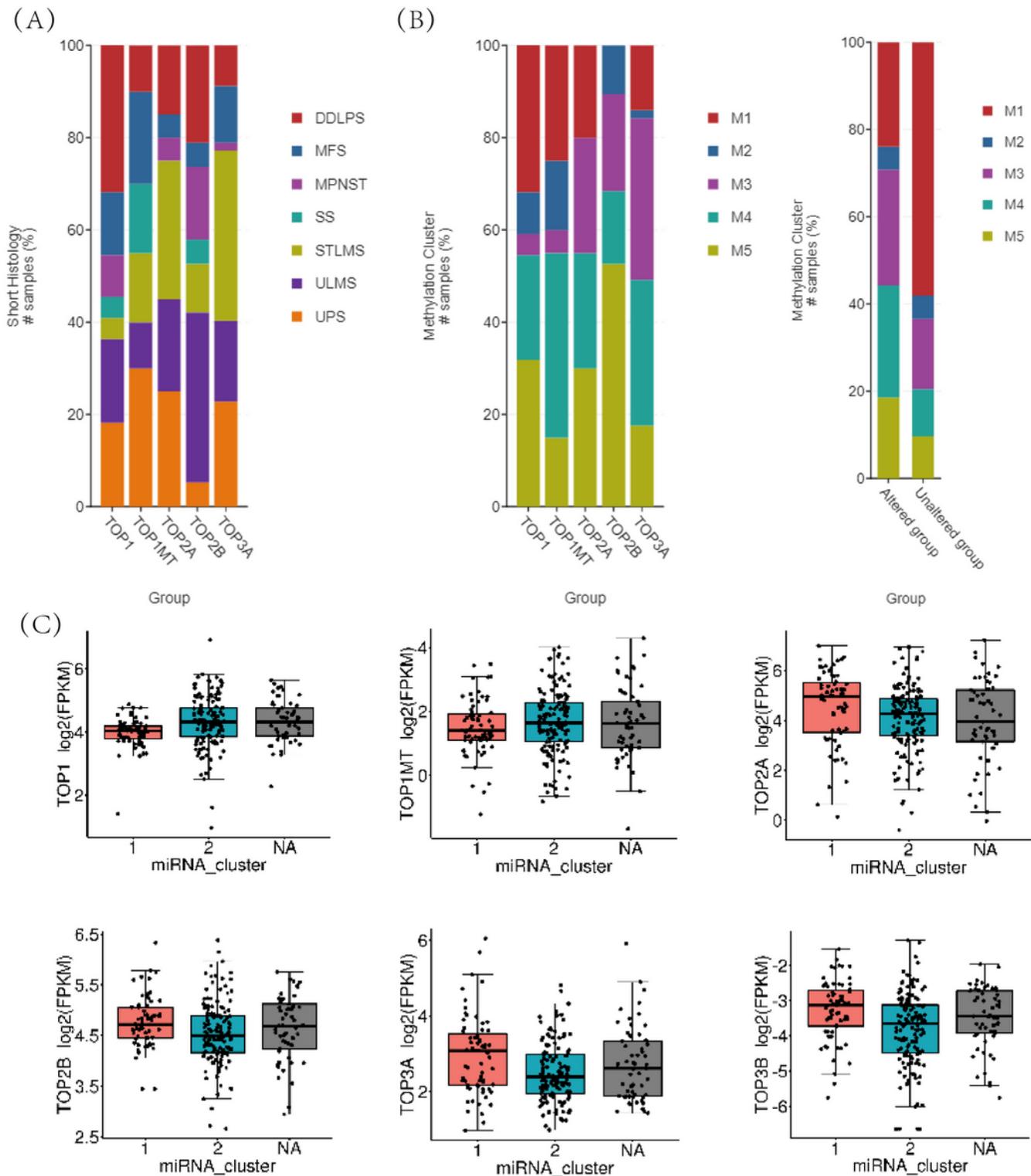


Figure 4

(A) The relationship between the expression of TOPs and sarcoma subtypes. Statistical test method: chi-squared test, Kruskal-Wallis test; *P-value < 0.05. (B) The relationship between the expression of TOPs and DNA methylation. Statistical test method: Wilcoxon signed-rank test; *P-value < 0.05; PAM clustering methods: the dendrogram obtained by divisive hierarchical clustering revealed five main sarcoma

subtypes. (C) The relationship between the expression of TOPs and the miRNA cluster. Statistical Test method: Wilcoxon signed-rank test; *P-value < 0.05.

