Oncogenetic Function and Prognostic Value of DNA Topoisomerase II Alpha (TOP2A) in Human Malignancies: A Pan-Cancer Analysis

Fulai Zhao  
Shanghai University of Traditional Chinese Medicine

Junli Chang  
Shanghai University of Traditional Chinese Medicine

Wenyi Wang  
Shanghai University of Traditional Chinese Medicine

Xingyuan Sun  
Shanghai University of Traditional Chinese Medicine

Xiaoping Ma  
Shanghai University of Traditional Chinese Medicine

Peng Zhao  
Shanghai University of Traditional Chinese Medicine

Yongjun Wang  
Shanghai University of Traditional Chinese Medicine

Yanping Yang (mailto:yanpingyang@shutcm.edu.cn)  
Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: TOP2A, pan-cancer analysis, oncogenesis, immune cell infiltration, genetic alteration, prognostic value

Posted Date: October 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-947966/v1

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

Increasing studies have revealed significant associations between TOP2A with oncogenesis and prognosis of human cancers; however, pan-cancer analysis has not been reported. Here, we explored the potential carcinogenic function, the association with clinical outcomes of TOP2A in 33 different human cancers. The results showed that TOP2A was amplified in 32 investigated cancers; TOP2A expression was significantly associated with metastasis of six different cancers, and significantly associated with the survivals of patients in ten different cancers; TOP2A encoded protein was obviously upregulated in five available cancers; phosphorylated TOP2A protein at S1106 was significantly upregulated in all six available cancers. Moreover, TOP2A expression was found to be associated with the cancer-associated immune cell infiltration, including fibroblasts, Tregs and macrophages. In addition, Kyoto encyclopedia of genes and genomes (KEGG) pathway and Gene Ontology (GO) enrichment analyses revealed a most significant association between TOP2A with Wnt signaling pathway, and DNA conformation change. This work provides a comprehensive knowledge of TOP2A in different cancers, including carcinogenic function, prognostic values for metastasis and clinical outcomes.

Highlights

1. TOP2A was amplified in 32 investigated cancers, and significantly associated with the survivals of patients in ten different cancers.
2. Phosphorylated TOP2A protein at S1106 was significantly upregulated in all six available cancers.
3. TOP2A expression was found to be associated with the cancer-associated immune cell infiltration.

Introduction

As the second most-common cause of death worldwide, cancer kills more than 8 million people each year with an expected increase of incidence by more than 50% over the coming decades and a poor 5-year survival rate\textsuperscript{1,2}. Therefore, it is urgent to identify new biomarkers for the diagnosis and treatment of cancers.

Given the heterogeneity arises and complexity of carcinogenesis, it is very important to perform pan-cancer analysis of a potential biomarker gene to promote the development of combination therapies and personalized medicine by evaluating its association with the clinical prognosis and involved molecular mechanisms\textsuperscript{3}.

DNA topoisomerase is a nuclear protein, which regulates the spatial dynamics of DNA and plays an important role in life activities, such as DNA replication, transcription, recombination, and chromosome separation\textsuperscript{4}. Topoisomerase has two isoenzymes of topoisomerase I (TOP1) and topoisomerase II (TOP2), the latter includes two subtypes of TOP2A and TOP2B\textsuperscript{5}. TOP2B mainly exists in terminally differentiated cells\textsuperscript{6}, while TOP2A expression is cell-dependent and necessary for dividing cell survival\textsuperscript{7}. The protein structure of TOP2A is conserved among different species (e.g., H. sapiens, P. troglodytes, M.
mulatta, etc.) with two common domains of DTHCT (pafam08070) and TOP4c (cd00187) (Figure 7), which regulates DNA topological state during transcription for chromosome condensation and chromatin separation. Upregulated TOP2A has been reported in different cancers and is identified to be associated with poor prognosis. A long-term clinical cohort study showed that high expression of TOP2A in breast cancer patients can be used as a marker for application of anthracycline-containing chemotherapeutics. However, current researches on TOP2A focus on single or limited cancer types. Therefore, pan-cancer evidence on biological function of TOP2A cross cancers based on big clinical data is urgently needed to be explored for revealing the potential molecular mechanism of TOP2A in the pathogenesis or clinical prognosis of different cancers, thus to develop more specific treatments of various cancer types.

In current study, a comprehensive pan-cancer analysis of TOP2A among 33 solid tumors was performed based on the public databases of TCGA and GEO, including the multiple factors of the gene/total protein expressions, the survival prognostic value, the genetic alteration, the protein mutation, the protein phosphorylation, the immune cell infiltration and the related signaling pathways of TOP2A.

