



OPEN

CCDC58 is a potential biomarker for diagnosis, prognosis, immunity, and genomic heterogeneity in pan-cancer

Kai Yang^{1,7}, Yan Ma^{2,7}, Weigang Chen^{1,7}, Lu Liu^{3,7}, Zelong Yang¹, Chaokui He⁴, Nanbei Zheng⁵, Xinyu Liu¹, Xin Cheng¹, Junbo Song¹, Yong Chen¹✉, Hongyu Qiao⁶✉ & Ruohan Zhang¹✉

Coiled-coil domain-containing 58 (CCDC58) is a member of the CCDC protein family. Similar to other members, CCDC58 exhibits potential tumorigenic roles in a variety of malignancies. However, there is no systematic and comprehensive pan-cancer analysis to investigate the diagnosis, prognosis, immune infiltration, and other related functions of CCDC58. We used several online websites and databases, such as TCGA, GTEx, UALCAN, HPA, CancerSEA, BioGRID, GEPIA 2.0, TIMER 2.0, and TISIDB, to extract CCDC58 expression data and clinical data of patients in pan-cancer. Then, the relationship between CCDC58 expression and diagnosis, prognosis, genetic alterations, DNA methylation, genomic heterogeneity, and immune infiltration level were determined. In addition, the biological function of CCDC58 in liver hepatocellular carcinoma (LIHC) was investigated. Pan-cancer analysis results showed that CCDC58 was differentially expressed in most tumors and showed excellent performance in diagnosis and prediction of prognosis. The expression of CCDC58 was highly correlated with genetic alterations, DNA methylation, and genomic heterogeneity in some tumors. In addition, the correlation analysis of CCDC58 with the level of immune infiltration and immune checkpoint marker genes indicated that CCDC58 might affect the composition of the tumor immune microenvironment. Enrichment analysis showed that CCDC58-related genes were mainly linked to mitosis, chromosome, and cell cycle. Finally, biological function experiments demonstrated that CCDC58 plays an important role in tumor cell proliferation and migration. CCDC58 was first identified as a pan-cancer biomarker. It may be used as a potential therapeutic target to improve the prognosis of patients in the future.

Keywords CCDC58, Pan-cancer, Biomarker, Immune infiltration, LIHC

Abbreviations

CCDC	Coiled-coil domain-containing
LIHC	Liver hepatocellular carcinoma
OS	Overall survival
DSS	Disease-specific survival
CI	Confidence intervals
ROC	Receiver operating characteristic
AUC	Area under curve
TMB	Tumor mutation burden
MATH	Mutant-allele tumor heterogeneity

¹Department of Hepatobiliary Surgery, Xi Jing Hospital, Air Force Medical University, Xi'an 710032, China. ²Department of Gynecology and Obstetrics, Xi Jing Hospital, Air Force Medical University, Xi'an 710032, China. ³College of Life Sciences, Northwest University, Xi'an 710000, China. ⁴Department of Oncology, The First Affiliated Hospital of Shanxi Medical University, Taiyuan 030000, China. ⁵Department of General Surgery, The First Affiliated Hospital of Jiamusi University, Jiamusi 154002, China. ⁶Department of Pediatrics, Xi Jing Hospital, Air Force Medical University, Xi'an 710032, China. ⁷These authors contributed equally: Kai Yang, Yan Ma, Weigang Chen, and Lu Liu. ✉email: gdwkcy@163.com; qiaohongyu.1990@163.com; zhangrh1990@outlook.com

MSI	Microsatellite instability
NEO	Neoantigen
HRD	Homologous recombination deficiency
LOH	Loss of heterozygosity
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genome
DCA	Decision curve analysis

Coiled-coil domain was a highly conserved supercoiled protein motif with several alpha-helices, usually two to six. It was estimated that about 10% of proteins contain coiled-coil domains^{1,2}. Coiled-coil domain-containing (CCDC) proteins were expressed in many tissues, and involved in many important physiological functions³. At present, it had been reported that CCDC proteins took key roles in intracellular signal transduction, genetic signal transcription, cell cycle regulation, and other functions⁴. In recent years, more and more scholars paid attention to the expression and effect of CCDC proteins in malignant tumors. For example, Zhen Liu et al.⁵ pointed out that CCDC19 might play a tumor suppressor role by inhibiting the proliferation, invasion, and migration of nasopharyngeal carcinoma cells. It had been suggested that CCDC67 was a tumor suppressor gene that was silenced by promoter CpG methylation in gastric cancer, and it might be involved in cell signaling and migration related to tumorigenesis⁶. With literature, circ-CCDC66 was up-regulated in renal cancer cell lines and could enhance the enrichment of tumor stem cells⁷. In addition, CCDC85B promotes the proliferation and invasion of non-small cell lung cancer cells by activating the AKT/GSK3 β / β -catenin oncogenic signaling pathway⁸. It had been confirmed that overexpression of the CCDC34 protein in esophageal squamous cell carcinoma was associated with tumor progression, angiogenesis, and poor survival⁹. In summary, studies had confirmed that some CCDC family proteins were involved in the occurrence and development of tumors.

The CCDC58 gene, also known as Mitochondrial Matrix Import Factor 23 (MIX23), was located on chromosome 3q21.1. It encodes CCDC58 protein, including 144 amino acids, with a molecular weight of 16620 Da. CCDC58 protein belonged to intracellular proteins located in mitochondria¹⁰. Previous studies indicated that the knockdown of CCDC58 could enhance Ca²⁺ retention and bioenergetics in mitochondria¹¹. Other scholars found that CCDC58 was significantly up-regulated when mitochondrial input was interfered with, suggesting that CCDC58 might be a stabilizer or regulator of the mitochondrial protein input mechanism¹². In triple-negative breast cancer, a miR-432-p/CCDC58 regulatory axis influenced tumor proliferation, migration, and invasion phenotypes¹³. CCDC58 had also been identified as a gene associated with tumor progression in endometrial cancer and ovarian cancer^{14,15}. Among the differentially expressed genes between diabetic patients and normal controls, CCDC58 ranked in the top ten, suggesting that CCDC58 might be related to the occurrence and development of diabetes¹⁶.

Given the lack of studies on CCDC58 and its potential role in tumorigenesis, we performed a pan-cancer analysis of CCDC58. Based on several reliable databases and combined with statistical analysis, we showed diverse results of CCDC58 expression, diagnostic and prognostic value, genomic heterogeneity, immune infiltration analysis, enrichment analysis, and so on. Finally, we explored the role of CCDC58 in progression of liver cancer cells, analyzed its correlation with clinical indicators, and established a prognostic nomogram in liver hepatocellular carcinoma (LIHC). It was concluded that CCDC58 was a reliable pan-cancer biomarker, and it had good diagnostic and prognostic value in a variety of cancers. The analysis and experimental results of this study provided a direction for tumor treatment.

Materials and methods

CCDC58 expression analysis

We downloaded RNA-seq data of pan-cancer in transcripts per million (TPM) format of The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) Project uniformly processed by Toil¹⁷ process from UCSC XENA (<https://xenabrowser.net/datapages/>)¹⁸. Types of cancer included adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangio carcinoma (CHOL), colon adenocarcinoma (COAD), lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), 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), acute myeloid leukemia (LAML), brain lower grade glioma (LGG), LIHC, lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), testicular germ cell tumors (TGCT), thyroid carcinoma (THCA), thymoma (THYM), uterine corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS) and uveal melanoma (UVM). The log₂(TPM + 1) data was applied for the log scale. R “stats” and “car” packages were used for statistical analysis according to the data characteristics, and R “ggplot2” package was used to visualize the statistical results^{19–21}. The results of the level of protein expression and the relationship between CCDC58 expression and individual cancer stage were obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) of the University of Alabama at Birmingham CANcer data analysis Portal (UALCAN, <http://ualcan.path.uab.edu/analysis-prot.html>)²². The Human Protein Atlas (HPA, <https://www.proteinatlas.org/>)²³ was further employed to show the expression of CCDC58 in tumor and normal tissue.

Survival and ROC curve analysis

The Kaplan–Meier (K–M) survival curves were used to display the correlation between the expression of CCDC58 and Overall Survival (OS) and Disease-Specific Survival (DSS) of cancers using TCGA datasets. Patients were divided into 0–50% and 50–100% groups according to the expression levels of CCDC58. Samples with no clinical information were removed²⁴. Log-rank *P* value and 95% confidence intervals (CI) were examined. R “survival” package was used for fitting survival regression, and the results were visualized using R “survminer” and “ggplot2” package²⁵. Based on Receiver Operating Characteristic (ROC) curve, the Area Under Curve (AUC) was often used to evaluate the diagnostic effect of variables in predicting outcomes. R “pROC” package and “timeROC” package were used for ROC analysis of the data from TCGA and GTEx, and the results were visualized with R “ggplot2” package^{26,27}.

Critical clinical characteristics analysis

R “stats” package was used to analyze the correlation between CCDC58 expression and clinical characteristics. Statistically significant results were presented in the form of a three-line table.

Genetic alteration and DNA methylation analysis

We obtained the Genetic Alteration of the CCDC58 gene in cBioPortal (<https://www.cbioportal.org/>)^{28,29}, as well as Structural variant data, Mutation data, and Copy Number Alteration (CNA) data. In addition, R “ggplot2” package was used to visualize RNA-seq data and Illumina human methylation 450 data of CCDC58 from TCGA.

