

High expression of CDCA7 predicts poor prognosis for clear cell renal cell carcinoma and explores its associations with immunity

Shouyong Liu

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Yi Wang

Affiliated Hospital of Nantong University

Chenkui Miao

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Qianwei Xing

Affiliated Hospital of Nantong University

Zengjun Wang (✉ zengjunwang@njmu.edu.cn)

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

<https://orcid.org/0000-0002-7583-4750>

Primary research

Keywords: CDCA7, Clear cell renal cell carcinoma, Prognosis, Survival, Immunity

Posted Date: February 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-143767/v2>

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

[Read Full License](#)

Version of Record: A version of this preprint was published on March 1st, 2021. See the published version at <https://doi.org/10.1186/s12935-021-01834-x>.

Abstract

Background: Cell division cycle-associated 7 (CDCA7), as a member of the cell division cycle associated family, was reported to be aberrantly expressed in both solid tumors and hematological tumors, suggesting its essential role in promoting tumorigenesis. Hence, we aimed to explore its comprehensive roles of overall survival (OS) in clear cell renal cell carcinoma (ccRCC) and emphasize its associations with immunity.

Methods: The RNA sequencing data and corresponding clinical information were downloaded from The Cancer Genome Atlas (TCGA) database. Gene set enrichment analysis (GSEA) was adopted to explore CDCA7 associated signaling pathways. Univariate and multivariate Cox regression analyses were carried out to assess independent prognostic factors. Furthermore, roles of CDCA7 in human immunity were also investigated.

Results: Our results suggested that CDCA7 was overexpressed in ccRCC and its elevated expression was related to shorter OS ($P < 0.01$). Univariate and multivariate Cox regression analyses identified CDCA7 as an independent prognostic factor (both $P < 0.05$). The prognostic nomogram integrating CDCA7 expression level and clinicopathologic variables was constructed to predict 1-, 3- and 5-year OS. GSEA indicated that high CDCA7 expression was related to the apoptosis pathway, cell cycle pathway, JAK-STAT pathway, NOD like receptor pathway, P53 pathway, T cell receptor pathway and toll like receptor pathway, etc. Moreover, CDCA7 was significantly related to microsatellite instability (MSI, $P < 0.001$) and tumor mutational burden (TMB, $P < 0.001$). As for immunity, CDCA7 was remarkably associated with immune infiltration, tumor microenvironment, immune checkpoint molecules and immune pathways.

Conclusions: CDCA7 could serve as an independent prognostic factor for ccRCC and it was closely related to MSI, TMB, and immunity.

Background

According to Cancer Statistics 2020 in the United States, renal cell carcinoma (RCC) is the 6th and 8th most common cancer in males and females respectively [1]. It is estimated that there will be 73,750 newly diagnosed cases with the male-to-female ratio approximately being 1.6:1.0, and 14,830 newly estimated deaths in 2020 [1]. Clear cell RCC (ccRCC) is the most common histological subtype, accounts for 75% of all kidney tumors [2, 3]. From 2009 to 2015, the 5-year relative survival rate for RCC in the United States is 74% [2]. In general, RCC has been increasingly recognized as a serious, worldwide public health concern.

Cell division cycle-associated 7 (CDCA7) gene, also called JPO1, located on chromosome 2q31, is a member of the cell division cycle associated family of genes and encodes a 47kDa nuclear protein consisting of 371 amino acids [4, 5]. JPO1 was first discovered and named by Dang et al. as a differentially expressed gene in fibroblasts transfected with c-Myc gene in a representational difference analysis (RDA) carried out to explore novel putative c-Myc target genes [5, 6]. Later the Human Genome Nomenclature Committee officially named the gene JPO1 as CDCA7 according to its periodic expression

in the cell cycle, reaching its peak at the G1 to S phase transition [7]. In human normal tissues, high expression levels of CDCA7 were found in small intestine, thymus, and colon whereas relatively low in bone marrow, lymph node, spleen, and peripheral leukocytes [4]. It was reported that CDCA7 was aberrantly expressed in both solid tumors and hematological tumors, including lung cancer, stomach cancer, breast cancer, colorectal cancer, lymphoma, acute myelogenous leukemia and so on [5, 8-13], suggesting its essential role in promoting tumorigenesis. CDCA7 was identified to behave as a direct c-Myc target gene [4], containing a leucine zipper motif and a cysteine rich region that suggested it could function as a DNA binding protein [5, 9]. Besides, CDCA7 was also reported to be a direct transcriptional target of transcription factor E2F1 [11]. Intriguingly, Ye et al. demonstrated that CDCA7 could increase the expression of EZH2, an important regulator in triple-negative breast cancer (TNBC), by binding the promoter of EZH2 to enhance its transcriptional activity [9]. All of these indicated the complex functional mechanisms of CDCA7 and more efforts were required to elucidate its role in tumorigenesis.

So far, however, there has been no detailed investigation of the role of CDCA7 in RCC. In the present study, we found CDCA7 was upregulated in ccRCC and significantly associated with some clinicopathologic parameters of patients via comprehensive and systematic bioinformatic analysis of RNA sequence data downloaded from The Cancer Genome Atlas (TCGA) database. CDCA7 also could serve as an independent prognostic factor of ccRCC patients. Gene Set Enrichment Analysis (GSEA) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were carried out to explore related enrichment genes and signaling pathways. Furthermore, we performed immune-related analysis of CDCA7 in ccRCC to reveal its potential function.

Materials And Methods

Data acquisition

The RNA sequence data and related clinical information of kidney renal clear cell carcinoma (KIRC) patients were searched from The Cancer Genome Atlas (TCGA) Data Portal (<http://cancergenome.nih.gov/>), containing 531 ccRCC tissues and 72 matched adjacent normal kidney tissues. International Cancer Genome Consortium (ICGC) dataset (N=45; T=91; <http://dcc.icgc.org>) and ArrayExpress dataset (E-MTAB-1980; T=99; <https://www.ebi.ac.uk/arrayexpress/>) were utilized as the external validation cohorts. Data were standardized by log₂ transformation and we set |log₂ fold change (FC)| ≥1 and false discovery rate (FDR) < 0.05 as the cut-off value.

RNA extraction, reverse transcription and quantitative real-time PCR (qRT-PCR)

16 pairs of ccRCC and corresponding adjacent normal kidney tissues were obtained from patients with primary ccRCC who had undergone radical nephrectomy at the Department of Urology of the First Affiliated Hospital of Nanjing Medical University. The clinical information of the 16 ccRCC patients was shown in **Table S1**. Total RNA was extracted from clinical samples using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and cDNA was synthesized using HiScript III RT SuperMix for qPCR (+gDNA wiper) (Vazyme, Nanjing, China) according to the manufacturer's instructions. The qRT-PCR was performed by

using StepOne Plus Real-time PCR system (Applied Biosystems, Foster City, CA, USA) with ChamQ SYBR qPCR Master Mix (High ROX Premixed) (Vazyme). The following primers were used for qRT-PCR: CDCA7, forward: 5'- GGGTGGCGATGAAGTTTCCA -3', reverse: 5'- GGGGATGTCTTCCACGGAAC -3'; β -actin, forward: 5'- CTCGCCTTTGCCGATCC -3', reverse: 5'- TTCTCCATGTCGTCCCAGTT -3'. Data analysis was performed with ABI Step One Software version 2.1 and the relative mRNA level was calculated using $2^{-\Delta\Delta C_t}$ methods.

The Kaplan Meier survival analysis and the receiver operating characteristic analysis

The median expression value of CDCA7 in the enrolled 531 ccRCC patients was set as the cut-off value and the patients were divided into a high-risk group and a low-risk group. The Kaplan Meier survival curve was plotted to analyze the different survival outcomes of the two groups. The receiver operating characteristic curves (ROC) were generated by means of the "survivalROC" package, and the area under the curve (AUC) values were calculated to assess the specificity and sensitivity of CDCA7 and associated clinicopathologic parameters including age, gender, race, grade, stage, and T, N, M.

Univariate and multivariate Cox hazard regression analysis

In order to identify significantly OS-related independent factors, univariate and multivariate Cox regression analyses were conducted to exclude clinical characteristics with little prognostic values from age, gender, ethnicity, grade, stage, T, N, M and CDCA7 expression level by R package.