**Results**

**Gene and total protein expressions of TOP2A in different malignancies, pathological stages, and metastatic status**

To investigate the oncogenic function of human TOP2A, the expression of TOP2A mRNA (NM_001067.4) between the cancerous and normal control tissues, or between the primary and metastasis cancer were analyzed using the TIMER2 online tool in 33 different cancers from TCGA database. As shown in Figure 1A, TOP2A expression in tumor tissues was significantly increased versus the corresponding normal control tissues, including bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pancreatic adenocarcinoma (PADD), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), uterine corpus endometrial carcinoma (UCEC), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), pheochromocytoma and paraganglioma (PCPG). Meanwhile, TOP2A mRNA expression was also found to be significantly increased in metastasis skin cutaneous melanoma than in the primary skin cutaneous melanoma.

Moreover, the expression of TOP2A mRNA between normal and tumor tissues for those cancers without data of normal control in the TCGA database were performed using the data of corresponding normal tissues from GTEx dataset as normal controls. As shown in Figure 1B, TOP2A mRNA expression was
significantly upregulated in tumor than in the normal control tissues, including adrenocortical carcinoma (ACC), lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), acute myeloid leukemia (LAML), brain lower grade glioma (LGG), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), sarcoma (SARC), skin cutaneous melanoma (SKCM), thymoma (THYM), uterine carcinosarcoma (UCS). However, TOP2A mRNA expression showed no significant difference between the testicular germ cell tumors (TGCT) and the normal control tissues.

To further evaluate the differences of TOP2A protein (NP_001058.2) expression between normal and tumor tissues, the protein expression patterns of TOP2A in cancer patients were first extracted from the Human Protein Atlas. As shown in Figure 1D, TOP2A protein was significantly upregulated in cancer than in normal control tissues, including breast, cervical, colon, endometrium, kidney, liver, lung, melanocytes, ovary, pancreas, thyroid gland, and testis, which were consistent with the expression panel of TOP2A mRNA in patients with same cancer based on TCGA database as shown in Figure 1B; meanwhile, the data from CPTAC dataset was analyzed using the online UALCAN portal, which also showed an increased expression of TOP2A protein in patients with primary breast invasive carcinoma, ovarian serous cystadenocarcinoma, colon adenocarcinoma, uterine corpus endometrial carcinoma, and lung adenocarcinoma than in normal tissues, and consistent with the expression panel of TOP2A mRNA in patients with same cancer; however, TOP2A protein showed a lower expression in kidney renal clear cell carcinoma (Figure 1C), which was inconsistent with the mRNA expression pattern in patients with kidney renal clear cell carcinoma based on TCGA database as shown in Figure 1A and protein expression pattern in patients with kidney renal clear cell carcinoma from Human Protein Atlas as shown in Figure 1D.

To analyze the association of TOP2A mRNA expression level with the metastatic abilities of different cancers, the “compare tumor, normal and metastasis” module of the TNMplot on line tool were applied. The results showed that, in breast, kidney, liver, lung, prostate, and skin cancers, TOP2A expression was significantly upregulated in the tumor than in the normal control tissues; as well as in the metastatic versus the tumor tissues (Figure 1E).

To further explore the association of TOP2A mRNA expression level with the pathological stages of different cancers, the “Pathological Stage Plot” module of HEPIA2 (Figure 1F) and Ualcan database (Figure 1G) were applied respectively, which revealed a significant difference in TOP2A expression among clinical stages of adrenocortical carcinoma, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, and breast invasive carcinoma. The highest mRNA expression of TOP2A in adrenocortical carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, and lung adenocarcinoma appeared in stage 4, while in lung squamous cell carcinoma, head and neck squamous cell carcinoma and breast invasive carcinoma appeared in stage 3 (Figure 1G).

**Prognostic Value Of Top2a In Different Cancer Patients**
To explore the prognostic value of TOP2A expression level in different cancers, the cancer cases were divided into the high-expression group and the low-expression group according to the expression levels of TOP2A mRNA, followed by analyzing the associations between the TOP2A expression levels with the survival status of different cancer patients, based on the databases of TCGA and GEO, respectively. The results were visualized as a heat map (Figure 2A, 2C) and the survival curves with Kaplan-Meier method (Figure 2B, 2D-2F). Our results identified that high TOP2A expression was significantly associated with the poor overall survival (OS) for cancers of adrenocortical carcinoma ($p=0.00014$), kidney renal clear cell carcinoma ($p=0.00021$), kidney renal papillary cell carcinoma ($p<0.001$), brain lower grade glioma ($p<0.001$), liver hepatocellular carcinoma ($p=0.0028$), lung adenocarcinoma ($p=0.011$), mesothelioma ($p<0.001$), and pancreatic adenocarcinoma ($p=0.036$) (Figure 2B), as well as significantly associated with the poor disease-free survival (DFS) for adrenocortical carcinoma ($p=0.0036$), kidney chromophobe ($p=0.015$), kidney renal clear cell carcinoma ($p=0.00072$), kidney renal papillary cell carcinoma ($p<0.001$), brain lower grade glioma ($p<0.001$), liver hepatocellular carcinoma ($p=0.00053$), mesothelioma ($p=0.011$), pancreatic adenocarcinoma ($p=0.00092$), prostate adenocarcinoma ($p<0.001$), sarcoma ($p=0.018$), thyroid carcinoma ($p=0.0023$), uveal melanoma ($p=0.0016$) based on TCGA database analyzed by GEPIA 2 online tool (Figure 2D).