Genomic heterogeneity analysis

We obtained the scores of 8 tumor genomic heterogeneity indicators from UCSC and TCGA. And based on R “maftools” package, we analyzed the Pearson correlation between them and the expression of CCDC58, including tumor mutation burden (TMB), mutant-allele tumor heterogeneity (MATH), microsatellite instability (MSI), neoantigen (NEO), PURITY, PLOIDY, homologous recombination deficiency (HRD), loss of heterozygosity (LOH)³⁰. TMB was highly correlated with the efficacy of immune checkpoint inhibitors and could predict the efficacy of immunotherapy in tumor patients to a certain extent³¹. MATH was able to represent a bias in the distribution of mutation annotation format (MAF) values at tumor-specific mutation sites, which were positively associated with tumor heterogeneity³². MSI was the spontaneous loss or acquisition of nucleotides in repetitive DNA, associated with defects in DNA mismatch repair, and was a diagnostic phenotype for a variety of malignancies³³. NEO were tumor-specific antigens derived from somatic cell mutations and were considered important targets for tumor immunotherapy^{34,35}. Tumor tissue contained tumor cells, immune cells, stromal cells, stromal cells, and other cells, which jointly affected the development of the tumor. Therefore, tumor PURITY was significantly correlated with clinical characteristics, genomic expression, and biological characteristics of tumor patients^{35,36}. PLOIDY was considered a hallmark of cancer and was closely related to chromosome instability in cancer development^{35,37}. Homologous recombination was a high-precision DNA repair mechanism, so HRD was a key indicator of multiple tumor treatment options and prognosis^{35,38}. LOH represented chromosomal events that could cause the loss of entire genes and nearby chromosomal regions^{35,39}. Finally, the correlation analysis results of the above 8 indicators were displayed in the form of radar charts.

Immune infiltration analysis

Based on single sample gene set enrichment analysis (ssGSEA) of R “GSVA” package⁴⁰, Spearman correlation analysis was conducted between CCDC58 and 24 immune cells markers of immune infiltration matrix data from TCGA datasets. 24 immune cells included activated DC (aDC), B cells, CD8 T cells, Cytotoxic cells, DC, Eosinophils, immature DC (iDC), Macrophages, Mast cells, Neutrophils, NK CD56bright cells, NK CD56dim cells, NK cells, Plasmacytoid DC (pDC), T cells, T helper cells, T central memory (Tcm), T effector memory (Tem), T follicular helper (TFH), T gamma delta (Tgd), Th1 cells, Th17 cells, Th2 cells, TReg⁴¹. Combined with previous reports³⁵, we also extracted the expression data of two types of immune checkpoint marker genes (inhibitory 24 and stimulatory 36) and analyzed their correlation with CCDC58 in pan-cancer. Finally, R “ggplot2” package was used for visualization.

Single-cell sequencing

In CancerSEA (<http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp>)⁴², the relevance of CCDC58 across 14 functional states in distinct cancers and functional relevance in distinct cell groups was obtained. T-SNE describes the distribution of cells, and the color of the dots represents the level of CCDC58 expression.

Related genes enrichment analysis

BioGRID (<https://thebiogrid.org/>)⁴³ was a biomedical interactive repository whose data was compiled through comprehensive administrative effort. The protein interaction network of CCDC58 can be obtained with it. In addition, Gene Expression Profiling Interactive Analysis (GEPIA 2.0, <http://gepia2.cancer-pku.cn/#index>)⁴⁴ web server was a high-quality resource for gene expression analysis based on tumor and normal samples from TCGA and GTEx databases. Tumor Immune Estimation Resource (TIMER 2.0, <http://timer.cistrome.org/>)⁴⁵ was a comprehensive resource for the systematic analysis of immune infiltration in different cancer types. In all tumor and normal tissue samples, the top 100 genes highly related to CCDC58 were obtained from GEPIA 2.0. We carried out a correlation analysis between the top 10 genes and CCDC58 in GEPIA 2.0 and TIMER 2.0, conducted Gene ontology (GO) and Kyoto encyclopedia of genes and genome (KEGG) enrichment analysis on the top 100

genes, based on R “clusterProfiler” package, to investigate the key phenotypes and signal pathways that CCDC58 might affect in pan-cancer⁴⁶. Finally, R “ggplot2”, “igraph” and “ggraph” packages were used for visualization^{47,48}.

Further analysis of CCDC58 expression in LIHC

TISIDB (<http://cis.hku.hk/TISIDB/index.php>)⁴⁹ was an integrated repository portal for tumor-immune system interactions. Based on CCDC58 expression, we obtained immune subtypes and molecular subtypes of LIHC in TISIDB. In addition, we analyzed correlations between CCDC58 expression and clinical variables in LIHC, GBM/LGG, SARC, and UCEC from TCGA, visualizing the results with boxplots and K–M survival curves. Based on R “survival” package, Cox regression analysis was performed for clinical variables including CCDC58, and variables with $P < 0.1$ were included in the multivariate analysis. According to the variables with $P < 0.05$ in the multivariate Cox regression analysis, R “rms” package was used to construct and visualize the OS prognosis nomogram. Finally, calibration curves were drawn and prognostic decision curve analysis (DCA)⁵⁰ was performed.

Cell lines and transfection

The human liver cancer cell line (MHCC97-H) was purchased from Zhong Qiao Xin Zhou Biotechnology Co., Ltd (Shanghai, China). Human liver cancer cell lines (Huh-7, HCCLM3, and Hep 3B) and human liver cell line (WRL-68) were purchased from Haixing Biological Technology Co., Ltd (Suzhou, China). The human liver cancer cell line (Hep G2) was purchased from Procell Life Science&Technology Co., Ltd (Wuhan, China). All cell lines were identified correctly by STR analysis. MHCC97-H, Huh-7, HCCLM3, and WRL-68 were cultured in Dulbecco’s Modified Eagle Medium (DMEM, Gibco, USA) + 10%Fetal Bovine Serum (FBS, Gibco, USA) + 1%Antibiotics (Penicillin streptomycin, Gibco, USA). Hep G2 and Hep 3B were cultured in Minimum Essential Medium (MEM, Gibco, USA) + 10%Fetal Bovine Serum (FBS, Gibco, USA) + 1%Antibiotics (Penicillin streptomycin, Gibco, USA). The cell incubator was incubated at 37 °C with 5% CO₂. Cells in 6-well plates were transiently transfected with small-interfering RNA (siRNA) of CCDC58 and its negative control at 70–80% density using Lipofectamine 3000 (Invitrogen, USA) according to the manufacturer’s description. The siRNA sequences were presented in Table S1. After 48 h transfection, cells were tested for knockdown efficiency of CCDC58 and used for functional experiments.

RNA isolation and real-time quantitative polymerase chain reaction

Total RNA was isolated with Trizol reagent (Invitrogen) according to the manufacturer’s instructions. The mRNA was reverse transcribed into cDNA using BeyoRT II M-MLV (D7160L, Beyotime). RT-qPCR was performed with 2 × Taq PCR MasterMix (PC1150, Solarbio) and SYBR Green (SY1020, Solarbio) in a Bio-Rad CFX Maestro 2.2 Real-Time PCR System (Bio-Rad, Hercules, CA, USA). β -actin was served as the internal control. The sequences of the primers used were listed in Table S2.

Western blotting and antibodies

Total cellular protein was extracted using Radio Immunoprecipitation Assay Lysis Buffer (RIPA, Beyotime, China) supplemented with Protease inhibitor cocktail (Beyotime, China) and quantified by Bicinchoninic Acid (BCA kit, Boster, China) assay. Target and reference proteins were separated by 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to the PVDF membrane (Millipore, USA). Final Western blot visualization was performed using the ECL kit (Beyotime, China). The antibodies involved in this study included anti-CCDC58 (diluted 1:1000, Novus, NBP2-14452, USA) and β -actin (diluted 1:20000, Proteintech, 66009-1-Ig, China).

Cell counting kit-8 assay and colony formation

Cell counting kit-8 (CCK-8) assay was used to detect cell viability and proliferation. Cells were seeded in 96-well plates at a density of 4000 cells per well and cultured overnight to allow cell adherence. The medium was replaced with fresh medium containing CCK-8 kit (No.C0005, TargetMol). After 4 h of incubation, absorbance was measured at a wavelength of 450 nm using a microplate reader. Cells were seeded into 6-well plates at a density of 1000 cells per well for the colony formation assay. The medium was changed every 3 days, and after 10 days the cells were fixed with 4% paraformaldehyde and stained with 0.5% crystal violet.

Wound healing and transwell assays

Wound healing assay and transwell assay were used to detect cell migration ability. Cells were seeded in 6-well plates at a density of 50% per well, 2 mL of medium was added and cultured overnight. After siRNA transfection, a straight wound was created in each well using a 200 μ L pipette tip. Subsequently, cell debris was washed off with PBS and cultured in DMEM containing 2%FBS. The effect of CCDC58 knockdown on cell migration was analyzed by comparing the wound healing images at 0 and 48 h. Tumor cells were seeded at a density of 5×10^4 cells per well onto the upper chamber (8 μ m pore size) containing 200 μ L of serum-free medium, while the lower chamber was filled with 500 μ L of medium containing 10%FBS. Cells migrated for 24 h. After methanol fixation and crystal violet staining, 5 random fields were selected to collect images, and the number of cells passing through the chamber reflected the migration ability of the cells.

Statistical analysis

Pearson and Spearman methods were used for correlation analysis. Different methods were selected to calculate the differences between groups according to the data characteristics, including the T-test, Welch t ‘test, Wilcoxon rank sum test, etc. All statistics were performed using R (4.2.1) via R Studio. $P < 0.05$ was defined as statistically

significant and all statistical analyses were two-sided tests ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$, “ns” was regarded as not statistically significant).

Results

Expression of CCDC58 in various tumors and normal tissues

First, based on TCGA RNA-seq data in STAR workflow, we conducted unpaired samples analysis of mRNA expression of CCDC58 pan-cancer (Fig. 1A). Except for tumors with little or no normal tissue samples, it was not difficult to find that in most cancers, the expression of CCDC58 in tumor tissues was statistically higher than that in normal tissues, including BLCA, BRCA, CESC, CHOL, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, STAD, UCEC. In some cancers, tumor tissue expression was lower, which was statistically significant only in PCPG and THCA. Secondly, based on the TCGA datasets, we extracted and analyzed the tumor and normal samples with corresponding number pairs (Fig. 1B). We observed that out of 23 cancers, 13 tumor tissues expressed higher than their paired normal tissues. This was consistent with the results of unpaired analysis, but there was no longer statistically significant CCDC58 expression in CESC and KIRC. In addition, the results in THCA were also consistent with the unpaired results, which were also lower in tumor tissue expression. Thirdly, we continued to analyze the difference in expression of CCDC58 in pan-cancer after being added to the control of normal tissue samples from GTEx (Fig. 1C). We presented all statistically significant results in the form of violin plots. Most results showed that CCDC58 expression was higher in tumor tissues, except for LAML. Moreover, the expression of CCDC58 in THCA was contrary to the previous result, that the expression in tumor tissue is higher.