Construction of nomogram model

In order to predict the possibility of OS and visualize the correlation between individual predictors (age, gender, ethnicity, grade, stage, T, M, N) and survival outcomes, we applied a nomogram model to help clinicians observe and predict the prognosis of ccRCC patients by using the R "rms" package. Points were divided to every parameter by performing the point scale in the nomogram, and we calculated the whole points by summing up the points of all factors.

Gene set enrichment analysis (GSEA)

To analyze the signaling pathways of CDCA7, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed with the help of "clusterProfiler" R package. Gene Set Enrichment Analysis (GSEA) is a powerful analytical method used to interpret gene expression data and analyze statistically significant and consistent differences between different groups with different biological states [14]. The permutation tests were carried out 1000 times for discovering significant critical biological pathway. It had been considered to be significantly enriched that nominal p value less than 0.05 and FDR less than 25%.

The evaluation of CDCA7 in microsatellite instability (MSI), tumor mutational burden (TMB) and neoantigen

With the help of MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>), we screened out all the autosomal microsatellite tracts containing five or more repeating subunits of 1-5 bp in length as described previously [15]. On the basis of the number of somatic nonsynonymous mutations (NSM), we carried out the mutation burden analysis via the comparison of sequence data between tumor tissues and its blood samples as previously described [16]. In addition, seq2HLA version 2.2 was applied to obtain 4-digit typing data for different cancers with no change of default settings in the TCGA database. Then we carried out pvac-seq to generate specific neoantigen on the samples, especially in ccRCC [17].

Correlation analysis of CDCA7 in immune infiltration and tumor microenvironment

To evaluate the performance of CDCA7 gene in immune infiltration, we performed the correlation analysis between CDCA7 and six immune cell infiltration with the help of the purity-adjusted Spearman. Moreover, we performed the ESTIMATE algorithm to assess tumor microenvironment-related scores including the estimate score, stromal score and immune score in ccRCC patients via normalized expression matrix [18]. P-values less than 0.05 were considered to be statistically significant.

CIBERSORT is a significant deconvolution algorithm used to predict the fractions of multiple cell types by analyzing gene expression profiles of admixtures [19]. We could estimate the cellular composition of the whole tissues on the basis of standardized gene expression data from TCGA, indicating the abundant specific cell types. We evaluated the expression level of each gene to figure out the composition of genes in each cell from three aspects including immune checkpoint molecules, immune pathways and mismatch repair protein, respectively.

Statistical analysis

We analyzed all of the statistical data and the figures by mean of SPSS 24.0 (IBM, Chicago, USA), R3.3.1 (<https://www.r-project.org/>) and GraphPad Prism 6.0 (San Diego, CA, USA). Pearson's correlation method was used to analyze the correlation between two different genes. We estimated the survival predictive performance of CDCA7 with the help of Kaplan-Meier curve and log-rank test. And univariate and multivariate Cox regression analyses were utilized to assess the correlation between different variables and OS. The nomogram was generated by using the rms package of R software. ROC and AUC were performed to evaluate the prognostic ability of RS by means of the package of "survivalROC" in R. It was considered to be statistically significant that nominal p value less than 0.05.

Results

Expression levels of CDCA7 in ccRCC

We analyzed the mRNA expression levels of CDCA7 in 531 ccRCC tissues and 72 normal kidney tissues from the TCGA dataset and found that CDCA7 was upregulated in ccRCC tissues compared to the normal tissues ($P < 0.001$, **Figure 1A-1B**). The pairwise boxplot of 72 pairs of ccRCC tissues and matched adjacent normal tissues from TCGA showed most of the cancer tissues exhibited a higher level of CDCA7

($P < 0.001$, **Figure 1C**). We further verified the high expression level of CDCA7 in the International Cancer Genome Consortium (ICGC) dataset as external verification ($P < 0.001$, **Figure 1D**). Furthermore, qRT-PCR results from 16 pairs of ccRCC and adjacent normal kidney tissues also exhibited a higher expression of CDCA7 in ccRCC tissues (**Figure 1E**). According to the CDCA7 expression levels of the 531 ccRCC patients, we set the median expression level as the cut-off value and divided these patients into a high- and low-risk group respectively. The Kaplan–Meier curve was plotted and showed that patients in the high-risk group had significantly poorer overall survival than those in the low-risk group ($P < 0.001$, **Figure 1F**), suggesting its potential to predict ccRCC patients' prognosis. Moreover, we plotted the Kaplan–Meier curve of CDCA7 in ccRCC patients with the help of ArrayExpress database (E-MTAB-1980) as external verification, which exhibited the same results ($P < 0.05$, **Figure 1G**). The ROC analysis was performed and its AUC for CDCA7 was 0.661 (**Figure 1H**), indicating the barely satisfactory prognosis predicted ability of CDCA7.

Association of CDCA7 expression with clinicopathologic parameters

To investigate the association between CDCA7 expression and related clinicopathologic parameters, we analyzed the CDCA7 expression levels in different groups of clinicopathologic characteristics by means of independent sample t-tests. The results showed that the CDCA7 exhibited higher expression levels in groups of higher grade ($P < 0.001$; **Figure 2A**), pathologic stage ($P < 0.001$; **Figure 2B**), T stage ($P < 0.001$; **Figure 2C**), and M stage ($P < 0.01$; **Figure 2D**). No significantly difference was observed of CDCA7 expression levels in age, gender, ethnicity, and N stage (date not shown).

Univariate and multivariate analysis of OS and construction of ccRCC prognostic prediction nomogram

We carried out univariate Cox and multivariate Cox regression analysis to investigate whether the CDCA7 expression was an independent prognostic factor correlated with the overall survival (OS) of ccRCC patients (**Table 1**). As showed in **Figure 3A**, in the univariate Cox regression analysis, CDCA7 expression, grade, age, pathological stage, T stage and M stage were all significantly associated with OS of ccRCC patients. However, in the multivariate Cox regression analysis, only CDCA7 expression, grade, pathological stage, and N stage showed significant correlation with OS of ccRCC patients (**Figure 3B**), and high CDCA7 expression predicted a poorer OS (HR= 1.125; $P < 0.001$). Based on the results above, CDCA7 could act as an independent prognostic factor of OS when adjusted by other related variables.

Then we carried out ROC curve analysis to assess the predictive ability of CDCA7 and other clinicopathologic parameters (**Figure 3C**). The AUC of the CDCA7 expression was 0.660, higher than that of age, gender, race and lymph nodes status, and lower than that of tumor grade, pathological stage, T stage and M stage. According to the AUC results, we realized that CDCA7 expression level alone could not sufficiently predict prognosis of ccRCC patients. We further established a prognostic nomogram by integrating CDCA7 and clinicopathologic parameters (**Figure 3D**). The nomogram could help to evaluate 1-, 3-, and 5-year survival probabilities to predict ccRCC patients' prognosis with a quantitative approach.

GSEA analysis of CDCA7

Though we discovered CDCA7 could act as an independent prognostic factor in ccRCC, how CDCA7 is involved in the ccRCC pathogenesis still remained unclear. We performed GSEA analysis to explore possible mechanisms and signaling pathways through which CDCA7 functioned to regulate ccRCC. On the basis of the normalized enrichment score (NES) and FDR q-val (FDR<0.01), the most significantly enriched biological pathways were exhibited (**Table 2, Figure 4**), which were apoptosis pathway, cell cycle pathway, JAK-STAT pathway, NOD like receptor pathway, P53 pathway, T cell receptor pathway and toll like receptor pathway (**Figure 4A~H**), to uncover the potential regulatory mechanism of CDCA7 in ccRCC.

PPI network construction and association of CDCA7 with MSI, TMB, Neoantigen in ccRCC

To explore the potential functional interaction of CDCA7, we constructed the PPI network by applying the STRING database and the Cytoscape software. As showed in **Figure 5A**, ten genes including SLBP, GINS2, HELLS, UHRF1, MCM2, MCM4, MCM5, NASP, TYMS, and CDC6 were significantly associated with CDCA7 functionally. Based on the ccRCC samples from TCGA database, we further investigated whether CDCA7 was relevant to MSI, TMB or neoantigen. The results suggested that CDCA7 was significantly related to MSI (P<0.001, **Figure 5B**) and TMB (P<0.001, **Figure 5C**), while it was not associated with neoantigen (P=0.95, **Figure 5D**).