To further confirm the prognostic value of TOP2A expression level in different cancers, Kaplan-Meier method (https://kmplot.com/analysis/) was used to plot the survival curves, which presented a significant correlation between the high TOP2A expression and the poor survival status. As shown in Figure 2E, high TOP2A expression was found to be significantly associated with the poor OS ($p<2.9e-07$), distant metastasis-free survival (DMFS) ($p=9.4e-10$), post-progression survival (PPS) ($p=0.0076$), and relapse-free survival (RFS) ($p<1E-16$) of breast cancer patients; significantly associated with the poor OS ($p=0.00012$), disease-specific survival (DSS) ($p=4.8e-05$), progress-free survival (PFS) ($p=3e-06$), and RFS ($p=0.00014$) of liver cancer patients; significantly associated with the poor OS ($p<1E-16$), first progression (FP) ($p=1.6e-12$), and PPS ($p=0.0091$) of lung cancer patients; as well as significantly associated with the poor OS of ovarian cancer ($p=0.039$). And then “The Human Protein Atlas” module of “PATHOLOGY” was used to analyze the association between TOP2A expression level and the prognosis of different cancer patients, which demonstrated that the high TOP2A expression was closely related to the poor prognosis of renal, liver, pancreatic, and lung cancers (Figure 2F, all $p<0.001$).

Genetic And Protein Alteration Characteristic Analysis

To explore the influences of genetic and amino acid (AA) alterations on carcinogenesis of different cancers, both the genetic alteration and AA mutation frequencies of TOP2A in all cancer patients from TCGA cohorts were analyzed using cBioPortalp FOR CANCER GENOMICS on line tool. Our results showed that among the 10953 cancer patients, 381 cases appeared TOP2A alteration with a frequency of 3% (381/10953), and amplification was the main genetic alteration types of TOP2A (Figure 3A, 3B). Furthermore, in TOP2A, the highest alteration frequency of “amplification” appeared in esophageal adenocarcinoma (>9%); of “mutation” appeared in skin cutaneous melanoma patients (>8%); and of
“deep deletion” appeared in adrenocortical carcinoma (>3%). It is worth noting that uveal melanoma patients showed a ~2% frequency of exclusive genetic alteration (deep deletion) in TOP2A (Figure 3B). As shown in Figure 3C, the 215 AA in the Tudor domain of TOP2A protein (T215P) showed the highest mutation frequency, which was detected in 7 ovarian serous cystadenocarcinoma cases.

Moreover, the potential association between TOP2A genetic alteration and the clinical survival prognosis of different cancer patients were analyzed based on the TCGA cohorts using the cBioPartalp FOR CANCER GENOMICS online tool. Patients with TOP2A genetic alteration showed worse prognoses in overall survival \( (p=2.632e-03) \), disease-specific survival \( (p=8.530e-03) \), and progression-free survival \( (p=2.237e-03) \), but not disease-free survival \( (p=0.222) \), versus those cases without TOP2A genetic alteration (Figure 3D).

**Protein Phosphorylation Analysis Data**

To identify the differences in TOP2A phosphorylation levels with specific phosphorylation sites between normal tissues and primary tumor tissues, the CPTAC module in UALCAN web resource was applied for analysis over six different tumors (breast cancer, colon cancer, kidney renal clear cell carcinoma, lung adenocarcinoma, ovarian cancer, and uterine corpus endometrial carcinoma). The results revealed a higher S1106 phosphorylation level of TOP2A in all primary tumor tissues than in the normal tissues (Figure 4A-4F).