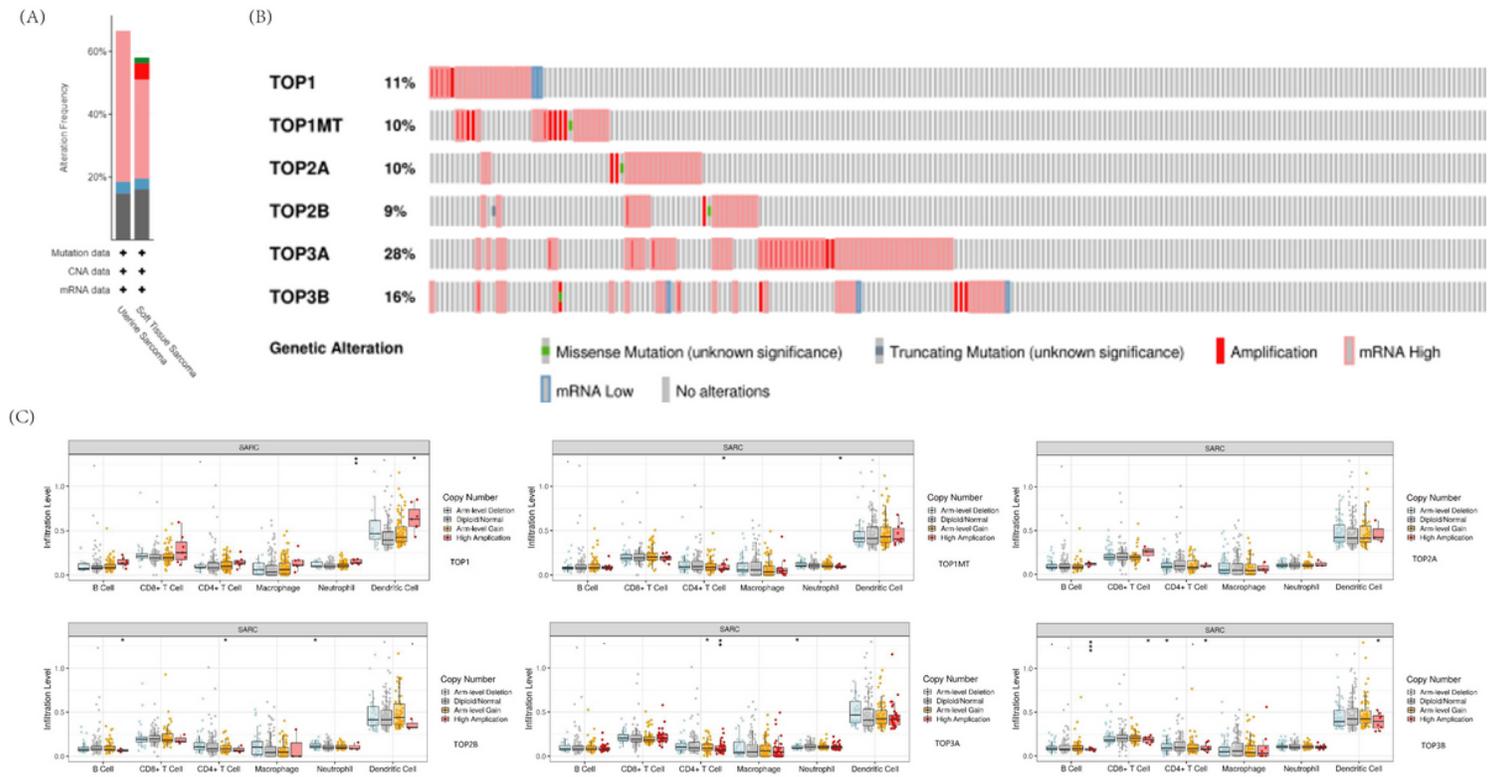


Figure 5

(A) Alteration frequency of TOPs in the database. The detection mutation rate was altered in 113 (55%) of queried patients/206 samples. (B) TOP gene mutation rates in sarcoma: TOP1 (11%), TOP1MT (10%), TOP2A (10%), TOP2B (9%), TOP3A (28%) and TOP3B (16%). (C) The relationship between the expression of TOPs and copy number variation. *P-value < 0.05.