Next, we analyzed the protein expression of CCDC58 in cancers from CPTAC (Fig. 2A). In our results, the protein expression of CCDC58 was different between the primary tumors and normal tissues in 8 cancers, including BRCA, KIRC, COAD, GBM, LUAD, OV, PAAD, and UCEC. The expression of the CCDC58 protein was higher in tumor tissues in BRCA, COAD, LUAD, OV, and UCEC, but lower in KIRC, GBM, and PAAD. The correlation between CCDC58 protein expression and tumor stage was also detected in CPTAC (Fig. 2B). CCDC58 protein expression was significantly correlated with individual cancer stages in BRCA, KIRC, COAD, LUAD, OV, PAAD, and UCEC. In addition, immunohistochemical images of CCDC58 in multiple organs were obtained from the HPA database (Fig. 2C). It could be seen from the images that CCDC58 protein expression was higher in the breast, colon, liver, and ovary, excluding the cerebral cortex and kidney. This was consistent with the antibody staining score in HPA, which was based on the combination of staining intensity and staining cell proportion. Western blotting results showed that the expression of CCDC58 in liver cancer cell lines was significantly higher than that in normal liver cell line, which was consistent with the data of various databases (Fig. 2D). The full-length blot image was included in Fig. S1.

Prognostic and diagnosis value of CCDC58 across cancers

To evaluate the correlation between CCDC58 expression and the prognosis of patients with various tumors, K–M survival curves were plotted, and log-rank tests were performed. We focused on OS and DSS and presented statistically significant results. As shown in Fig. 3A, CCDC58, a high risk factor, linked to poor OS of patients in ACC ($P = 0.008$), GBM/LGG ($P < 0.001$), HNSC ($P = 0.008$), KICH ($P = 0.020$), LIHC ($P < 0.001$), PAAD ($P = 0.014$), SARC ($P < 0.001$) and UCEC ($P < 0.001$). Similarly, in the DSS analysis, CCDC58 was also a high risk factor in ACC ($P = 0.015$), GBM/LGG ($P < 0.001$), HNSC ($P = 0.008$), KICH ($P = 0.059$), LIHC ($P < 0.001$), PAAD ($P = 0.073$), SARC ($P = 0.008$) and UCEC ($P < 0.001$) (Fig. 3B). Detailed statistical results were presented in Tables S3 and S4.

On this basis, we further evaluated the diagnostic value of CCDC58 using ROC curves in the above 8 cancers. The value range of AUC was generally between 0.5 and 1. The closer the AUC was to 1, the better the diagnostic effect of CCDC58 was in predicting the outcome. As shown in Fig. 3C, CCDC58 showed good diagnostic accuracy in GBM/LGG (AUC: 0.971), HNSC (AUC: 0.864), LIHC (AUC: 0.853), PAAD (AUC: 0.944) and UCEC (AUC: 0.956), and average diagnostic accuracy in ACC (AUC: 0.553), KICH (AUC: 0.635) and SARC (AUC: 0.559). In addition, we evaluated the accuracy of CCDC58 in predicting 1-year, 3-year, and 5-year survival using time-dependent ROC curves. In addition, after adding GTEx normal samples, we evaluated the accuracy of CCDC58 in predicting 1-year, 3-year, and 5-year survival using time-dependent ROC curves (Fig. 3D). It showed that CCDC58 predicted survival with better diagnostic accuracy in LIHC (1-year AUC: 0.728, 3-year AUC: 0.676, 5-year AUC: 0.656) and KICH (1-year AUC: 0.949, 3-year AUC: 0.856, 5-year AUC: 0.825).

Correlation between CCDC58 expression and critical clinical characteristics

We analyzed the correlation between CCDC58 expression and critical clinical characteristics (Table S5). The expression of CCDC58 in GBM/LGG was correlated with age, WHO grade, and isocitrate dehydrogenase (IDH) status. The higher the expression of CCDC58, the older the patient, the higher the WHO grade, and the lower the mutation rate of IDH. In HNSC, CCDC58 was more highly expressed in male patients, and its high expression was indicative of a higher pathologic stage. And in PAAD, CCDC58 was indicative of a higher pathologic T stage.

Genetic alteration and DNA methylation of CCDC58

Detailed gene alteration types and frequency of CCDC58, including mutation, structural variant, amplification, deep deletion, and multiple alterations, were obtained from cBioPortal (Fig. 4A). We observed that CCDC58 amplification accounted for the largest proportion. The top three tumors in alteration frequency were LUSC (mutation 0.41% + amplification 6.37% + deep deletion 0.21%), CESC (mutation 0.34% + amplification 5.39%), and ESCA (mutation 0.55% + amplification 2.75% + deep deletion 0.55%). Figure 4B showed altered/profiled = 156/10950 (1.4%) in CCDC58. In addition, based on cBioPortal, we found that X60_splice was the most

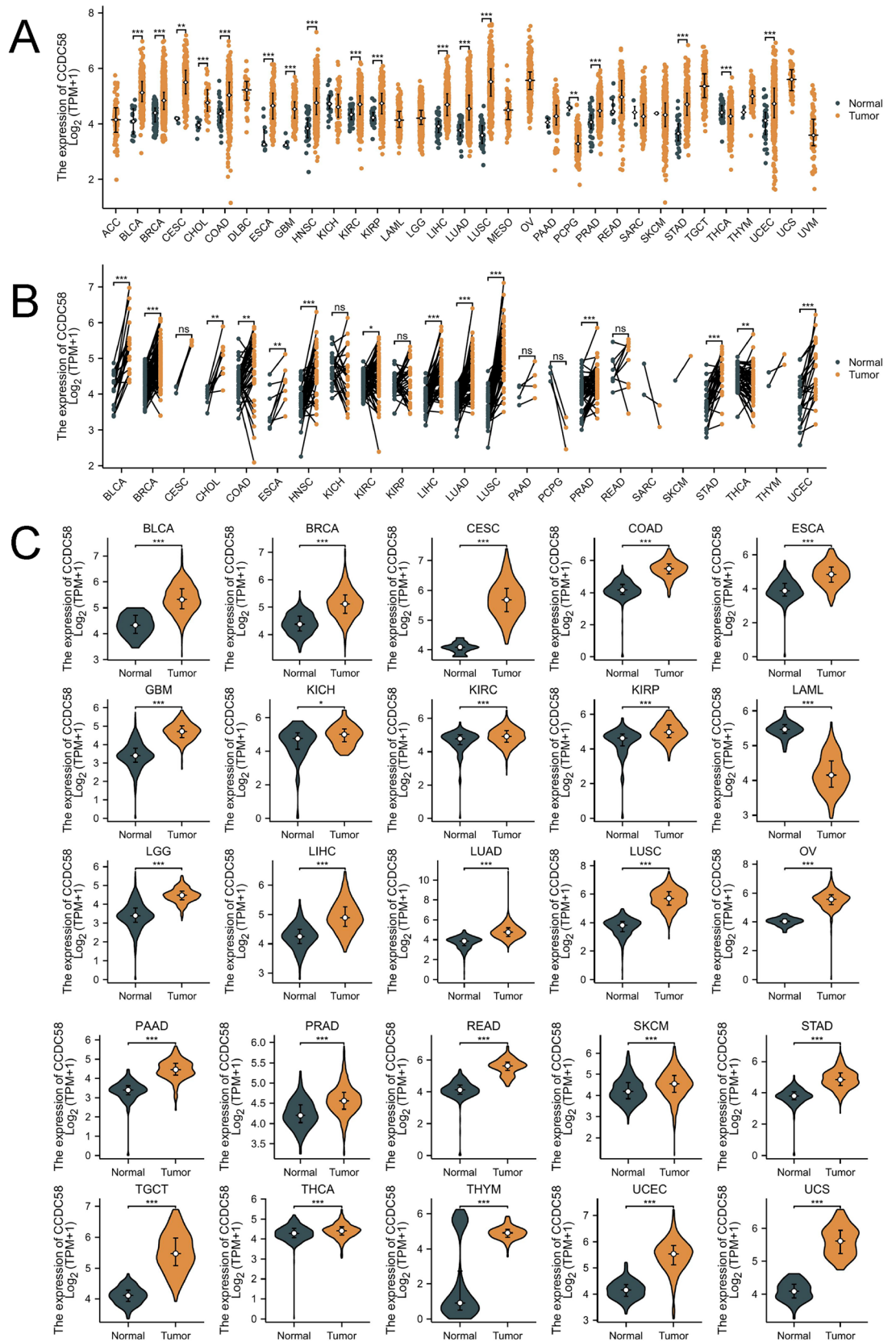


Figure 1. (A) The expression of CCDC58 in pan-cancer from TCGA datasets. (B) The expression of CCDC58 using paired tumor and normal samples from TCGA datasets. (C) The expression of CCDC58 in pan-cancer from TCGA and GTEx.