**Immune Cell Infiltration Analysis Data**

Since the importance of tumor-immune cell infiltrations in initiation, development and metastasis of cancers, the potential relationship between TOP2A expression level with tumor-immune cell infiltration, including cancer-associated fibroblasts, Tregs, and macrophages, were analyzed using eight different algorithms of TIDE, TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, XCELL, MCPCounter, and EPIC based on the TCGA database. Our results demonstrated a statistical positive correlation of TOP2A expression with the estimated infiltration value of cancer-associated fibroblasts in cervical squamous cell carcinoma and endocervical adenocarcinoma, human papillomavirus negative head and neck squamous cell carcinoma, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, brain lower grade glioma, mesothelioma, as well as thyroid carcinoma (Figure 5A); and the representative scatterplot data of one algorithm (EPIC) were provided in Figure 5B. Meanwhile, a statistical positive correlation of TOP2A expression with the estimated infiltration value of Tregs in kidney chromophobe, kidney renal clear cell carcinoma, liver hepatocellular carcinoma, pheochromocytoma and paraganglioma, and thyroid carcinoma were revealed (Figure 5C); and the representative scatterplot data of one algorithm (CIBERSORT) were provided in Figure 5D. We also observed a statistical positive correlation of TOP2A expression and the estimated infiltration value of macrophages in bladder urothelial carcinoma, kidney renal clear cell carcinoma, prostate adenocarcinoma, and thyroid carcinoma (Figure 5E); and the representative scatterplot data of one algorithm (TIMER) were provided in Figure 5F.
Enrichment Analysis Of Top2a-related Partners

To further investigate the involved molecular mechanisms of TOP2A gene in tumorigenesis, the KEGG pathway and GO enrichment analyses were performed for the TOP2A-binding proteins. Based on the STRING on line tool, 50 experimental validated TOP2A-binding proteins were identified and the protein-protein interaction was visualized with Cytoscape (Figure 6A). The KEGG pathway and GO enrichment analyses were then performed using the online bioinformatics tool “HILOT”, which showed that “Wnt signaling pathway” might be the most important signaling pathway involved in TOP2A trigged tumor pathogenesis(Figure 6B); and the GO enrichment analysis data further indicated that most of these genes were linked to DNA structure, biological complex, and protein kinase activity, such as DNA conformation change, PcG protein complex, protein serine/threonine kinase activity (Figure 6C).

Discussion

Increasing studies have revealed the importance of TOP2A in carcinogenesis, metastasis and prognosis of different human cancers \(^{14-17}\). Pan-cancer analysis of TOP2A, in terms of association with the clinical prognosis and involved molecular mechanisms based on profiling data, is necessary to understand the tumorigenesis mechanisms and explore the potential diagnostic biomarkers and treatment targets in clinical practice.

Our results based on TCGA database showed that TOP2A mRNA was amplified in 32 out of 33 available tumor tissues except for testicular germ cell tumors, which only showed an increased trend without statistical significance in the tumor than in the normal control tissues. These findings were further confirmed by analyzing the TOP2A mRNA expression profile based on GEO database using the TNMplot on line tool, furthermore, our results showed that TOP2A mRNA expression was more higher in the metastatic tissues than in the paired tumor tissues of breast cancer, kidney cancer, liver cancer, lung cancer, prostate cancer, and skin cancer tissues.

Moreover, our results showed that high TOP2A mRNA expression was significantly associated with the clinical stages of patients with adrenocortical carcinoma, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, or breast invasive carcinoma; meanwhile, high TOP2A mRNA expression was also significantly associated with poor overall survival and diseases free survival of patients with adrenocortical carcinoma, breast invasive carcinoma, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, brain lower grade glioma, liver hepatocellular carcinoma, lung adenocarcinoma, mesothelioma, or pancreatic adenocarcinoma. Consistent with mRNA expression profile, TOP2A protein expression determined by immunohistochemistry assay was also significantly up-regulated in 12 available tumor tissues. These findings indicate that TOP2A can be used as an independent prognostic predictor of cancers mentioned above.
Studies have shown that genomic alterations, such as mutations, small nucleotide polymorphisms, translocations, deletions and insertions, often occur in human cancers, which may involve in carcinogenesis \(^{18,19}\). Our results identified a 3% mutation rate of TOP2A in all 10953 tested cancer patients, and TOP2A mutation was significantly associated with poor overall survival, disease-specific survival, and progression free survival in patients with ovarian serous cystadenocarcinoma, while the associations between TOP2A mutation with survival prognosis of other cancers needs to be further investigated.

Protein phosphorylation is a post-translational modification, which is one of the most important regulatory mechanisms in cells\(^{20}\). Phosphorylation of TOP2A specific sites occurs in a cell cycle-dependent manner and also regulates the sensitivity to TOP2A targeted drugs \(^{21}\). In this study, we used the CPTAC dataset to explore the molecular mechanisms of TOP2A protein in breast cancer, clear cell renal cell carcinoma, lung adenocarcinoma, ovarian cancer, and uterine corpus endometrial carcinoma in terms of total protein and phosphoprotein levels. Our findings revealed higher expressions of both total and S1106 phosphorylated TOP2A proteins in the primary tumors than in the normal control tissues, suggesting the importance of S1106 phosphorylated TOP2A protein in tumorigenesis and drug resistance, and the necessary of exploring S1106 phosphorylated TOP2A protein related mechanisms to design new therapeutic strategies for cancers resistant to TOP2A targeted drugs.