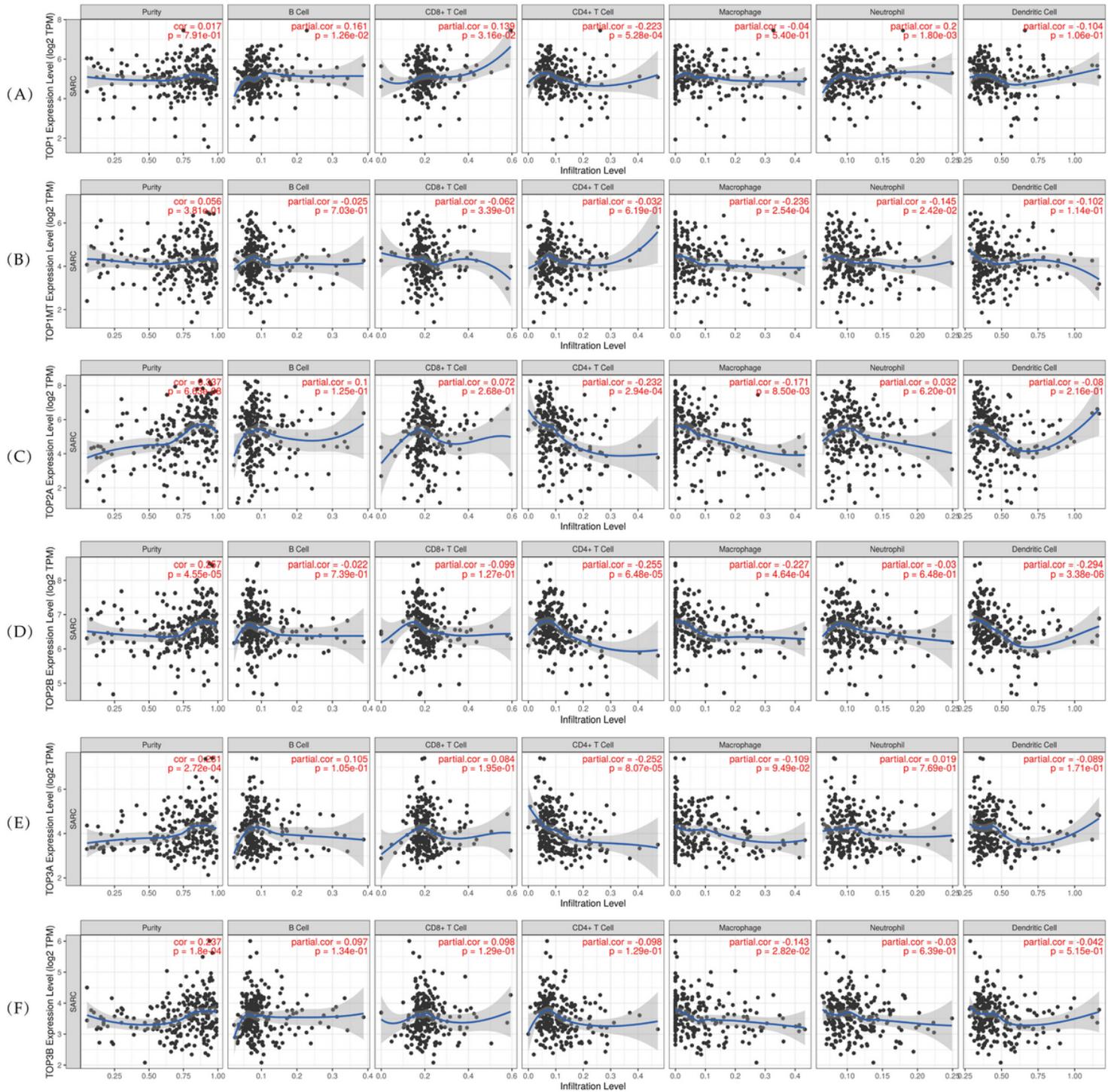


Figure 6

The relationship between the expression of TOPs and immune cell infiltration (TIMER). Correlation between immune cell abundance and the expression of (A) TOP1, (B) TOP1MT, (C) TOP2A, (D) TOP2B, (E) TOP3A, (F)TOP3B