Associations of CDCA7 with immune infiltrations, tumor microenvironment and methyltransferase in ccRCC

Through analyzing the correlation of CDCA7 and six immune cell infiltration levels in ccRCC via online analysis TIMER, we found that in ccRCC, CDCA7 was in close connection with the immune infiltration including B cell infiltration, CD4⁺ T cell infiltration, CD8⁺ T cell infiltration, neutrophil infiltration, macrophage infiltration, and dendritic cell infiltration (P<0.001, **Figure 6A**). As to the tumor microenvironment, CDCA7 was shown to be involved in immune cells, stromal cells and both of them (P<0.001, **Figure 6B**). In addition, the DNA methyltransferase DNMT1 (P<0.001), DNMT2 (P<0.05), DNMT3 (P<0.001), DNMT4 (P<0.001) were also significantly associated with the CDCA7 expression level in ccRCC (**Figure 6C**).

Associations of CDCA7 with immune checkpoint molecules, immune pathways and mismatch repair proteins

In this study, we analyzed the associations of CDCA7 and 47 immune checkpoint molecules in ccRCC and finally found 25 significantly related molecules including CTLA4, CD274, LAG3 and so on (**Figure 7A**). We further explored the expression levels of these 25 genes between normal kidney tissues and ccRCC tissues by utilizing the TCGA dataset and eventually identified 10 genes including LAG3, CD27, CD44, CD86, CD276, HHLA2, LAIR1, LGALS9, TIGIT, TNFRSF14 (**Figure S1**). Besides, relationships between CDCA7 and immune pathways displayed that CDCA7 was closely linked to associated immune cells like activated CD4 T cell, regulatory T cell, memory B cell, macrophage, monocyte and so on (**Figure 7B**). We also found that CDCA7 was markedly related to mismatch repair proteins including MLH1, MSH2, MSH6, PMS2 in ccRCC (**Figure 7C**).

Discussion

CDCA7 is a member of the cell division cycle associated family of genes, located on chromosome 2q31, encodes a 47kDa nuclear protein consisting of 371 amino acids [4, 5] and is involved in embryonic development [20]. It was found that CDCA7 was expressed in many human tissues including small intestine, thymus, colon, bone marrow, lymph node, spleen, and peripheral leukocytes [4]. Previous researches had reported its close correlations to malignant tumors. As reported, CDCA7 was markedly upregulated in TNBC, the most aggressive subtype of breast cancer, related to tumor proliferation and metastatic relapse status, and predicted poor prognosis [9]. Wang etc. discovered that CDCA7 was preferentially elevated in lung adenocarcinoma (LUAD) and overexpression of CDCA7 could enhance cell proliferation in LUAD through G1 phase promotion [8]. And CDCA7 was essential for invasion and migration of lymphoma cells [10]. However, up to date, no data had been reported about the expression level and potential function of CDCA7 in ccRCC.

In our study, we carried out a systematic analysis on CDCA7 in ccRCC by analyzing the RNA sequence data downloaded from TCGA database. Comparing to the expression level of CDCA7 in normal kidney tissues, we found CDCA7 was elevated in ccRCC tissues. The enrolled ccRCC patients from TCGA were divided into a high- and low-expression group according to the median CDCA7 expression value of all the patients. And we discovered that patients in the low-expression group had longer OS than those in high. Meantime, higher expression of CDCA7 was significantly associated with higher disease grade, stage, T and M period. Then we conducted univariate and multivariate Cox regression analysis on CDCA7 and found that CDCA7 could serve as an independent prognostic factor of ccRCC. Moreover, to help clinicians predict the prognosis of ccRCC patients, we constructed a predictive nomogram based on the CDCA7 expression level and relative clinicopathological parameters.

Gene Set Enrichment Analysis (GSEA) is a computational method utilized to evaluate whether a prior defined set of genes display statistically significant, consistent differences between two biological states [21]. To explore possible signaling pathways and mechanisms CDCA7 could functioned through, we conducted GSEA analysis and discovered signal pathways CDCA7 might be involved in including apoptosis pathway, cell cycle pathway, JAK-STAT pathway, NOD like receptor pathway, P53 pathway, T cell receptor pathway and toll like receptor pathway. Previous researches indicated that down-expression of A100A4 could reduce cell growth of RCC via NF- κ B-dependent MMP-2 and bcl-2 Pathway [22]. Fang etc. discovered that simvastatin could lead to the inhibition of cell growth of RCC via AKT/mTOR, ERK and JAK2/STAT3 pathway [23]. It was also reported that Tropomyosin-1, a widely expressed actin-binding protein, could promote cancer cell apoptosis via the p53-mediated mitochondrial pathway in ccRCC [24], and the MDM2 inhibitor MI-319 could induce RCC cell apoptosis mainly dependent on p53 overexpression [25].

Then we constructed the PPI network of CDCA7 to explore potential protein-protein interaction. It showed that ten genes were significantly functional associated with CDCA7, including SLBP, GINS2, HELLS, UHRF1, MCM2, MCM4, MCM5, NASP, TYMS, and CDC6. Researchers have discovered the important roles

of these genes in the physiological process and tumor progression. For instance, SLBP (Stem-loop-binding protein) is evolutionarily conserved and involved in the processing, translation, and degradation of canonical histones mRNAs including H1, H2A, H2B, H3, and H4 [26]. With regard to GINS2, it was reported that GINS2 could promote cancer cell proliferation, migration and invasion of non-small-cell lung cancer (NSCLC) via facilitating epithelial-to-mesenchymal transition and modulating PI3K/Akt and MEK/ERK signal pathway [27]. Previous reports revealed that mutations in CDCA7 and HELLS respectively could cause immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome types 3 and 4 [28]. And the ZBTB24-CDCA7 axis could facilitate DNA methylation by regulating HELLS enrichment at centromeric satellite repeats [29]. As to the minichromosome maintenance (MCM) proteins family, they were mainly involved in the initiation and elongation of DNA replication [30] and the formation of the pre-replicative complex (preRC), the replication fork and the initial steps of DNA synthesis [31]. MCM2, MCM4, and MCM5 served as essential components of a hexameric, ring-shaped complex that acted as one of the pre-replication factors and were related to the chromatin and the proteins of the origin recognition complex at the M-G1 transition [30]. The important roles of these genes in the PPI network of CDCA7 all indicated the possible essential functions of CDCA7.

We also discovered the association of CDCA7 with immunity. CDCA7 was in close connection with the immune infiltration including B cell infiltration, CD4⁺ T cell infiltration, CD8⁺ T cell infiltration, neutrophil infiltration, macrophage infiltration, and dendritic cell infiltration. Moreover, CDCA7 was significantly correlated with tumor-related immunosuppressive molecules including PDCD1, CD274, CTLA4, BTLA and LAG3. PDCD1 (programmed cell death 1), CTLA-4 (cytotoxic T-lymphocyte-associated antigen 4) and BTLA (B and T lymphocyte attenuator) are members of immunoglobulin-related receptors family associated with various aspects of T cell immune regulation [32]. PDCD1 (PD-1, CD279) and CD274 (PD-L1) axis has been discovered as a worthy therapeutic target for its important role not only in physiological immune homeostasis, but also in the way through which cancer cells evade the immune system [33]. Overexpression of PD-1 and PD-L1 in tumors is closely correlated with poor disease outcome in some human cancers [34]. The researches of PD-1 or PD-L1 inhibitors have greatly promoted the development of the treatment of cancer [35]. As to CTLA4, it is a transmembrane receptor with inhibitory function expressed by T lymphocytes. CTLA4 could inhibit costimulation through competing for CD28 ligands [36]. In addition to CTLA4 and PD1, BTLA is also an immune checkpoint related to suppress immune responses [37], containing two immunoreceptor tyrosine-based inhibitory motifs in its cytoplasmic region. BTLA could modulate T cell responses and attenuate B cell function to play its inhibitory roles in multiple diseases [38, 39]. Lymphocyte Activation Gene-3 (LAG3, CD223) served as another potential cancer immunotherapeutic target. It could inhibit T cells function and mediate a state of exhaustion in combination with PD1 [40]. The association of CDCA7 with these tumor-related immunosuppressive molecules suggested its important potential functions in human immunity.

Before fully understanding this article, several limitations should not be ignored. Firstly, clinical information was limited, due to the retrospective data from TCGA. Secondly, the sample sizes of normal renal tissue samples were relatively small (N = 72) in TCGA and this might lead to some biased in our

conclusions. Thirdly, we currently could not tell the possible mechanism of CDCA7. Is it tumor intrinsic (related to tumor proliferation, migration, invasion?) or extrinsic (related to immune suppression?) or both? Future studies with larger sample sizes and sufficient clinical information were required to correct our results.