Infiltration of different types of immune cells is closely associated with initiation, development and metastasis of cancers, thus a promising source of novel diagnostic and prognostic biomarkers \(^{22,23}\). Many challenges still remain in extensive application of tumor immunotherapy \(^{24,25}\), therefore, it is important to fully understand the status of tumor-immune cell infiltration and explore new biomarkers for improving tumor immunotherapy. At present, the association between TOP2A expression and tumor-immune cell infiltration has not been fully reported. By using multiple immune deconvolution methods to detect the correlation between TOP2A expression and the infiltration levels of different immune cells, including cancer-associated fibroblasts, Tregs, and macrophages, in available cancers, we found out that TOP2A expression was significantly associated with tumor-immune cell infiltration, indicating that TOP2A expression can reflect the status of tumor-immune cell infiltration, and further basic experiments and clinical trials are worth performing to validate our findings and promote the clinical translation.

KEGG and GO enrichment analysis of TOP2A functional related genes can help to reveal their potential functions and involved molecular mechanisms. Our results identified the significant association with Wnt signaling pathway, thyroid hormone signaling pathway, and DNA conformation changes. These findings are consistent with the literature reports that TOP2A induces the malignant characteristics of tumors by activating the Wnt/\(\beta\)-catenin signaling pathway \(^{26,27}\).

In conclusion, our pan-cancer analysis of TOP2A showed that TOP2A was highly expressed in most cancers, and significantly correlated with cancer development and patient survival prognosis. These findings provide a comprehensive and systematic understanding of TOP2A in tumorigenesis from the
perspective of clinical tumor samples; a promising diagnostic and prognostic biomarker and a therapeutic target for multiple cancers.

**Materials And Methods**

**Gene expression Exploration**

To identify the differently expressed TOP2A gene between various cancers/specifc cancer subtypes and paired paracancerous normal tissues based on TCGA projects, “TOP2A” was used as the input term for the online “Gene_DE” module of TIMER2 (tumor immune estimation resource, version 2) (http://timer.cistrome.org/) \(^{28}\). For those cancers with extremely less or without paracancerous normal tissues [such as TCGA-ACC (Adrenocortical carcinoma) and TCGA-DLBC (Lymphoid Neoplasm Diffuse Large B-cell Lymphoma)], “Expression analysis-Box Plots” module of Gene Expression Profiling Interactive Analysis version 2 (GEPIA2) online tool (http://gepia2.cancer-pku.cn/#analysis) \(^{29}\) was used to achieve the box plots of differently expressed genes between cancerous and corresponding normal tissues from the Genotype-Tissue Expression (GTEx) database based on the cutoff values: log2FC (fold change) \(=1\), \(p\)-value \(= 0.01\), and “Match TCGA normal and GTEx data”. Moreover, the violin or box plots were obtained from log2 [TPM (Transcripts per million)+1] transformed expression data; meanwhile, the violin plots were also achieved from TNMplot (differential gene expression analysis in Tumor, Normal and Metastatic tissues) online tool (http://www.tnmplot.com) \(^{30}\) and the Gene-chip data, to discover TOP2A expression in tumor tissues, normal tissues, and metastatic tissues with Kruskal-Wallis test for significance assay of Gene-chip data.

The violin plots of TOP2A expression at diverse pathological stages (stage-I, -II, -III, -IV, and -I) of all cancers in TCGA database were got by “Pathological Stage Plot” module of HEPIA2. The UALCAN portal website (http://ualcan.path.uab.edu/analysis-prot.html) is used to analyze the mRNA expression of TOP2A in primary tissues and its association with clinical staging.

Protein expression investigation based on the CPTAC (Clinical proteomic tumor analysis consortium) dataset was performed using online UALCAN portal, which provides an easy access to the publicly available cancer OMICS data \(^{31}\). Therefore, we respectively identifed the expressions of total or phosphorylated TOP2A (NP_001058.2) (with phosphorylation sites of S1106, S1374, S1213, S1247, S1351, S1354, S1377, S1393, S1525, S1374S1377, S1351S1354, and T1343S1351S1354) between primary cancer and normal tissues using “TOP2A” as input. Six cancer datasets were fnally included in this study, including lung adenocarcinoma (LUAD), uterine corpus endometrial carcinoma (UCEC), Renal cell carcinoma (clear cell RCC), breast cancer, ovarian cancer, and colon cancer.