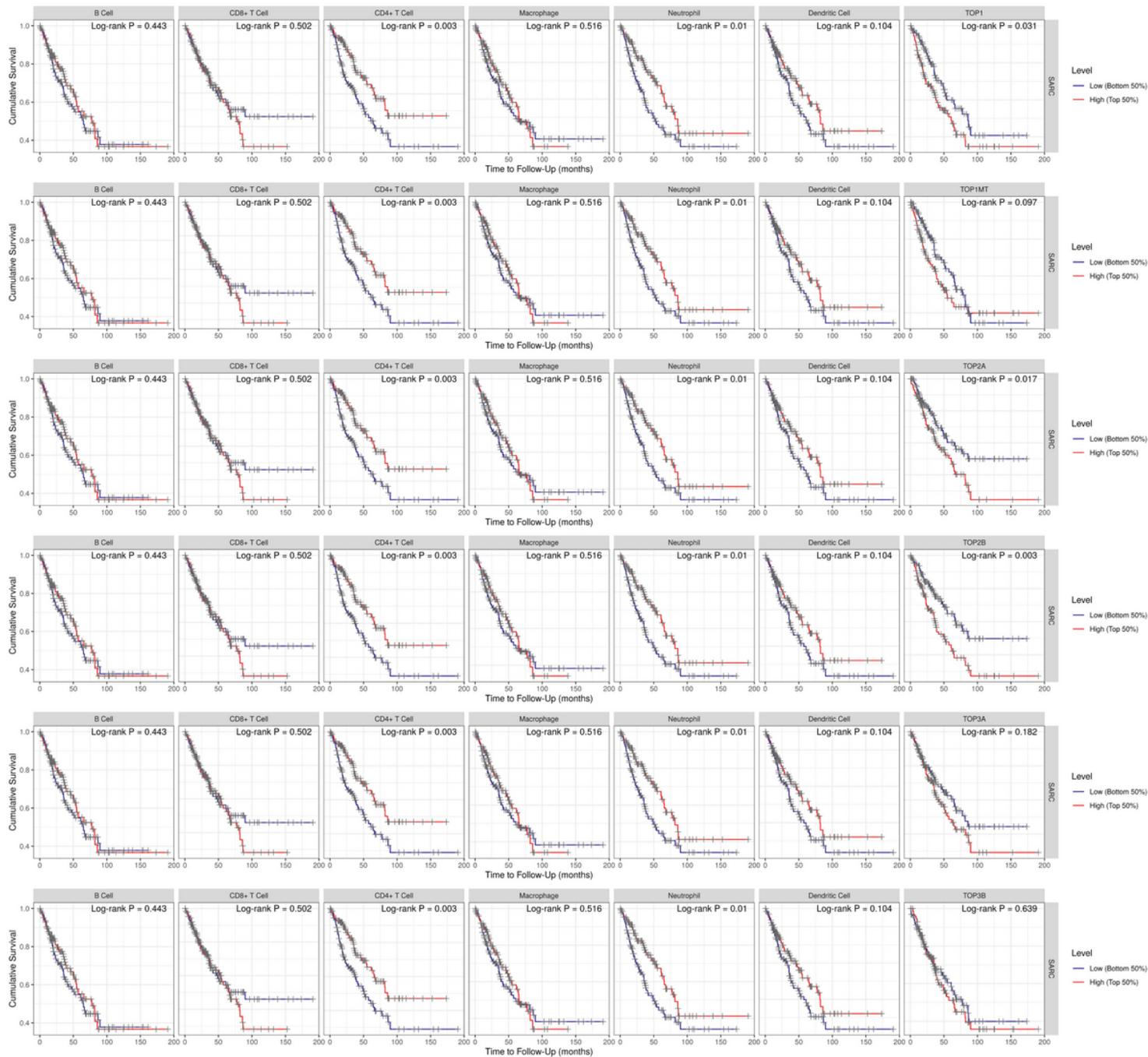