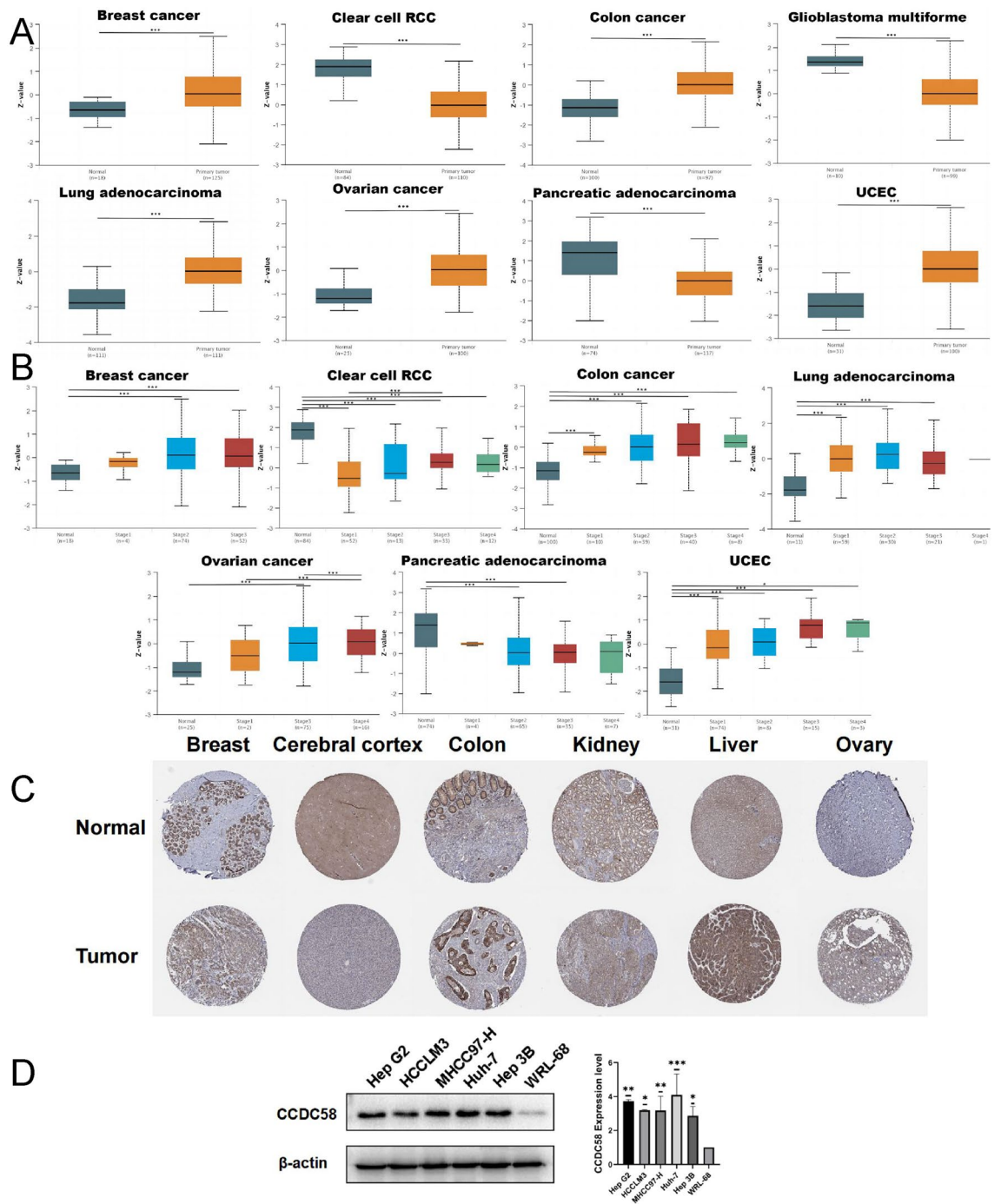


Figure 2. (A) The expression of CCDC58 protein in BRCA, Clear cell RCC, COAD, GBM, LUAD, OV, PAAD and UCEC from CPTAC. (B) Based on individual cancer stage, the expression of CCDC58 protein in BRCA, KIRC, COAD, LUAD, OV, PAAD and UCEC. (C) The expression of CCDC58 protein in tumor and normal tissue of breast, cerebral cortex, colon, kidney, liver and ovary from HPA. (D) The expression of CCDC58 protein in cell lines by western blotting.

important site of the CCDC58 mutation known to date. It was supported by three samples, 2 cases of OV and 1 case of LAML (Fig. 4C).

Through analyzing RNA-seq data from TCGA and Illumina human methylation 450 methylation data using two probes (cg16509192 and cg13010990), we found that CCDC58 expression was negatively correlated with its promoter methylation in HNSC, KICH, and LIHC, positively correlated in PAAD and UCEC (Fig. 4D).

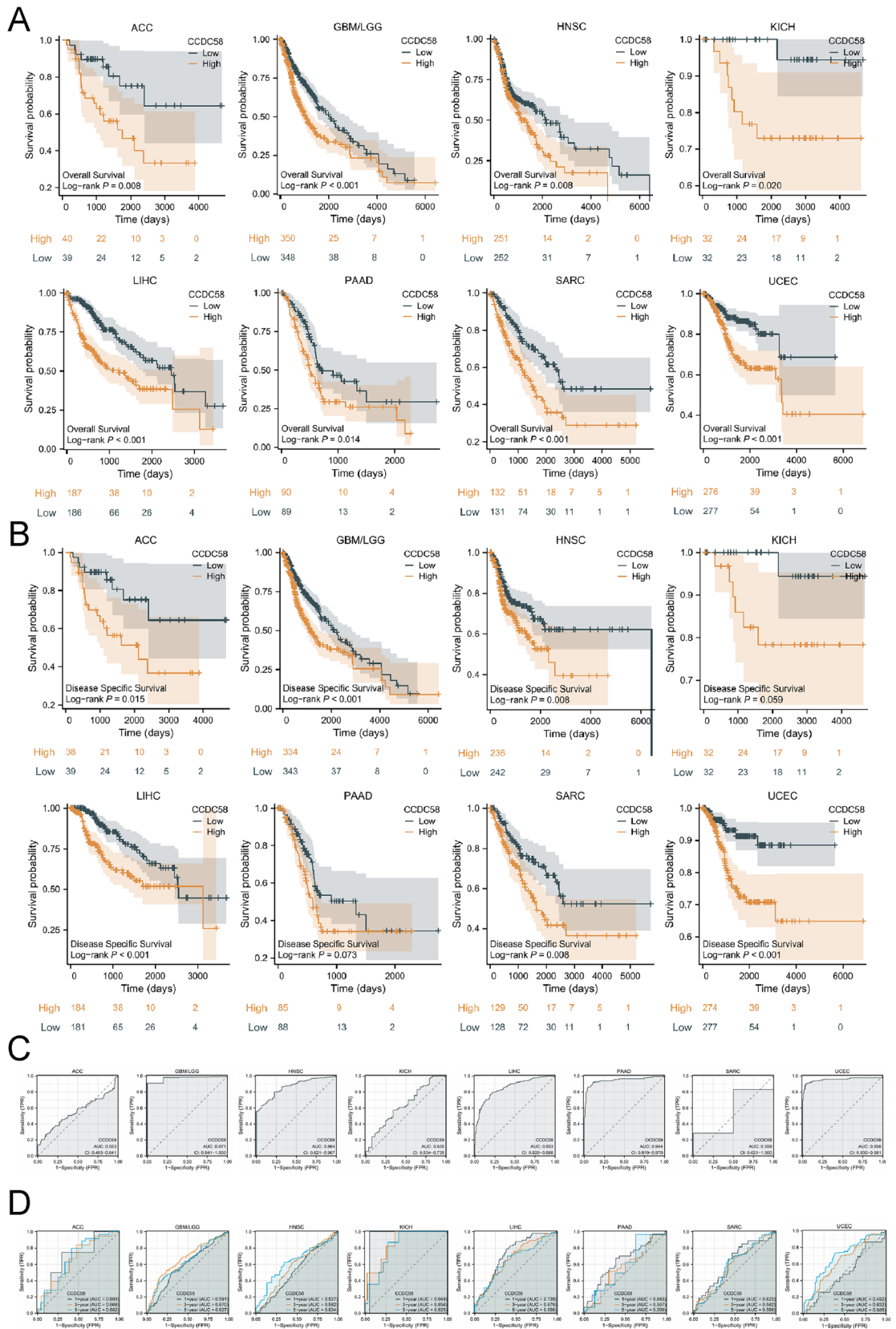


Figure 3. (A) OS analysis of CCDC58 genes from TCGA datasets. (B) DSS analysis of CCDC58 genes from TCGA datasets. (C) ROC curve of CCDC58 genes from TCGA and GTEx (GBM/LGG, HNSC and SARC did not add GTEx). (D) Time-dependent ROC curve of CCDC58 genes from TCGA database.

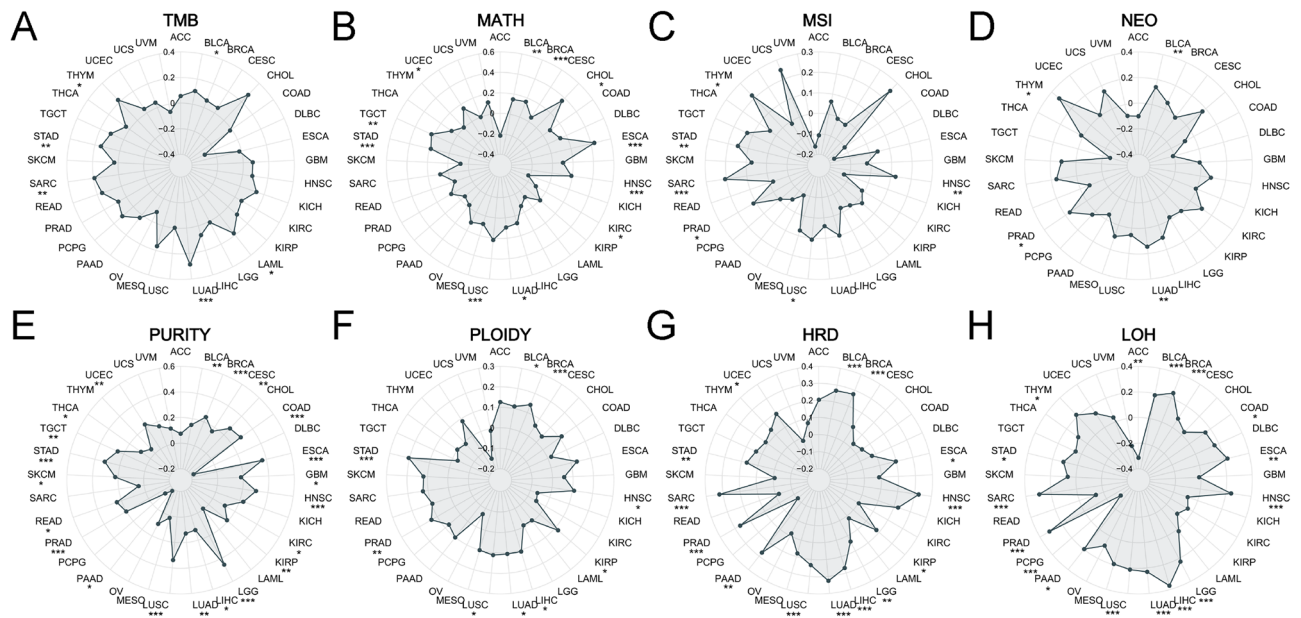


Figure 5. (A–H) Radar chart of correlation between CCDC58 expression and TMB, MATH, MSI, NEO, PURITY, PLOIDY, HRD and LOH in pan-cancer respectively.