**Prognosis Investigation**

GEPIA2 online “Survival Map” module \(^{29}\) was applied to analyze Disease-free survival (DFS) and overall survival (OS) significance map data of TOP2A among all cancers in TCGA dataset. Survival curves were
plotted by GEPIA2 “Survival Analysis” module, using log-rank in hypothesis test and median expression values as cutoff of high- and low-expression cohorts.

Genetic Alteration Assay

TOP2A genetic alteration features were analyzed using “TOP2A” as the querying term in “Quick select” section of “TCGA Pan-Cancer Atlas Studies” in the cBioPortal online tool (https://www.cbioportal.org/). Mutation type, alteration frequency, and copy number alteration (CNA) across all cancers in TCGA database were obtained based on the “Cancer Types Summary” module. Meanwhile, the information of TOP2A genetic mutation site can be displayed by the “Mutations” module in schematic diagram of protein structure (Figure 3C). Differences of overall survival, progression-free survival and disease-free survival for cancer cases in TCGA database without or with TOP2A genetic alteration were analyzed with “Comparison” module. Kaplan-Meier method with log-rank $p$-value was used for plotting the survival curves.

Investigation of Tumor-immune Cell Infiltration

Online “Immune-Gene” module of TIMER2 was applied to investigate association between TOP2A gene expression and tumor-immune cell infiltrates across all cancers in TCGA database, including cancer-associated fibroblasts, Tregs, and macrophages. The immune cell infiltration was evaluated with TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, XCELL, MCPCOUNTER, and EPIC, the seven different algorithms. $p$- and partial correlation (cor)-values were calculated with purity-adjusted Spearman’s rank correlation test. Data visualization was achieved by heat maps and scatter plots.

Enrichment Analysis of TOP2A Related Genes

A single protein name (“TOP2A”) AND an organism (“Homo sapiens”) were first queried online using STRING (https://string-db.org/) followed by setting the following parameters of active interaction sources (“experiments”), max number of interactors to show (“no more than 50 interactors” in 1st shell), meaning of network edges (“evidence”) and minimum required interaction score (“Low confidence (0.150)”) to collect the available TOP2A-binding proteins validated by experiments. Then the obtained experimentally confirmed TOP2A-binding proteins were further visualized using Cytoscape. Moreover, using the online bioinformatics tool “HIPLON” (https://hiplot.com.cn/), the Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis, and the Gene Ontology (GO) enrichment analysis including cellular component, molecular function and biological process of two sets of data after being combined. Two-tailed $p<0.01$ indicated statically significance.

Declarations

Funding

This work was supported by (1) National Key R&D Program of China (2018YFC1704300 and 2020YFE0201600); (2) National Nature Science Foundation (81973877, 82174408).
References

1 Consortium, I. T. P.-C. A. o. W. G. Pan-cancer analysis of whole genomes. *Nature* **578**, 82-93, doi:10.1038/s41586-020-1969-6 (2020).

2 Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer* **103**, 356-387, doi:10.1016/j.ejca.2018.07.005 (2018).

3 NIH Completes Genomic Analysis of 33 Cancer Types. *J Nucl Med* **59**, 11N-12N (2018).

4 Champoux, J. J. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* **70**, 369-413 (2001).

5 McKinnon, P. J. Topoisomerases and the regulation of neural function. *Nat Rev Neurosci* **17**, 673-679, doi:10.1038/nrn.2016.101 (2016).

6 Isik, S. *et al.* DNA topoisomerase IIbeta as a molecular switch in neural differentiation of mesenchymal stem cells. *Ann Hematol* **94**, 307-318, doi:10.1007/s00277-014-2209-7 (2015).

7 Cobb, J., Reddy, R. K., Park, C. & Handel, M. A. Analysis of expression and function of topoisomerase I and II during meiosis in male mice. *Mol Reprod Dev* **46**, 489-498 (1997).

8 Jain, M. *et al.* TOP2A is overexpressed and is a therapeutic target for adrenocortical carcinoma. *Endocr Relat Cancer* **20**, 361-370, doi:10.1530/ERC-12-0403 (2013).

9 Depowski, P. L. *et al.* Topoisomerase IIalpha expression in breast cancer: correlation with outcome variables. *Mod Pathol* **13**, 542-547 (2000).

10 Erguden, H. C. *et al.* The association of topoisomerase 2alpha expression with prognosis in surgically resected non-small cell lung cancer (NSCLC) patients. *J Thorac Dis* **4**, 352-357, doi:10.3978/j.issn.2072-1439.2012.08.04 (2012).

11 Parker, A. S. *et al.* Higher expression of topoisomerase II alpha is an independent marker of increased risk of cancer-specific death in patients with clear cell renal cell carcinoma. *Eur Urol* **66**, 929-935, doi:10.1016/j.eururo.2013.12.017 (2014).