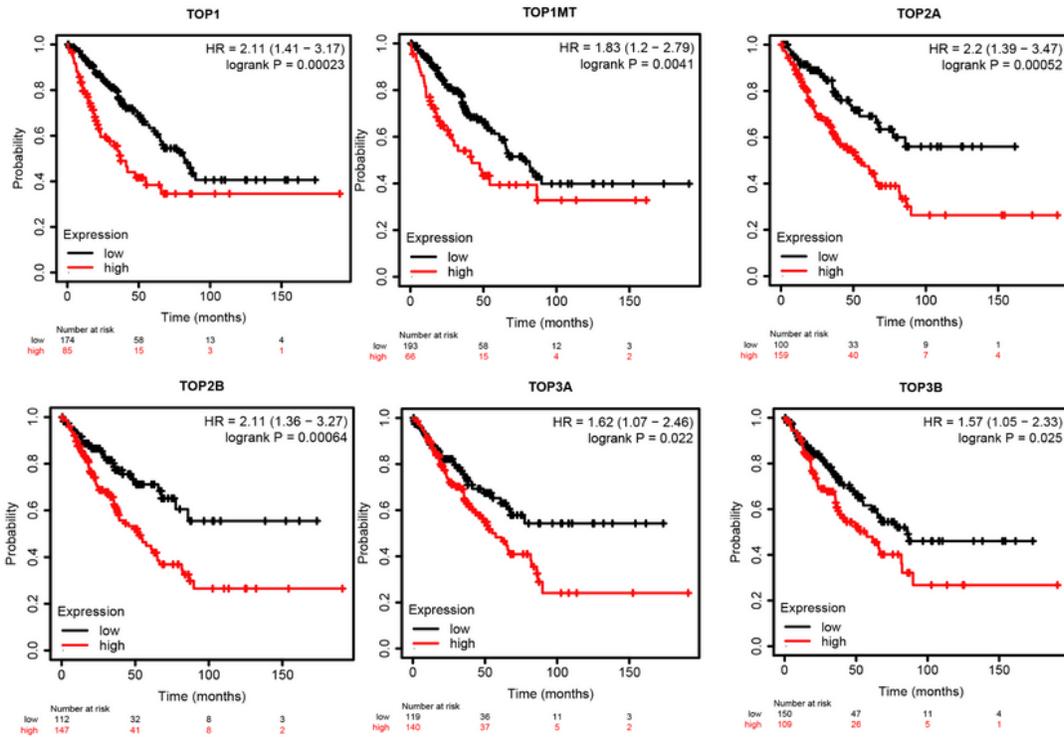