in BLCA, BRCA, CESC, COAD, ESCA, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, PRAD, READ, SKCM, STAD, TGCT, THCA, and UCEC, negatively correlated in PAAD (Fig. 5E). CCDC58 expression levels were positively correlated with PLOIDY in BLCA, BRCA, HNSC, KIRP, LIHC, LUAD, LUSC, PRAD, and STAD (Fig. 5F). CCDC58 expression levels were positively correlated with HRD in BLCA, BRCA, ESCA, HNSC, KIRP, LGG, LIHC, LUAD, LUSC, PAAD, PRAD, SARC, STAD, and UCEC (Fig. 5G). CCDC58 expression levels were positively correlated with LOH in BLCA, BRCA, COAD, ESCA, HNSC, LGG, LIHC, LUAD, LUSC, PAAD, PRAD, SARC, STAD, and THYM, negatively correlated in ACC and PCPG (Fig. 5H).

Immune cell infiltration and immune checkpoint genes analysis

In pan-cancer, to determine the value of CCDC58 in influencing the tumor immune microenvironment, we analyzed the correlation between CCDC58 expression and immune cell infiltration in cancers (Fig. 6). The results showed that CCDC58 expression was negatively correlated with most immune cells in ACC, HNSC, LIHC, and PAAD. In GBM/LGG, KICH, SARC, and UCEC, CCDC58 expression was positively correlated with some immune cells. In addition, we established a heatmap of the correlation between CCDC58 and two types of immune checkpoint marker genes. In Fig. 7, we could intuitively see that CCDC58 expression levels were positively correlated with VEGFA, IL12A, CD276, and HMGB1, negatively correlated with C10orf54, EDNRB, CX3CL1, CD27, PRF1, SELP, ENTPD1, and TLR4. In TGCT, LUSC, THCA, OV, CESC, HNSC, ESCA, BLCA, BRCA, LUAD, STAD, and CCDC58 expression was highly correlated with multiple immune checkpoint marker genes.

Relevance of CCDC58 across functional states in cancers

Based on single-cell sequencing data from CancerSEA, we established a heat map of the correlation between CCDC58 expression and 14 functional states (Fig. 8A). We observed that CCDC58 expression was highly correlated with these functional states in acute myeloid leukemia (AML), head and neck cancer (HNSCC), retinoblastoma (RB), and uveal melanoma (UM). In AML, almost all of them were positively correlated with CCDC58 expression, except cell cycle, DNA damage, DNA repair, and invasion. In HNSCC, except for cell cycle, DNA repair, and stemness, almost all were negatively correlated with CCDC58 expression. In RB, CCDC58 expression was positively correlated with angiogenesis, differentiation, inflammation, metastasis, quiescence, and stemness, and negatively correlated with apoptosis, cell cycle, DNA damage, DNA repair, epithelial-mesenchymal transition (EMT), and proliferation. In UM, almost all of them were negatively correlated with CCDC58 expression, except stemness. In addition, the expression profiles of CCDC58 at the single-cell levels were presented as a t-SNE diagram (Fig. 8B).

Results of enrichment analysis of CCDC58 related genes

We performed an enrichment analysis of CCDC58-related genes to explore their potential molecular mechanism in the tumor microenvironment. Firstly, the protein interaction network diagram of CCDC58, including 34 related proteins, was obtained in BioGRID (Fig. 9A). Secondly, the top 100 genes co-expressed with CCDC58 in pan-cancer were obtained in GEPIA2.0 (Table S6). Correlation of UMPS, MRPL3, MRPS22, RUVBL1, TRMT10C, MRPL47, POLR2H, RFC4, ACTL6A, SNRPG with CCDC58 scatter diagram as shown in Fig. 9B. Almost all of these genes showed a high correlation with CCDC58 in pan-cancer. Similar correlations were shown in Fig. 9C

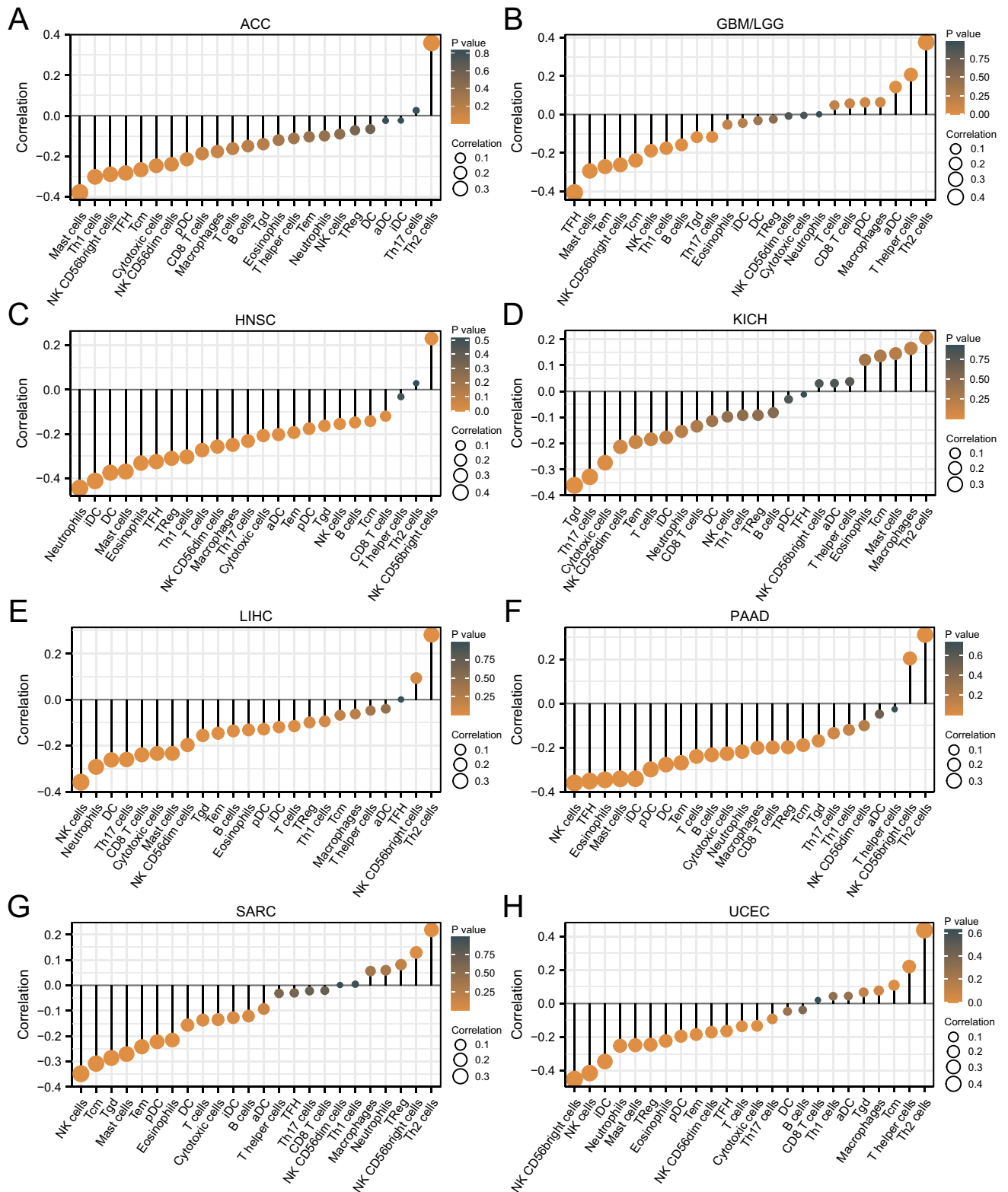


Figure 6. (A–H) Correlation analysis between CCDC58 and 24 immune cells in ACC, GBM/LGG, HNSC, KICH, LIHC, PAAD, SARC, UCEC.

from TIMER2.0. Finally, GO/KEGG enrichment analysis was performed for these 10 CCDC58-related genes (Fig. 9D). GO analysis included biological process (BP), cellular component (CC), and molecular function (MF). The BP results showed that CCDC58 was mainly related to organelle fission, chromosome segregation,