12 Leo, A. D. *et al.* HER2 and TOP2A as predictive markers for anthracycline-containing chemotherapy regimens as adjuvant treatment of breast cancer: a meta-analysis of individual patient data. *The Lancet Oncology* **12**, 1134-1142, doi:10.1016/s1470-2045(11)70231-5 (2011).

13 Nagy, A., Munkacsy, G. & Gyorffy, B. Pancancer survival analysis of cancer hallmark genes. *Sci Rep* **11**, 6047, doi:10.1038/s41598-021-84787-5 (2021).

14 Gao, Y. *et al.* TOP2A Promotes Tumorigenesis of High-grade Serous Ovarian Cancer by Regulating the TGF-beta/Smad Pathway. *J Cancer* **11**, 4181-4192, doi:10.7150/jca.42736 (2020).
15 Kou, F. et al. TOP2A Promotes Lung Adenocarcinoma Cells' Malignant Progression and Predicts Poor Prognosis in Lung Adenocarcinoma. *J Cancer* **11**, 2496-2508, doi:10.7150/jca.41415 (2020).

16 Aumayr, K. et al. HER2 and TOP2A Gene Amplification and Protein Expression in Upper Tract Urothelial Carcinomas. *Pathol Oncol Res* **24**, 575-581, doi:10.1007/s12253-017-0260-0 (2018).

17 Scott, A. T. et al. Gene Expression Signatures Identify Novel Therapeutics for Metastatic Pancreatic Neuroendocrine Tumors. *Clin Cancer Res* **26**, 2011-2021, doi:10.1158/1078-0432.CCR-19-2884 (2020).

18 Palangat, M. et al. The splicing factor U2AF1 contributes to cancer progression through a noncanonical role in translation regulation. *Genes Dev* **33**, 482-497, doi:10.1101/gad.319590.118 (2019).

19 Vedham, V. & Verma, M. Cancer-associated infectious agents and epigenetic regulation. *Methods Mol Biol* **1238**, 333-354, doi:10.1007/978-1-4939-1804-1_18 (2015).

20 Zolnierowicz, S. & Bollen, M. Protein phosphorylation and protein phosphatases. De Panne, Belgium, September 19-24, 1999. *EMBO J* **19**, 483-488 (2000).

21 Chikamori, K. et al. Phosphorylation of serine 1106 in the catalytic domain of topoisomerase II alpha regulates enzymatic activity and drug sensitivity. *J Biol Chem* **278**, 12696-12702, doi:10.1074/jbc.M300837200 (2003).

22 Fridman, W. H. et al. Immune infiltration in human cancer: prognostic significance and disease control. *Curr Top Microbiol Immunol* **344**, 1-24, doi:10.1007/82_2010_46 (2011).

23 Steven, A. & Seliger, B. The Role of Immune Escape and Immune Cell Infiltration in Breast Cancer. *Breast Care (Basel)* **13**, 16-21, doi:10.1159/000486585 (2018).

24 Zhang, Y. & Zhang, Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol* **17**, 807-821, doi:10.1038/s41423-020-0488-6 (2020).

25 Murciano-Goroff, Y. R., Warner, A. B. & Wolchok, J. D. The future of cancer immunotherapy: microenvironment-targeting combinations. *Cell Res* **30**, 507-519, doi:10.1038/s41422-020-0337-2 (2020).

26 Liu, S. L. et al. DGCR5 is activated by PAX5 and promotes pancreatic cancer via targeting miR-3163/TOP2A and activating Wnt/beta-catenin pathway. *Int J Biol Sci* **17**, 498-513, doi:10.7150/ijbs.55636 (2021).

27 Pei, Y. F., Yin, X. M. & Liu, X. Q. TOP2A induces malignant character of pancreatic cancer through activating beta-catenin signaling pathway. *Biochim Biophys Acta Mol Basis Dis* **1864**, 197-207, doi:10.1016/j.bbadis.2017.10.019 (2018).
28 Li, T. et al.TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* 48, W509-W514, doi:10.1093/nar/gkaa407 (2020).

29 Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 47, W556-W560, doi:10.1093/nar/gkz430 (2019).

30 Bartha, A. & Gyorffy, B. TNMplot.com: A Web Tool for the Comparison of Gene Expression in Normal, Tumor and Metastatic Tissues. *Int J Mol Sci* 22, doi:10.3390/ijms22052622 (2021).

31 Chen, F., Chandrashekar, D. S., Varambally, S. & Creighton, C. J. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. *Nat Commun* 10, 5679, doi:10.1038/s41467-019-13528-0 (2019).

32 Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).

33 Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6, pl1, doi:10.1126/scisignal.2004088 (2013).