To conclude, our study elucidated that CDCA7 could act as a favorable prognostic factor for ccRCC. Moreover, apoptosis pathway, cell cycle pathway, JAK-STAT pathway, NOD like receptor pathway, P53 pathway, T cell receptor pathway and toll like receptor pathway might be the primary pathways regulated by CDCA7. Furthermore, CDCA7 could serve as an independent prognostic factor for ccRCC and it was closely related to MSI, TMB, and immunity. Further advanced researches in vivo and in vitro were required to verify our findings.

Declarations

Funding

This study was supported by the National Natural Science Foundation of China (81771640).

Author's Contribution

ZJ.W: Conceptualization, Methodology, Project administration;

QW.X: Data collection and management, Formal analysis;

CK.M: Formal analysis, Writing - Review & Editing;

SY.L, Y.W: Software, Writing - Original Draft;

Conflicts of Interest

The authors confirm that there are no conflicts of interest.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by Ethics Committee of The First Affiliated Hospital of Nanjing Medical University.

Data Availability Statement

The RNA-sequencing data and corresponding clinical information were downloaded from the Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>).

Acknowledgements

We would like to thank the researchers and study participants for their contributions.

References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA Cancer J Clin* 2020, **70**(1):7-30.
2. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM: **The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours**. *European urology* 2016, **70**(1):93-105.
3. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Znaor A, Bray F: **Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods**. *Int J Cancer* 2019, **144**(8):1941-1953.
4. Prescott JE, Osthus RC, Lee LA, Lewis BC, Shim H, Barrett JF, Guo Q, Hawkins AL, Griffin CA, Dang CV: **A novel c-Myc-responsive gene, JPO1, participates in neoplastic transformation**. *The Journal of biological chemistry* 2001, **276**(51):48276-48284.
5. Osthus RC, Karim B, Prescott JE, Smith BD, McDevitt M, Huso DL, Dang CV: **The Myc target gene JPO1/CDCA7 is frequently overexpressed in human tumors and has limited transforming activity in vivo**. *Cancer research* 2005, **65**(13):5620-5627.
6. Lewis BC, Shim H, Li Q, Wu CS, Lee LA, Maity A, Dang CV: **Identification of putative c-Myc-responsive genes: characterization of rcl, a novel growth-related gene**. *Molecular and cellular biology* 1997, **17**(9):4967-4978.
7. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO *et al*: **Identification of genes periodically expressed in the human cell cycle and their expression in tumors**. *Molecular biology of the cell* 2002, **13**(6):1977-2000.
8. Wang H, Ye L, Xing Z, Li H, Lv T, Liu H, Zhang F, Song Y: **CDCA7 promotes lung adenocarcinoma proliferation via regulating the cell cycle**. *Pathology, research and practice* 2019, **215**(11):152559.

9. Ye L, Li F, Song Y, Yu D, Xiong Z, Li Y, Shi T, Yuan Z, Lin C, Wu X *et al*: **Overexpression of CDCA7 predicts poor prognosis and induces EZH2-mediated progression of triple-negative breast cancer.** *International journal of cancer* 2018, **143**(10):2602-2613.
10. Martín-Cortázar C, Chiodo Y, Jiménez RP, Bernabé M, Cayuela ML, Iglesias T, Campanero MR: **CDCA7 finely tunes cytoskeleton dynamics to promote lymphoma migration and invasion.** *Haematologica* 2020, **105**(3):730-740.
11. Goto Y, Hayashi R, Muramatsu T, Ogawa H, Eguchi I, Oshida Y, Ohtani K, Yoshida K: **JPO1/CDCA7, a novel transcription factor E2F1-induced protein, possesses intrinsic transcriptional regulator activity.** *Biochimica et biophysica acta* 2006, **1759**(1-2):60-68.
12. Gill RM, Gabor TV, Couzens AL, Scheid MP: **The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation.** *Molecular and cellular biology* 2013, **33**(3):498-513.
13. Li D, Jiang X, Zhang X, Cao G, Wang D, Chen Z: **Long noncoding RNA FGD5-AS1 promotes colorectal cancer cell proliferation, migration, and invasion through upregulating CDCA7 via sponging miR-302e.** *In vitro cellular & developmental biology Animal* 2019, **55**(8):577-585.
14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(43):15545-15550.
15. Timmermann B, Kerick M, Roehr C, Fischer A, Isau M, Boerno ST, Wunderlich A, Barmeyer C, Seemann P, Koenig J *et al*: **Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis.** *PLoS One* 2010, **5**(12):e15661.
16. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, Schrock A, Campbell B, Shlien A, Chmielecki J *et al*: **Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden.** *Genome Med* 2017, **9**(1):34.
17. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, Griffith M: **pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens.** *Genome Med* 2016, **8**(1):11.
18. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA *et al*: **Inferring tumour purity and stromal and immune cell admixture from expression data.** *Nat Commun* 2013, **4**:2612.
19. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.** *Nat Methods* 2015, **12**(5):453-457.
20. Guiu J, Bergen DJ, De Pater E, Islam AB, Ayllón V, Gama-Norton L, Ruiz-Herguido C, González J, López-Bigas N, Menendez P *et al*: **Identification of Cdca7 as a novel Notch transcriptional target involved in hematopoietic stem cell emergence.** *J Exp Med* 2014, **211**(12):2411-2423.
21. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP: **GSEA-P: a desktop application for Gene Set Enrichment Analysis.** *Bioinformatics* 2007, **23**(23):3251-3253.

22. Yang XC, Wang X, Luo L, Dong DH, Yu QC, Wang XS, Zhao K: **RNA interference suppression of A100A4 reduces the growth and metastatic phenotype of human renal cancer cells via NF- κ B-dependent MMP-2 and bcl-2 pathway.** *Eur Rev Med Pharmacol Sci* 2013, **17**(12):1669-1680.
23. Fang Z, Tang Y, Fang J, Zhou Z, Xing Z, Guo Z, Guo X, Wang W, Jiao W, Xu Z *et al*: **Simvastatin inhibits renal cancer cell growth and metastasis via AKT/mTOR, ERK and JAK2/STAT3 pathway.** *PLoS One* 2013, **8**(5):e62823.
24. Tang C, Wang J, Wei Q, Du YP, Qiu HP, Yang C, Hou YC: **Tropomyosin-1 promotes cancer cell apoptosis via the p53-mediated mitochondrial pathway in renal cell carcinoma.** *Oncol Lett* 2018, **15**(5):7060-7068.
25. Liu QJ, Shen HL, Lin J, Xu XH, Ji ZG, Han X, Shang DH, Yang PQ: **Synergistic roles of p53 and HIF1 α in human renal cell carcinoma-cell apoptosis responding to the inhibition of mTOR and MDM2 signaling pathways.** *Drug Des Devel Ther* 2016, **10**:745-755.
26. Marzluff WF, Wagner EJ, Duronio RJ: **Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail.** *Nat Rev Genet* 2008, **9**(11):843-854.
27. Liu X, Sun L, Zhang S, Zhang S, Li W: **GIN52 facilitates epithelial-to-mesenchymal transition in non-small-cell lung cancer through modulating PI3K/Akt and MEK/ERK signaling.** *J Cell Physiol* 2019.
28. Thijssen PE, Ito Y, Grillo G, Wang J, Velasco G, Nitta H, Unoki M, Yoshihara M, Suyama M, Sun Y *et al*: **Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome.** *Nat Commun* 2015, **6**:7870.
29. Hardikar S, Ying Z, Zeng Y, Zhao H, Liu B, Veland N, McBride K, Cheng X, Chen T: **The ZBTB24-CDCA7 axis regulates HELLS enrichment at centromeric satellite repeats to facilitate DNA methylation.** *Protein Cell* 2020, **11**(3):214-218.
30. Dubois ML, Bastin C, Lévesque D, Boisvert FM: **Comprehensive Characterization of Minichromosome Maintenance Complex (MCM) Protein Interactions Using Affinity and Proximity Purifications Coupled to Mass Spectrometry.** *J Proteome Res* 2016, **15**(9):2924-2934.
31. Giaginis C, Vgenopoulou S, Vielh P, Theocharis S: **MCM proteins as diagnostic and prognostic tumor markers in the clinical setting.** *Histol Histopathol* 2010, **25**(3):351-370.
32. Rowshanravan B, Halliday N, Sansom DM: **CTLA-4: a moving target in immunotherapy.** *Blood* 2018, **131**(1):58-67.
33. Sun C, Mezzadra R, Schumacher TN: **Regulation and Function of the PD-L1 Checkpoint.** *Immunity* 2018, **48**(3):434-452.
34. Ohaegbulam KC, Assal A, Lazar-Molnar E, Yao Y, Zang X: **Human cancer immunotherapy with antibodies to the PD-1 and PD-L1 pathway.** *Trends Mol Med* 2015, **21**(1):24-33.
35. Ribas A, Wolchok JD: **Cancer immunotherapy using checkpoint blockade.** *Science* 2018, **359**(6382):1350-1355.
36. Khailaie S, Rowshanravan B, Robert PA, Waters E, Halliday N, Badillo Herrera JD, Walker LSK, Sansom DM, Meyer-Hermann M: **Characterization of CTLA4 Trafficking and Implications for Its Function.** *Biophys J* 2018, **115**(7):1330-1343.