Figure 7

Immune cell infiltration of TOPs and survival curves. The immune cell infiltration of TOP1, TOP2A, and TOP2B is associated with poor prognosis. *P-value < 0.05.

(A)



(B)

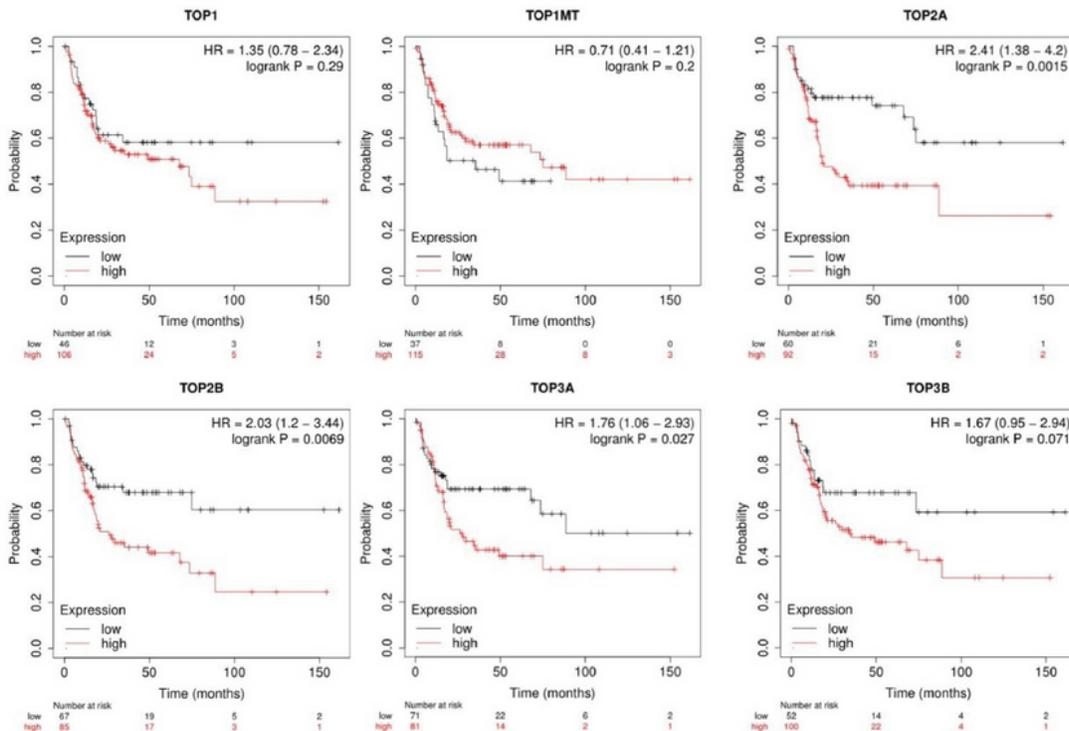
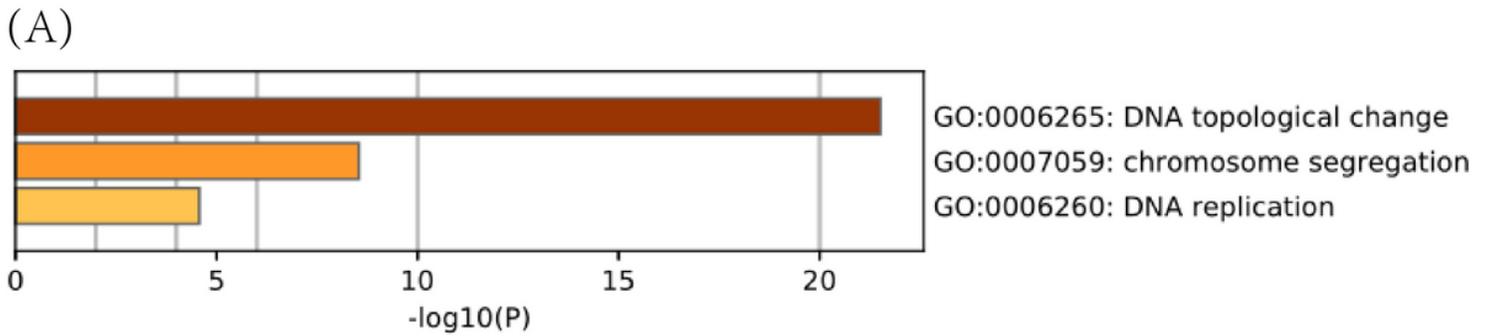


Figure 8

(A) The relationship between the expression of TOPs and OS. (B) The relationship between the expression of TOPs and DFS. The results from the Kaplan-Meier analysis and log-rank test showed that TOP gene expression changes were associated with shorter OS ($P = 3.06E-2$) and DFS ($P = 4.03E-4$) in patients with sarcoma. These results suggested that alterations in the TOP genes may significantly affect the prognosis of patients with sarcoma. * P -value < 0.05.



(B)

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0006265	GO Biological Processes	DNA topological change	6	100.00	-21.52	-17.20
GO:0007059	GO Biological Processes	chromosome segregation	5	83.33	-8.54	-4.70
GO:0006260	GO Biological Processes	DNA replication	3	50.00	-4.57	-1.20

Figure 9

(A) (B) Enriched biological processes (BPs), cellular components (CCs) and molecular functions (MFs) of TOPs. *P-value < 0.05.