34 Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, D607-D613, doi:10.1093/nar/gky1131 (2019).

35 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284-287, doi:10.1089/omi.2011.0118 (2012).

**Figures**
Figure 1

TOP2A gene and total protein expressions in different malignancies, pathological stages, and metastatic status. (A) Differently expressed TOP2A gene between various malignancies with/without specific malignant subtypes and the available normal control using the TIMER2 online tool based on TCGA database were presented as box plots. (B) Differently expressed TOP2A gene between cancers of ACC, BRCA, DLBC, LAML, LGG, OV, PAAD, SARC, SKCM, THYM, UCS and TGCT from TCGA database and the
available normal control from GTEx database were presented as box plots. (C) Differently expressed TOP2A total protein between the primary cancer (LUAD, colon cancer, ovarian cancer, breast cancer, UCEC, and the clear cell RCC) and the paired normal tissues based on the CPTAC dataset. (D) Representative immunohistochemistry images of TOP2A protein expressions in different cancer and normal control tissues from Human Protein Atlas. (E) TNMplot for differently expressed TOP2A gene expressions in normal, cancerous, and metastatic cancerous tissues of breast, kidney, liver, lung, prostate, and skin. (F) Differently expressed TOP2A gene among various pathological stages (I, II, III, IV, and V) of ACC, BRCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD and LUSC based on TCGA database. Log2 (TPM+1) was used for log-scale. (G) Boxplots showing TOP2A expression in stage-1, -2, -3 and -4 of ACC, BRCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD and LUSC with or without normal control based on TCGA database. *p<0.05, ** p<0.01; *** p<0.001.
Prognostic association between TOP2A gene expression and survival of cancer patients based on TCGA database. GEPIA2 online tool analyzed association between TOP2A gene expression with overall survival (A and B), (A) Survival map of TOP2A gene with significant associations in ACC, KIRC, KIRP, LGG, LIHC, LUAD, MESO, and PAAD; (B) Survival curves of TOP2A gene with significant associations in ACC, CESC, KIRC, KIRP, LGG, LIHC, LUAD, MESO, PAAD, and THYM. GEPIA2 online tool analyzed association between
TOP2A gene expression with disease-free survival (C and D), (C) Disease-free map of TOP2A gene with significant associations in ACC, KICH, KIRC, KIRP, LGG, LIHC, MESO, PAAD, PRAD, SARC, THCA, and UVM; (D) Disease-free survival curves of TOP2A gene with significant associations in ACC, KICH, KIRC, KIRP, LGG, LIHC, MESO, PAAD, PRAD, SARC, THCA, and UVM. (E) Correlation analysis between TOP2A gene expression and survivals (OS, DMFS, RFS, PPS, PFS, DSS, and FP) of cancer patients using Kaplan-Meier plotter. (F) The prognostic value of TOP2A gene in renal cancer, liver cancer, pancreatic cancer, and lung cancer in HPA dataset.
Figure 3

TOP2A mutations in all cancers from TCGA database. The cBioPortal tool was applied to analyze the mutation characteristics of TOP2A in TCGA database, the (A) Genomic alteration frequencies and types; (B) genetic alteration frequencies (381/10953; 3%) and types of TOP2A gene in 10953 cancer patients with 10967 samples; (C) the alteration sites of TOP2A protein; (D) associations between the TOP2A alteration status of TOP2A with the overall survival, the disease specific survival, the progression free survival and the disease-free survival in OV patients were analyzed by the cBioPortal tool.
Figure 4

Phosphorylation level of TOP2A protein in various cancers. Expression levels of TOP2A protein with different phosphorylation sites (NP_001058.2, S1106, S1374, S1213, S1247, S1351, S1354, S1377, S1393, S1525, S1374S1377, S1351S1354 or T1343S1351S1354) between selected primary cancer and normal tissues analyzed by UALCAN on line tool based on the CPTAC dataset. Box plots for (A)BRCA; (B)LUAD; (C)UCEC; (D)COAD; (E)KIRC; (F)OV.
Figure 5

Association analysis between TOP2A gene expressions with cancer associated fibroblast, Tregs and macrophage infiltration. Potential associations between TOP2A gene expressions and infiltration levels of cancer-associated fibroblasts, Tregs and macrophages across all cancers in TCGA database were explored by different algorithms.

Figure 6
Enrichment analysis of TOP2A related genes. Experimentally confirmed TOP2A-binding proteins obtained using the STRING online tool under the setting of “no more than 50 interactors”, (A) PPI network with Cytoscape; (B) KEGG pathway analysis; (C) Gene Ontology (GO) analysis.

### Figure 7

Conserved TOP2A (A) genes and (B) protein domains in different species.