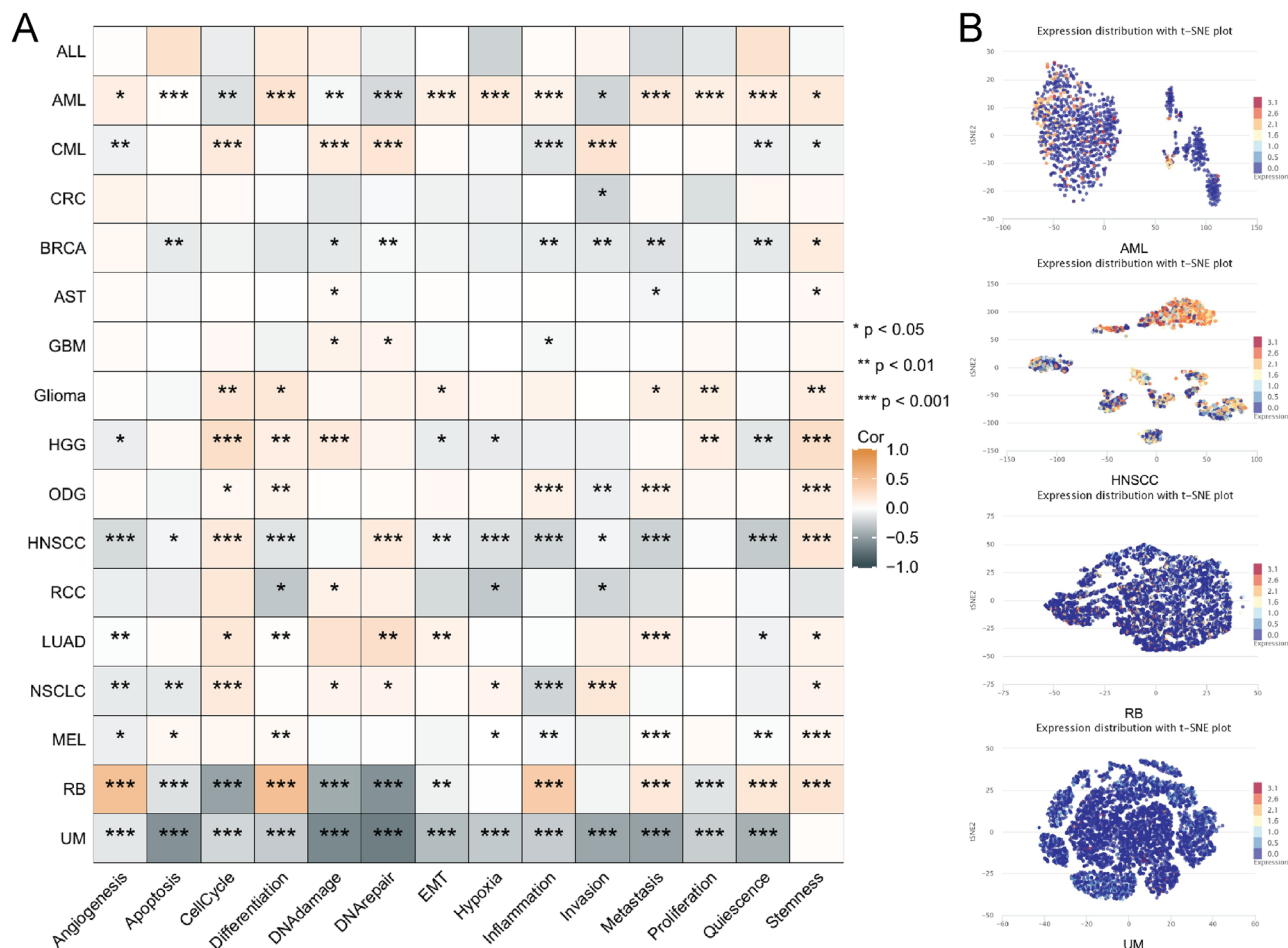


Figure 8. (A) Average correlations between CCDC58 and 14 functional states in different cancers. (B) CCDC58 expression distribution shown as t-SNE in AML, HNSCC, RB and UM.

and mitotic nuclear division. CC of CCDC58 was linked to the chromosomal region, chromosome, centromeric region, and U2 snRNP. In addition, the MF results showed that CCDC58 was also associated with protein kinase regulator activity, cyclin-dependent protein serine/threonine kinase regulator activity, and cyclin-dependent protein serine/threonine kinase activator activity. Finally, the pathways of KEGG enrichment were mainly related to spliceosome, cell cycle, and progesterone-mediated oocyte maturation. Figure 9E showed the enrichment analysis network diagram of CCDC58.

The expression difference and prognostic value of CCDC58

We further analyzed the differences in CCDC58 expression in different clinical variable subgroups of LIHC and showed the results with statistical differences. Statistical results showed that CCDC58 expression was higher in patients with alpha-fetoprotein (AFP) > 400 ng/ml (Fig. S2A), higher histologic grade (Fig. S2B), worse OS (Fig. S2C) and vascular invasion (Fig. S2E), lower in patients with prothrombin time > 4 s (Fig. S2D) and weight > 70 kg (Fig. S2F). Furthermore, in LIHC, the correlation between CCDC58 expression and molecular and immune subtypes was obtained from TISIDB (Fig. S2G). The results showed that CCDC58 expression was independent of the molecular subtype of LIHC, and was lower in C3 (inflammatory) and higher in C2 (IFN-gamma dominant) and C4 (lymphocyte depleted) in the immune subtype of LIHC.

Subgroup K-M survival curves based on CCDC58 expression levels were plotted to evaluate the role of CCDC58 in predicting survival value in LIHC patients with different variable groups, including age (Fig. S3A), gender (Fig. S3B), body mass index (BMI, Fig. S3C), albumin (Fig. S3D), AFP (Fig. S3E), prothrombin time (Fig. S3F), tumor status (Fig. S3G), pathologic T stage (Fig. S3H), pathologic stage (Fig. S3I). Most of the grouping results were consistent with OS analysis of LIHC, namely, the higher the expression of CCDC58, the worse the prognosis. However, we also found that CCDC58 was of modest value in predicting survival in patients with the female, AFP ≤ 400 ng/ml, prothrombin time ≤ 4 s, pathologic T2 stage, and pathologic stage II.

Cox regression analysis regarding OS was performed using data from LIHC patients in TCGA. Univariate analysis showed that pathologic T stage, Tumor status, and CCDC58 expression were associated with OS. When variables with *P* < 0.1 in univariate results were included, multivariate analysis showed that pathologic T stage, Tumor status, and CCDC58 expression levels were independent risk factors for OS (Table S7). The results of multivariate Cox regression are presented as forest plots (Fig. 10A). The independent risk factors for OS of LIHC

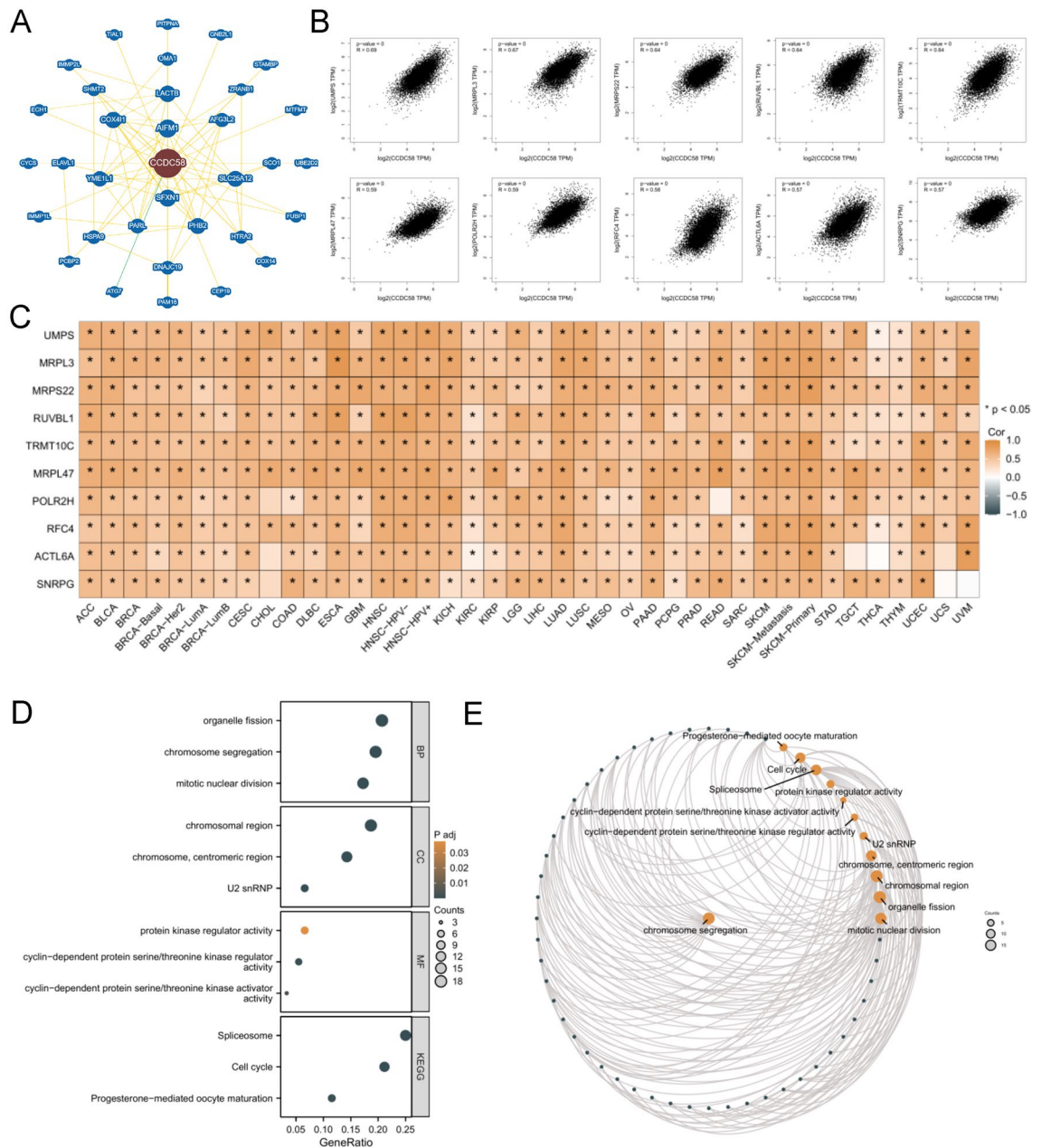


Figure 9. (A) Protein interaction network based on CCDC58 from BioGRID (green: association with genetic evidence, yellow: association with physical evidence). (B) Scatter plot of correlation analysis of CCDC58 related genes. (C) Correlation analysis heatmap of CCDC58 related genes in pan-cancer. (D) GO/KEGG enrichment analysis of CCDC58 related genes. (E) Enrichment analysis network diagram of CCDC58 related genes.

patients were drawn into the nomogram model, which could better predict the survival rate of patients at 1-year, 3-year, and 5-year, and the C-index was 0.692 (95%CI: 0.668–0.717) (Fig. 10B). The prognostic calibration result showed that the three fitting curves were well-fitted to the ideal line (Fig. 10C). Finally, DCA was performed to evaluate the clinical utility or patient benefit of the prognostic model, incorporating CCDC58 expression and risk scores generated from Cox regression analysis (Fig. 10D).