37. Pardoll DM: **The blockade of immune checkpoints in cancer immunotherapy.** *Nat Rev Cancer* 2012, **12**(4):252-264.
38. Watanabe N, Gavrieli M, Sedy JR, Yang J, Fallarino F, Loftin SK, Hurchla MA, Zimmerman N, Sim J, Zang X *et al.* **BTLA is a lymphocyte inhibitory receptor with similarities to CTLA-4 and PD-1.** *Nat Immunol* 2003, **4**(7):670-679.
39. Iwata A, Watanabe N, Oya Y, Owada T, Ikeda K, Suto A, Kagami S, Hirose K, Kanari H, Kawashima S *et al.* **Protective roles of B and T lymphocyte attenuator in NKT cell-mediated experimental hepatitis.** *J Immunol* 2010, **184**(1):127-133.
40. Turnis ME, Andrews LP, Vignali DA: **Inhibitory receptors as targets for cancer immunotherapy.** *Eur J Immunol* 2015, **45**(7):1892-1905.

Tables

Table 1: Univariate and multivariate Cox regression analysis of CDCA7 expression level and clinicopathologic variables in ccRCC;

Variable	Univariate analysis				Multivariate analysis ^c			
	HR	HR.95L	HR.95H	pvalue	HR	HR.95L	HR.95H	pvalue
Age(year)	1.0333	1.0197	1.0471	0.0000	1.0399	1.0249	1.0551	0.0000
Gender	0.9333	0.6797	1.2815	0.6696	0.9843	0.7090	1.3665	0.9245
Ethnicity	1.1931	0.7160	1.9881	0.4981	1.2535	0.7149	2.1977	0.4303
Grade	1.9669	1.6388	2.3606	0.0000	1.3483	1.0755	1.6903	0.0096
Stage	1.8556	1.6436	2.0950	0.0000	1.5974	1.1365	2.2454	0.0070
T	1.9976	1.6891	2.3625	0.0000	1.1300	0.8606	1.4836	0.3791
M	2.0996	1.6607	2.6546	0.0000	0.9606	0.5203	1.7735	0.8978
N	0.8630	0.7389	1.0079	0.0628	0.8328	0.7084	0.9790	0.0267
CDCA7	1.2742	1.1930	1.3609	0.0000	1.2754	1.1727	1.3871	0.0000

Table 2: Gene set enrichment analysis (GSEA) of CDCA7 in ccRCC;

GeneSet name	NES	Nominal p-value	FDR q-value
KEGG_APOPTOSIS	1.8310	0.0231	0.0556
KEGG_CELL_CYCLE	2.4277	0.0000	0.0008
KEGG_JAK_STAT_SIGNALING_PATHWAY	2.0902	0.0000	0.0115
KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	2.0426	0.0038	0.0153
KEGG_P53_SIGNALING_PATHWAY	2.2361	0.0000	0.0043
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	2.1507	0.0039	0.0090
KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	2.0946	0.0038	0.0132

Figures

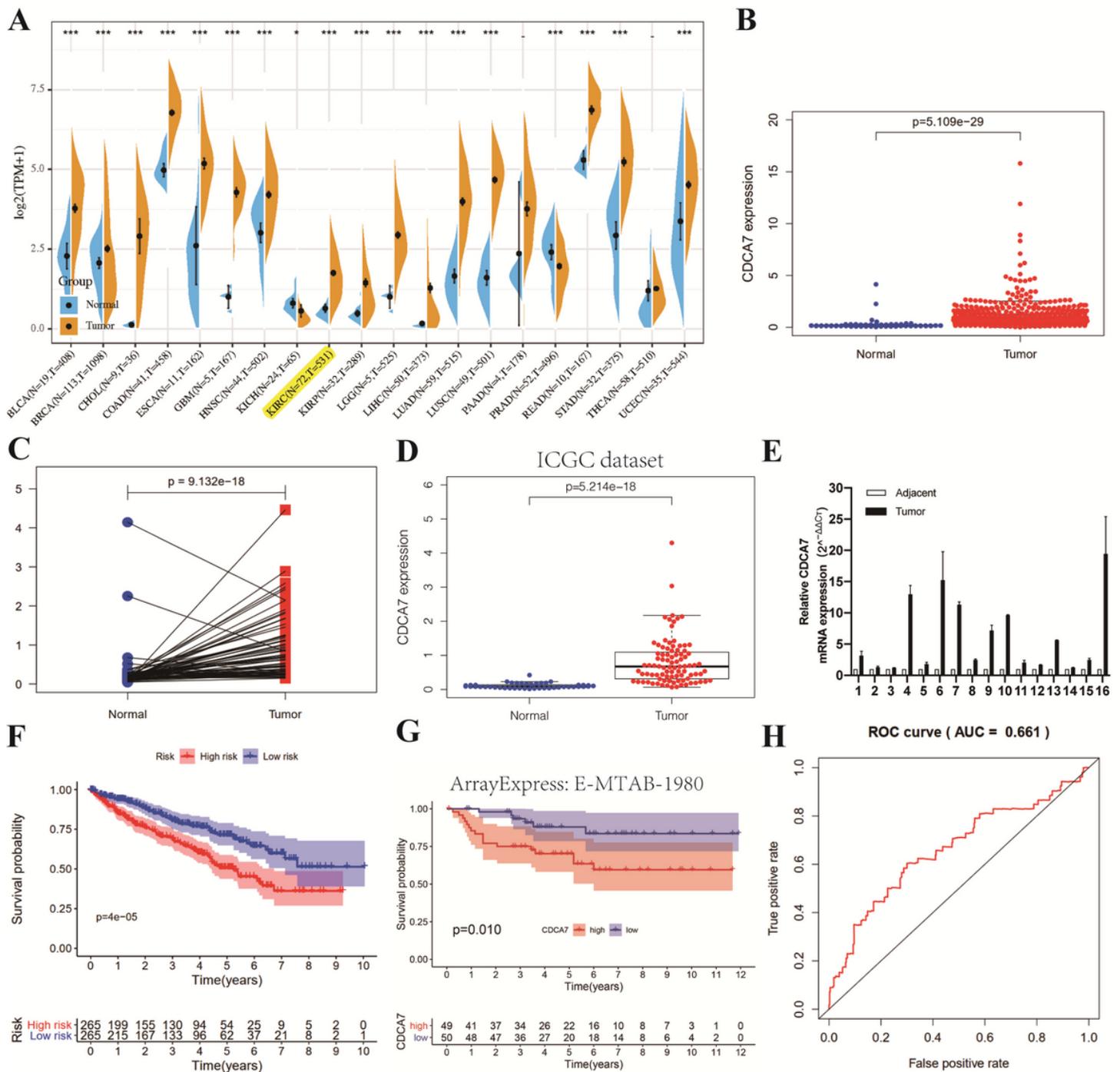


Figure 1

CDCA7 was overexpressed in ccRCC and associated with overall survival (OS). (A) The expression of CDCA7 in various cancers in TCGA dataset; (B) Relative expression levels of CDCA7 in ccRCC and normal kidney tissues in TCGA dataset (N=72; T=531); (C) Pairwise boxplot of CDCA7 expression between ccRCC and matched normal tissues in TCGA dataset (N=72; T=72); (D) External validation of relative expression levels of CDCA7 in ccRCC and normal kidney tissues in ICGC dataset (N=45; T=91); (E) CDCA7 mRNA expression in 16 pairs of ccRCC tissues and matched adjacent non-cancerous tissues; (F) Kaplan-Meier curve for low-risk and high-risk groups in TCGA database; (G) Kaplan-Meier curve for low-risk and high-risk groups in ArrayExpress database (E-MTAB-1980; T=99); (H) ROC curve of CDCA7.