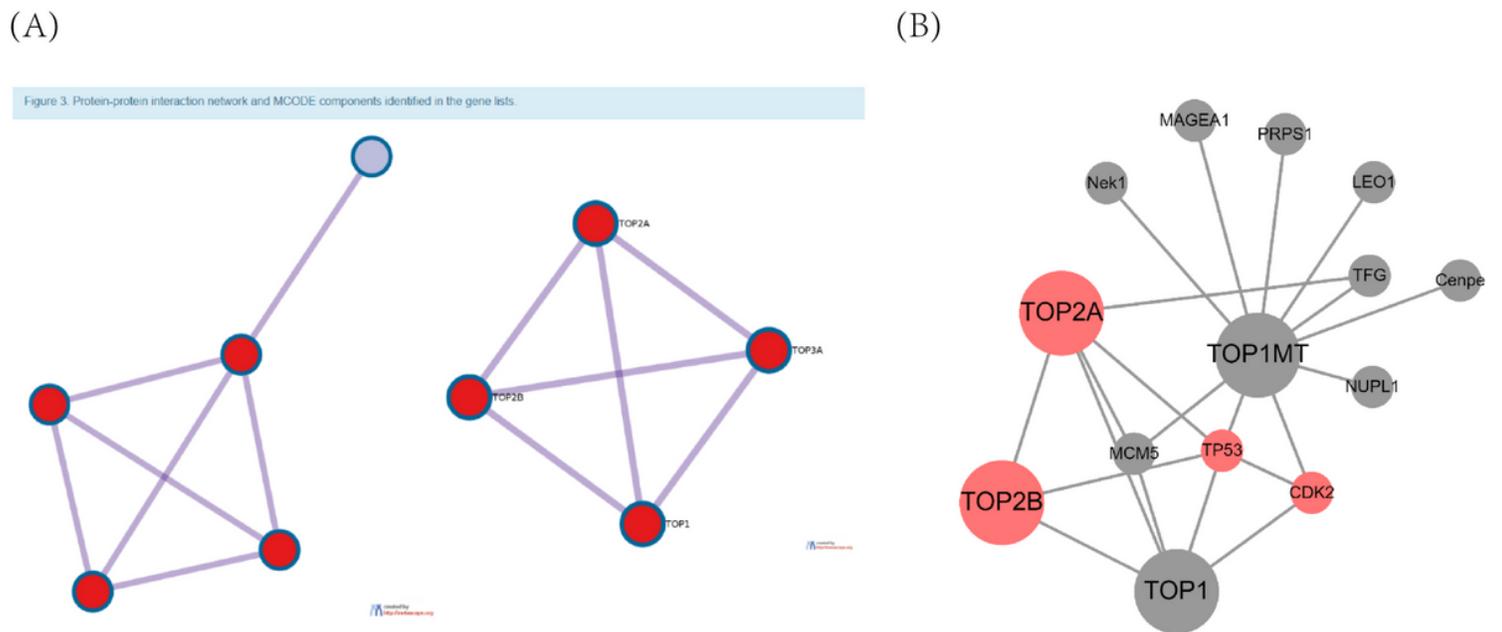


Figure 10

(A) Co-expressed genes of TOPs (Metascape). (B) Gene annotations of TOPs (NTA analysis); *P-value < 0.05.