Subgroup K-M survival curves of different clinical variables in GBM/LGG, SARC and UCEC also showed an overall trend of poor prognosis with higher CCDC58 expression. However, the predictive value of CCDC58 was different in different intervals of clinical variables. In GBM patients of age ≤ 60 years, high CCDC58 expression pointed to a worse prognosis, while this trend was not evident in patients of age > 60 years (Fig. S4A). Similar patterns were observed for race (Fig. S4C), 1p/19q codeletion (Fig. S4D), IDH status (Fig. S4E), except gender (Fig. S4B). In SARC, the groups in which CCDC58 expression was associated with poor prognosis included: female (Fig. S5B), white (Fig. S5C), leiomyosarcoma (Fig. S5D), non-multifocal tumor (Fig. S5E), R0 residual tumor (Fig. S5F), and not receiving radiotherapy (Fig. S5G). In patients over 60 years old (Fig. S5A), although the statistical significance of log-rank test was not significant, the K-M curve still reflected the trend of high

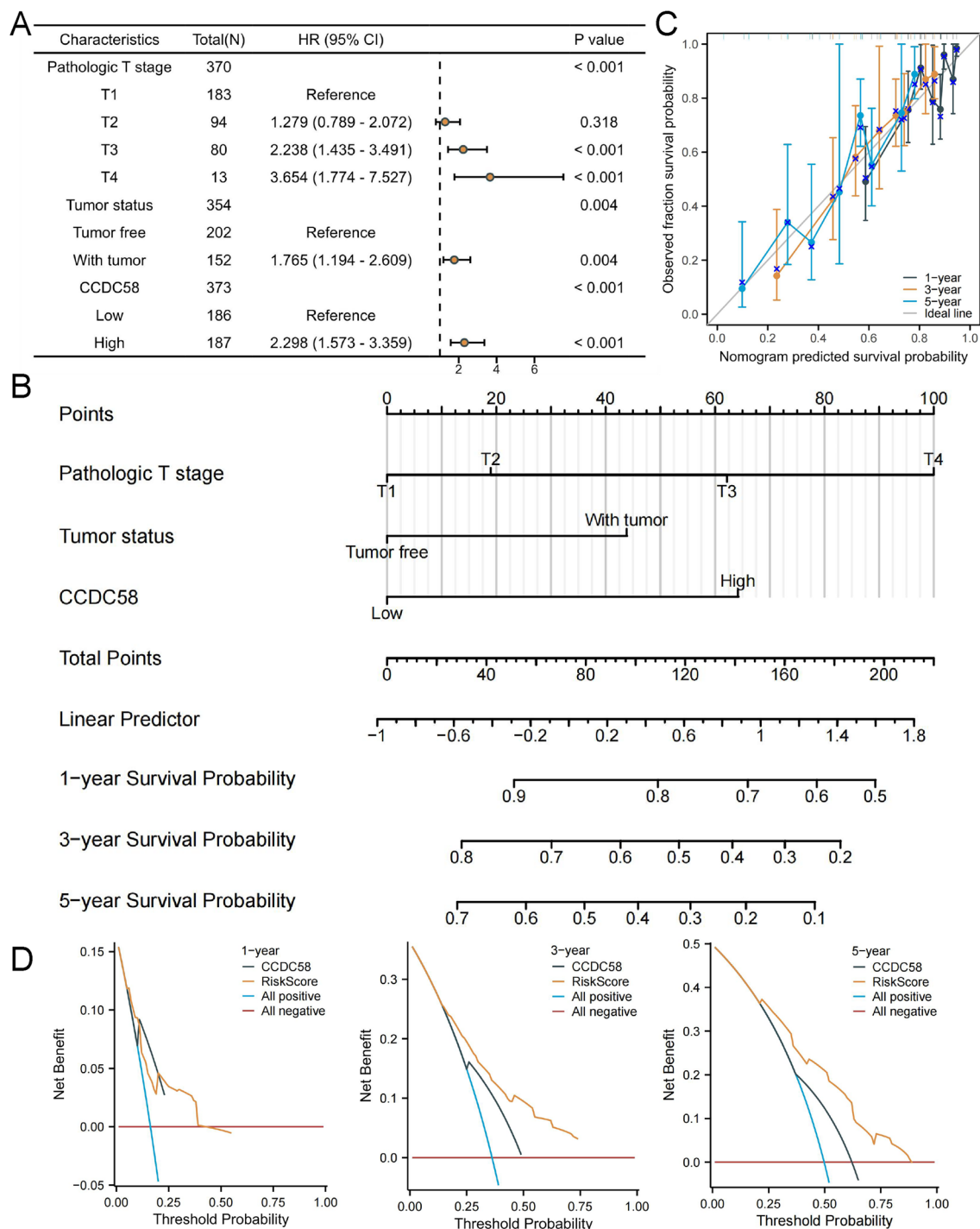


Figure 10. (A) Forest plots of multivariate Cox regression results. (B) Nomogram model for 1-year, 3-year, 5-year OS of LIHC patients. (C) The prognostic calibration curves of nomogram model. (D) DCA for 1-year, 3-year, 5-year results.

expression of CCDC58 with poor prognosis. In UCEC, patient groups in which CCDC58 expression level was associated with poor prognosis included: BMI ≤ 30 (Fig. S6A), non-diabetes (Fig. S6B), endometrioid (Fig. S6C), post-menopause (Fig. S6D), not treated with radiotherapy (Fig. S6E), white race (Fig. S6F), and weight ≤ 80 (Fig. S6G).

CCDC58 affected the biological behaviors of LIHC cells in vitro

Few studies investigated the role of CCDC58 in the biological behavior of tumor cells, so we designed siRNA to down-regulate CCDC58 gene. Based on the results of CCDC58 expression in western blot experiments, siRNA was transfected into Huh-7 cells. Since the transfection efficiency of siRNA-3 was less than 70%, siRNA-1 and siRNA-2 were selected for further study (Fig. S7A). CCK-8 assay was used to detect the proliferation of Huh-7

cells. Decreased expression of CCDC58 significantly inhibited the proliferation of Huh-7 cells (Fig. S7B). The number of colonies was reduced in Huh-7 cells after down-regulation of CCDC58 expression (Fig. S7C). Knock-down of CCDC58 also inhibited Huh-7 cells migration in a wound healing assay (Fig. S7D). Transwell assay was used to investigate the migration ability of tumor cells, and CCDC58 knockdown significantly reduced the number of cells passing through the chamber (Fig. S7E). In conclusion, the CCDC58 gene plays an important role in LIHC cell proliferation and migration.

Discussion

Cancer remained the leading cause of death in all countries in the world⁵¹. According to statistics, in 2020, there were about 19.3 million new cancer cases and about 10 million cancer deaths worldwide⁵². There were many reasons for the poor prognosis of cancer patients, including tolerance to immunotherapy and differences in the tumor microenvironment⁵³. We found that CCDC58 was a powerful pan-cancer biomarker that had rarely been reported. It could be used to assist in the diagnosis and predict the prognosis of patients with a variety of cancers, as well as to evaluate the immune infiltration of tumors. CCDC58 was highly conserved in evolution and a homolog of MIX23. It was widely expressed in various tissues and organs of the human body⁵⁴. At the earliest, CCDC58 was identified in the mitochondrial membrane space as Ybl107c (Mic23) and renamed to avoid confusion with the mitochondrial contact site and cristae organizing system (MICOS)^{55,56}. Moreover, other authors had demonstrated that mutations in CCDC58 increased resistance to Ehrlichiae in *Drosophila melanogaster*⁵⁷.

In the beginning, we investigated the expression of CCDC58 at the transcriptional and translational levels using data from various databases. The results showed that mRNA and protein levels of CCDC58 were significantly upregulated in a variety of tumors compared with adjacent normal tissues. The previous report backed up our results. Some scholars had mentioned that CCDC58 was highly expressed in tumor tissues in triple-negative breast cancer. In addition, MTT and transwell assays showed that overexpression of CCDC58 increased the proliferation, migration, and invasion of tumor cells¹³. The results of western blotting also supported the conclusion of our analysis. It was worth mentioning that the differences in mRNA and protein level results might be due to the metabolism and modification of proteins and different data sources or processing methods in different databases⁵⁸. Second, we evaluated the prognostic and diagnostic value of CCDC58. The expression level of CCDC58 was associated with OS and DSS in 8 cancers, and CCDC58 had good diagnostic efficacy in these cancers, including ACC, GBM/LGG, HNSC, KICH, LIHC, PAAD, SARC, UCEC. Furthermore, CCDC58 had excellent diagnostic performance in GBM/LGG, HNSC, LIHC, PAAD, and UCEC (AUC > 0.85). All in all, the results indicated that CCDC58 had a good predictive ability for the prognosis of patients.