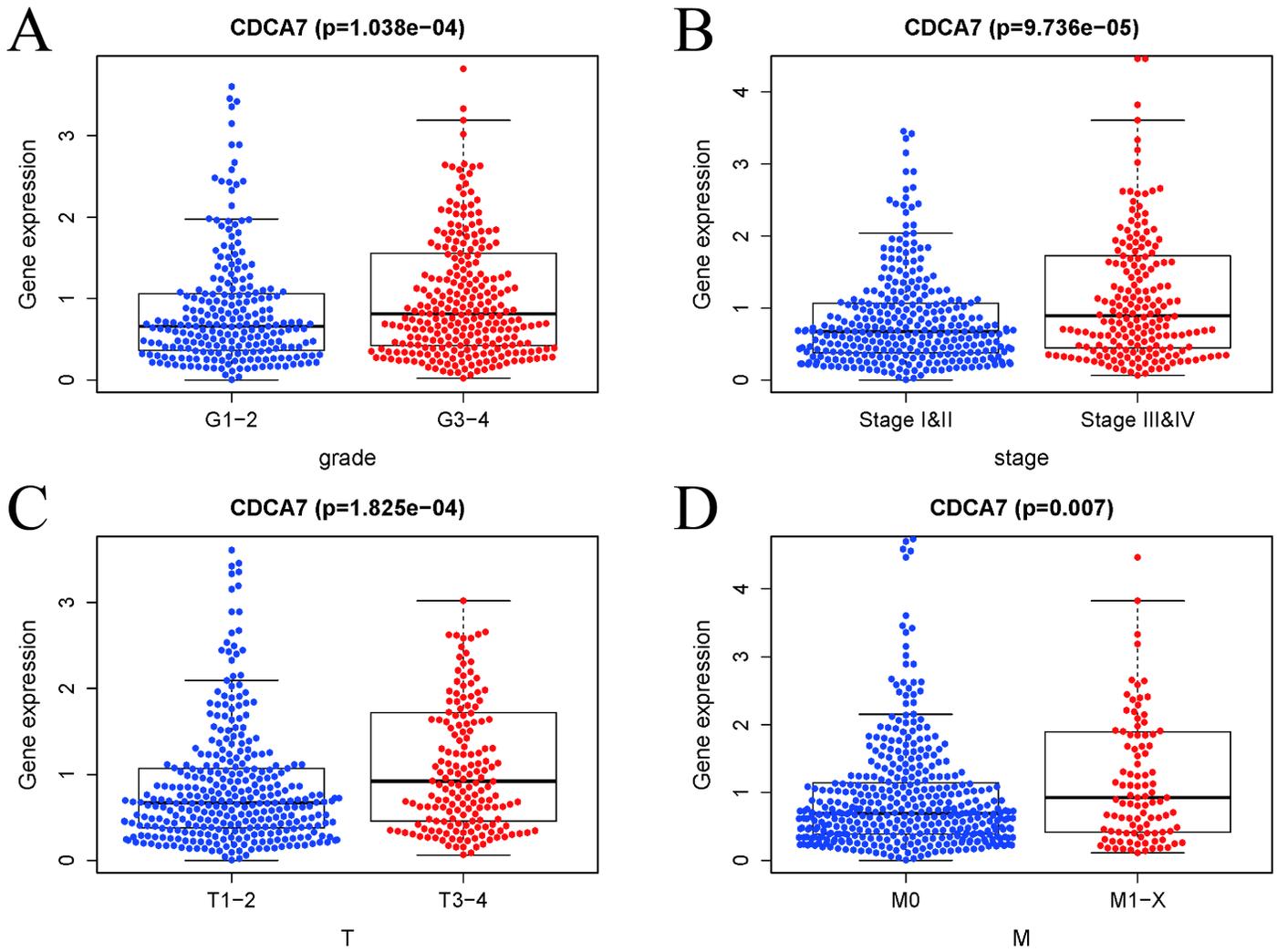


Figure 2

Association of CDCA7 expression with clinicopathologic characteristics. (A) Grade; (B) Stage; (C) T stage; (D) M stage.

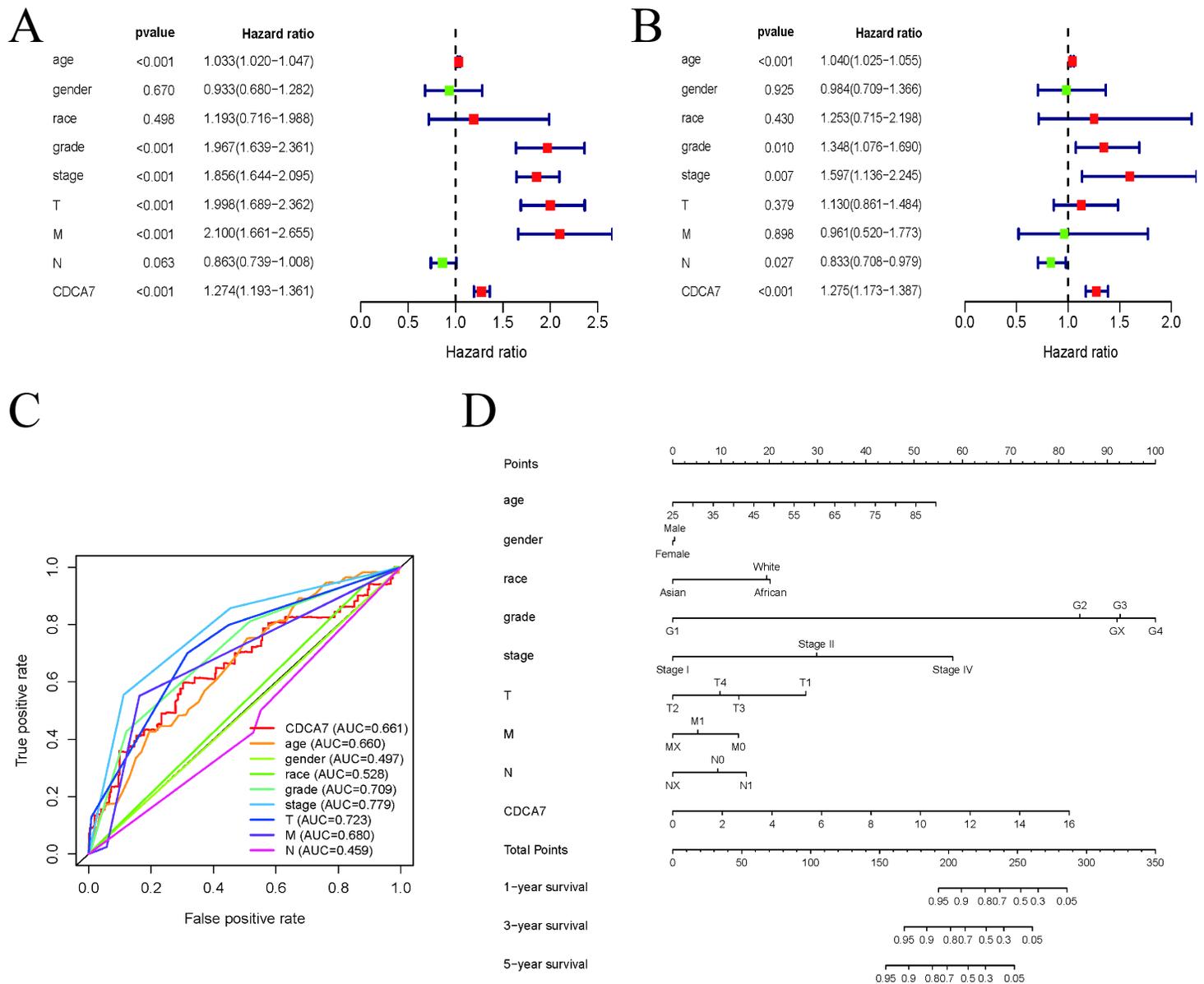


Figure 3

CDCA7 could serve as an independent prognostic factor and a prognostic nomogram was constructed. (A, B) Univariate and multivariate Cox regression analysis of clinicopathologic variables and CDCA7 in ccRCC; (C) ROC curves of CDCA7 and other clinicopathologic parameters; (D) The prognostic nomogram constructed to predict 1-, 3-, 5-year survival of ccRCC patients. ROC: the receiver operating characteristic curves.