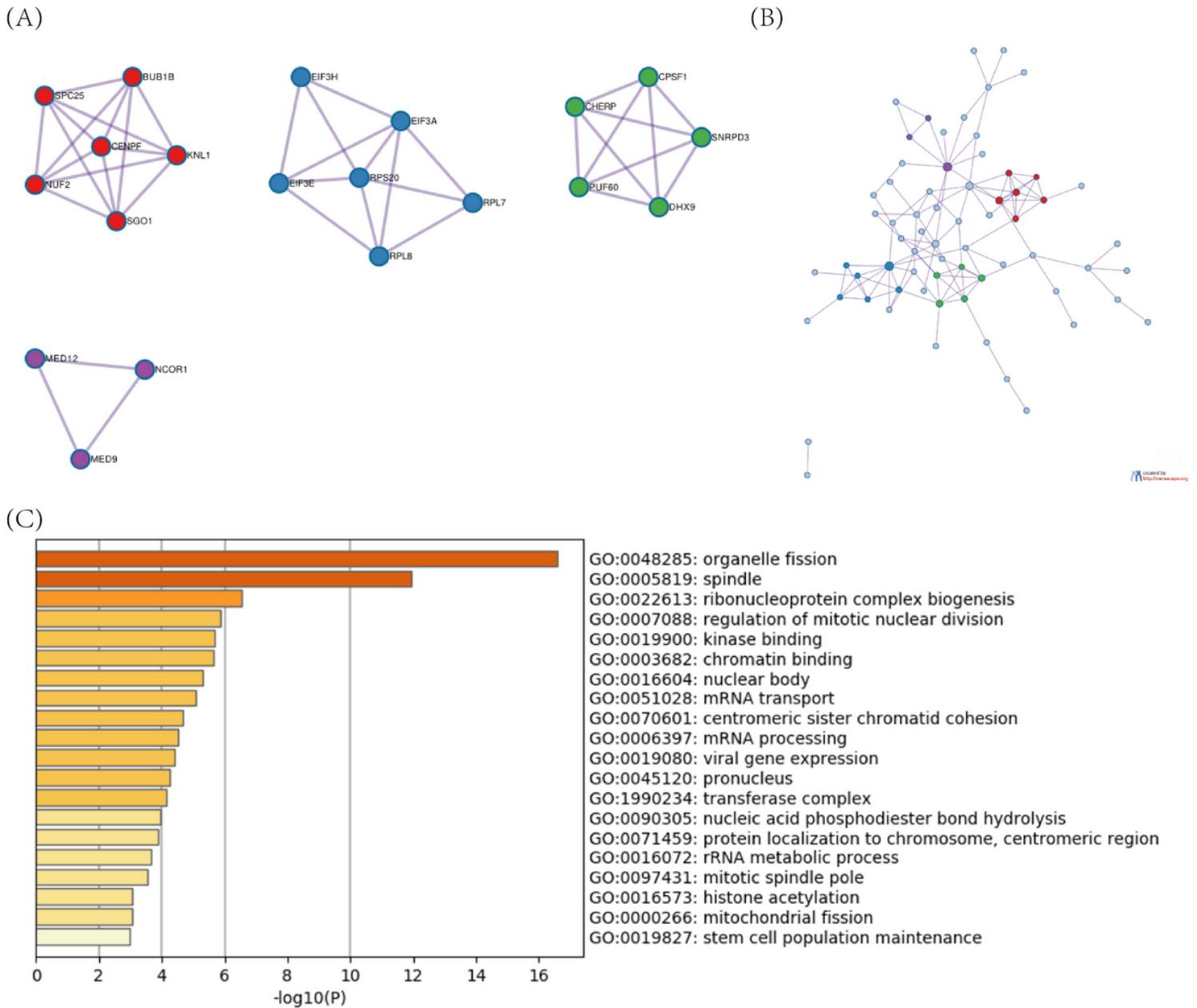


Figure 11

(A) (B) BUB1B, CENPF, SPC25, NUF2, KNL1SGO1, RPS20, SNRPD3, and MED9 are significantly correlated with TOP mutations (CBioportal); correlation=0.9, high correlation. *P-value < 0.05. (C) The GO and KEGG analysis results indicate that the TOP-like genes were associated with organelle fission, spindle, ribonucleoprotein complex biogenesis, regulation of mitotic nuclear division, and so on in sarcoma. Bar graph of enriched terms across the input gene list. *P-value < 0.05.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [additionalfile4TOPlikedgeneAnalysisReport.html](#)

- [additionalfile4metascaperesult.xlsx](#)
- [additionalfile4TOPlikedgenePPINetwork.html](#)
- [additionalfile1metascapeAnalysisReport.html](#)
- [additionalfile2coexpressedgeneReportTOP1MT.html](#)
- [additionalfile2coexpressedgeneReportTOP2A.html](#)
- [additionalfile2coexpressedgeneReportTOP2B.html](#)
- [additionalfile2coexpressedgeneReportTOP3A.html](#)
- [additionalfile2coexpressedgeneReportTOP3B.html](#)
- [additionalfile2linkedcoexpressedgeneReportTOP1.html](#)
- [additionalfile2project.pdf](#)
- [Datasetssupportingtheconclusions.docx](#)