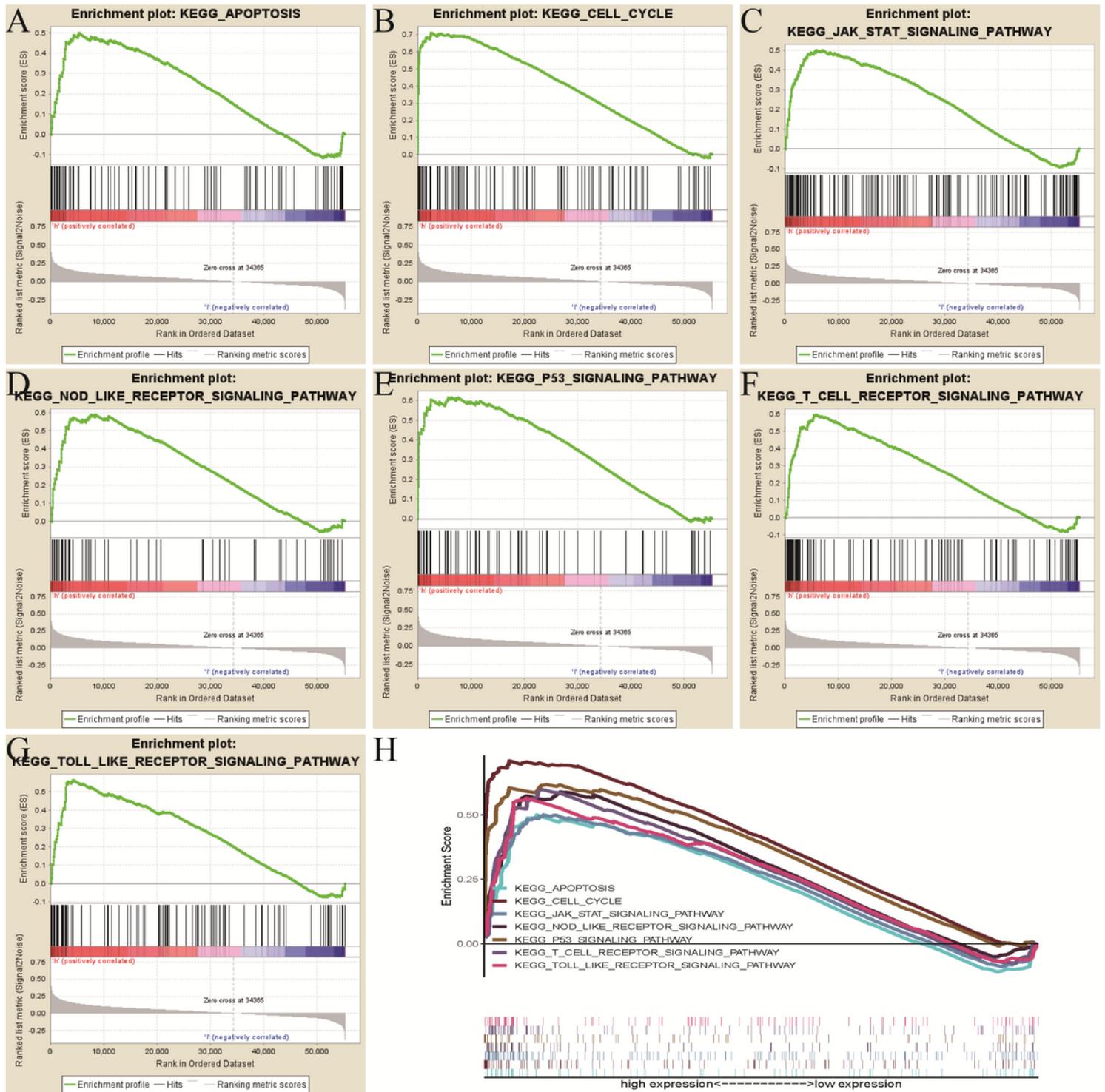


Figure 4

Enrichment plots from gene set enrichment analysis (GSEA). (A) Apoptosis pathway; (B) Cell cycle pathway; (C) JAK-STAT pathway; (D) NOD like receptor pathway; (E) P53 pathway; (F) T cell receptor pathway; (G) Toll like receptor pathway; (H) The seven most significantly enriched signaling pathways based on their normalized enrichment score and the expression map.

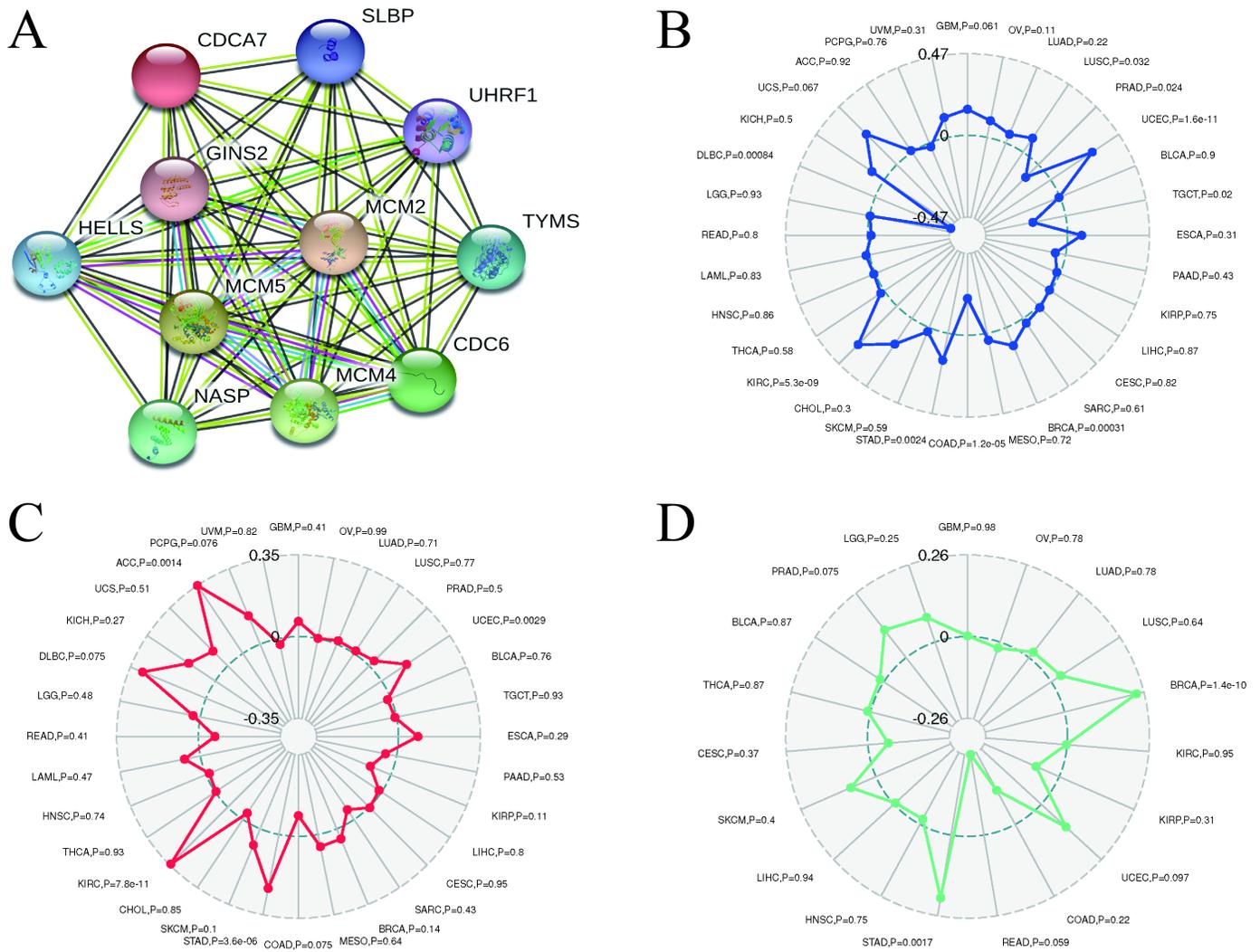


Figure 5

Protein-protein interaction network of CDCA7 (A) and associations of CDCA7 with MSI (B), TMB (C), and neoantigen (D). MSI: microsatellite instability; TMB: tumor mutational burden.

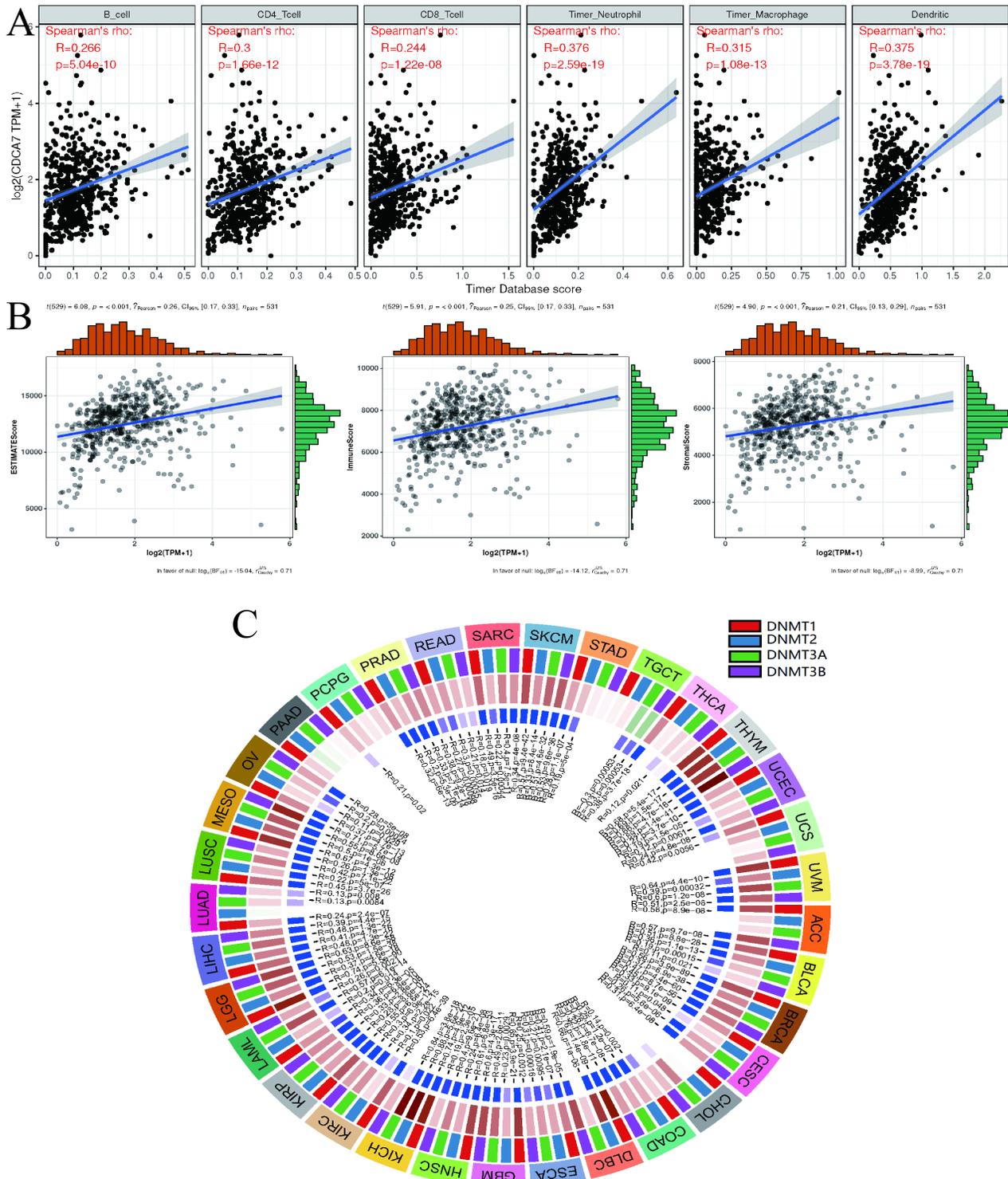


Figure 6

Associations of CDCA7 with immune infiltrations (A), immune microenvironment including immune cells, stromal cells and both of them (B), and methyltransferase (C).

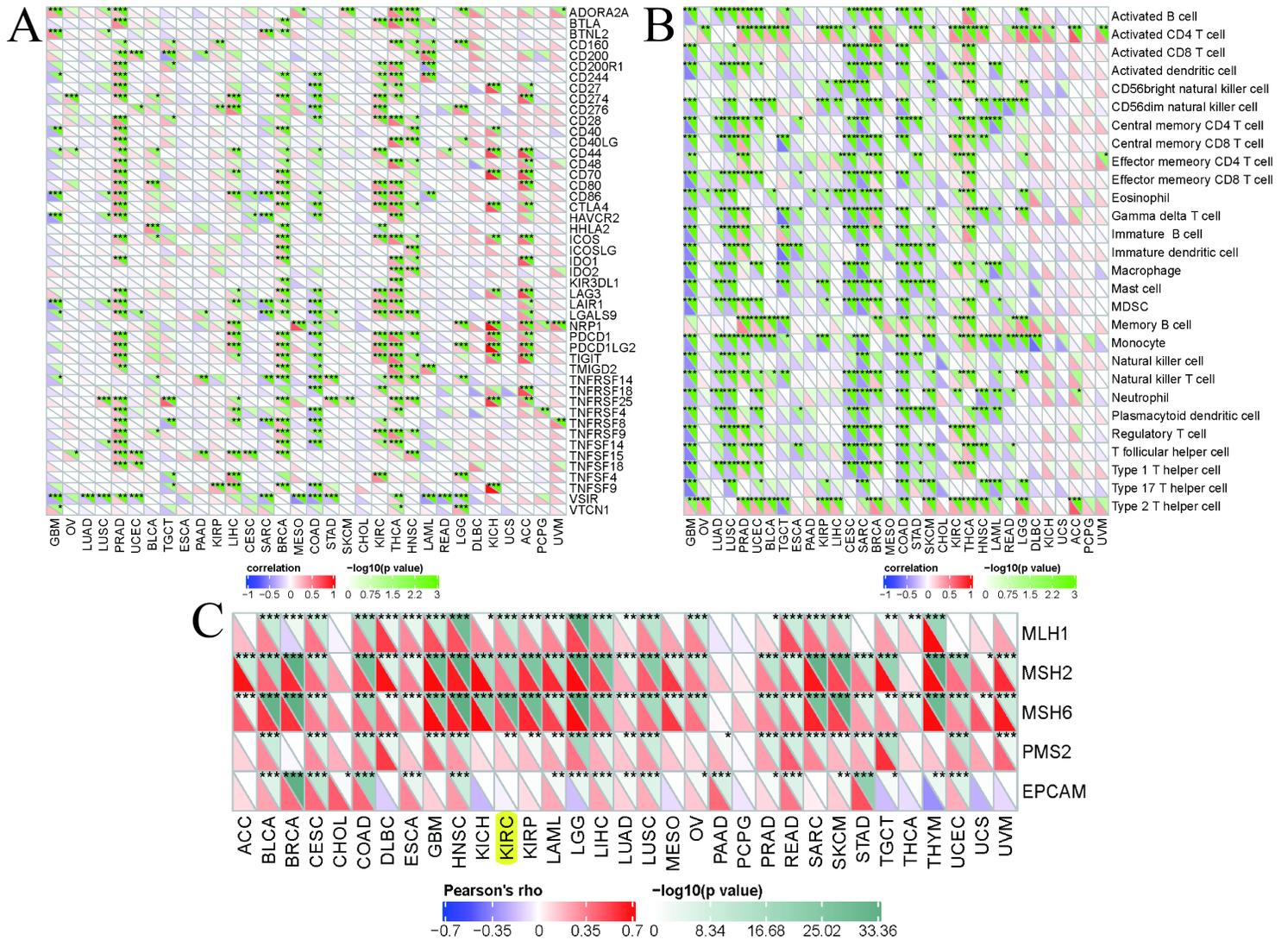


Figure 7

Associations of CDCA7 with immune checkpoint molecules (A), immune cells (B), and mismatch repair proteins (C).

Supplementary Files

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

- [TableS1.docx](#)
- [FigureS1.tif](#)