Over the course of a person's life, spontaneous mutations in somatic cells gradually accumulate. Such mutations can lead to aging and even cancer⁵⁹. Therefore, the Structural variant data, Mutation data, and CNA data of CCDC58 were obtained, and the correlation between CCDC58 and promoter methylation was analyzed. The obtained information showed that in most tumors, the gene alteration types of CCDC58 mainly included mutation, amplification, and deep deletion. DNA methylation is a form of epigenetic regulation. Abnormal DNA methylation includes hypomethylation and hypermethylation, both of which are involved in tumor regulation^{60,61}. Our correlation analysis results showed that CCDC58 expression was negatively correlated with DNA methylation in HNSC, KICH, and LIHC, and positively correlated with DNA methylation in PAAD and UCEC. In addition, 8 genomic heterogeneity indicators, including TMB, MATH, MSI, NEO, PURITY, PLOIDY, HRD, and LOH, were also included, which were related to the prognosis or treatment effect of cancer patients. TMB could be used to predict the efficacy of immune checkpoint inhibitors and was a widely recognized predictor^{62,63}. Moreover, it has been suggested that MSI status in malignant tumors represents a worse prognosis, such as COAD⁶⁴. Several studies showed that the MATH score could be used to evaluate the effect of patient treatment, such as neoadjuvant chemotherapy for breast cancer patients and EGFR TKI for lung adenocarcinoma patients^{65,66}. The results of this study indicated that CCDC58 was associated with multiple tumor heterogeneity indicators and affected the survival and treatment response of patients in a variety of cancers.

The metabolism and secretion of tumor cells can alter the function of the surrounding microenvironment, which is extremely important for tumor progression. Immune cells are an important part of the tumor microenvironment, and the infiltration of immune cells greatly affects the growth and metastasis of tumor cells^{67,68}. The results of this study showed that CCDC58 expression was positively correlated with the infiltration of Th2 cells, T helper cells, and NK CD56bright cells in most tumors. NK cells played an important role in monitoring the tumor immune microenvironment and could directly lyse tumor cells without sensitization⁶⁹. NK cells secreted chemokines and cytokines to regulate the immune response, which had great potential in tumor immunotherapy, and the safety of its clinical application had been confirmed^{70,71}. It had been well known that Th2 cells mediated immune responses in various allergic diseases, such as asthma⁷². In fact, Th2 cells could initiate an anti-tumor response through the MHC II complex pathway, and their immune response to tumors was largely mediated by the secretion of Th2 cytokines^{72,73}. In addition, multiple immune checkpoint genes were also associated with CCDC58 expression. Therefore, we suggested that CCDC58 might affect the sensitivity of cancer patients to immunotherapy by interfering with the tumor immune microenvironment, and it might be a potential therapeutic target and a new biomarker for predicting the prognosis of patients.

The cell cycle was regulated through a complex interaction network, which was closely related to the occurrence and development of tumors⁷⁴. Interference with the regulation of the cell cycle is one of the key mechanisms of continuous tumor cell division⁷⁵. Enrichment results highly suggested that CCDC58 was associated with mitosis, chromosome, cell cycle, and its related proteins. This suggested that CCDC58 might play an important role in the proliferation, apoptosis, and other phenotypes of tumor cells. In addition, we found that CCDC58 was associated with a variety of important clinical variables in LIHC and led to differential prognoses in different subgroups.

Using multiple data sources, this analysis provides a comprehensive picture of the role and underlying mechanism of CCDC58 in pan-cancer, but there are some limitations. First of all, more sequencing results from multi-center clinical samples were needed to improve the reliability of CCDC58 in prognosis and diagnosis. In addition, the regulatory mechanisms of CCDC58 need to be further explored to investigate its effect on tumor progression or immune escape.

Data availability

All data involved in this study were obtained from open databases (where specific sources were indicated in the article) or were presented in the Supplementary materials.

Received: 9 July 2023; Accepted: 8 April 2024

Published online: 13 April 2024

References

- Rose, A., Schraegle, S. J., Stahlberg, E. A. & Meier, I. Coiled-coil protein composition of 22 proteomes—differences and common themes in subcellular infrastructure and traffic control. *Bmc Evol. Biol.* **5**, 66. <https://doi.org/10.1186/1471-2148-5-66> (2005).
- Burkhard, P., Stetefeld, J. & Strelkov, S. V. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82–88. [https://doi.org/10.1016/s0962-8924\(00\)01898-5](https://doi.org/10.1016/s0962-8924(00)01898-5) (2001).
- Truebestein, L. & Leonard, T. A. Coiled-coils: The long and short of it. *Bioessays* **38**, 903–916. <https://doi.org/10.1002/bies.20160062> (2016).
- Lupas, A. N. & Gruber, M. The structure of alpha-helical coiled coils. *Adv. Protein Chem.* **70**, 37–78. [https://doi.org/10.1016/S0065-3233\(05\)70003-6](https://doi.org/10.1016/S0065-3233(05)70003-6) (2005).
- Liu, Z. *et al.* Decreased expression of updated NESG1 in nasopharyngeal carcinoma: its potential role and preliminarily functional mechanism. *Int. J. Cancer* **128**, 2562–2571. <https://doi.org/10.1002/ijc.25595> (2011).
- Park, S. J. *et al.* Epigenetic alteration of CCDC67 and its tumor suppressor function in gastric cancer. *Carcinogenesis* **33**, 1494–1501. <https://doi.org/10.1093/carcin/bgs178> (2012).
- Yang, J., Yang, L., Li, S. & Hu, N. HGF/c-met promote renal carcinoma cancer stem cells enrichment through upregulation of cir-CCDC66. *Technol. Cancer Res Treat* **19**, 1078168762. <https://doi.org/10.1177/1533033819901114> (2020).
- Feng, Y. *et al.* CCDC85B promotes non-small cell lung cancer cell proliferation and invasion. *Mol. Carcinog.* **58**, 126–134. <https://doi.org/10.1002/mc.22914> (2019).
- Hu, D. D., Li, P. C., He, Y. F., Jia, W. & Hu, B. Overexpression of coiled-coil domain-containing protein 34 (CCDC34) and its correlation with angiogenesis in Esophageal squamous cell carcinoma. *Med Sci Monit* **24**, 698–705. <https://doi.org/10.12659/msm.908335> (2018).
- Stelzer, G. *et al.* The genecards suite: From gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.* **54**, 1–30. <https://doi.org/10.1002/cpbi.5> (2016).
- Ramachandran, K. *et al.* SARS-CoV-2 infection enhances mitochondrial PTP complex activity to perturb cardiac energetics. *Iscience* **25**, 103722. <https://doi.org/10.1016/j.isci.2021.103722> (2022).
- Zoller, E. *et al.* The intermembrane space protein Mix23 is a novel stress-induced mitochondrial import factor. *J. Biol. Chem.* **295**, 14686–14697. <https://doi.org/10.1074/jbc.RA120.014247> (2020).
- Wang, Z. *et al.* Circ-TRIO promotes TNBC progression by regulating the miR-432-5p/CCDC58 axis. *Cell Death Dis.* **13**, 776. <https://doi.org/10.1038/s41419-022-05216-7> (2022).
- Kunitomi, H. *et al.* LAMC1 is a prognostic factor and a potential therapeutic target in endometrial cancer. *J. Gynecol. Oncol.* **31**, e11. <https://doi.org/10.3802/jgo.2020.31.e11> (2020).
- Li, X. *et al.* Analysis of competitive endogenous mechanism and survival prognosis of serum exosomes in ovarian cancer patients based on sequencing technology and bioinformatics. *Front. Genet.* **13**, 850089. <https://doi.org/10.3389/fgene.2022.850089> (2022).
- Elsherbini, A. M. *et al.* Decoding diabetes biomarkers and related molecular mechanisms by using machine learning, text mining, and gene expression analysis. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph192113890> (2022).
- Vivian, J. *et al.* Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* **35**, 314–316. <https://doi.org/10.1038/nbt.3772> (2017).
- Goldman, M. J. *et al.* Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678. <https://doi.org/10.1038/s41587-020-0546-8> (2020).
- Team RC Data from: R: A Language and Environment for Statistical Computing
- J F Data from: An R Companion to Applied Regression
- Wickham H Data from: ggplot2: Elegant Graphics for Data Analysis
- Chandrashekar, D. S. *et al.* UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia* **25**, 18–27. <https://doi.org/10.1016/j.neo.2022.01.001> (2022).
- Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**(6220), 1260419. <https://doi.org/10.1126/science.1260419> (2015).
- Liu, J. *et al.* An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416. <https://doi.org/10.1016/j.cell.2018.02.052> (2018).
- T T Data from: A Package for Survival Analysis in R
- Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *Bmc Bioinform.* **12**, 77. <https://doi.org/10.1186/1471-2105-12-77> (2011).
- Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397. <https://doi.org/10.1002/sim.5958> (2013).
- Cerami, E. *et al.* The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404. <https://doi.org/10.1158/2159-8290.CD-12-0095> (2012).
- Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**, 11. <https://doi.org/10.1126/scisignal.2004088> (2013).
- Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756. <https://doi.org/10.1101/gr.239244.118> (2018).
- Jardim, D. L., Goodman, A., de Melo, G. D. & Kurzrock, R. The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell* **39**, 154–173. <https://doi.org/10.1016/j.ccell.2020.10.001> (2021).
- Mroz, E. A. & Rocco, J. W. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol.* **49**, 211–215. <https://doi.org/10.1016/j.oraloncology.2012.09.007> (2013).
- Hause, R. J., Pritchard, C. C., Shendure, J. & Salipante, S. J. Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* **22**, 1342–1350. <https://doi.org/10.1038/nm.4191> (2016).

34. Li, L., Goedegebuure, S. P. & Gillanders, W. E. Preclinical and clinical development of neoantigen vaccines. *Ann. Oncol.* **28**, xii11–17. <https://doi.org/10.1093/annonc/mdx681> (2017).
35. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48**, 812–830. <https://doi.org/10.1016/j.immuni.2018.03.023> (2018).
36. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971. <https://doi.org/10.1038/ncomms9971> (2015).
37. Zanetti, M. & Mahadevan, N. R. Cancer. Immune surveillance from chromosomal chaos?. *Science* **337**, 1616–17. <https://doi.org/10.1126/science.1228464> (2012).
38. Hoppe, M. M., Sundar, R., Tan, D. & Jeyasekharan, A. D. Biomarkers for homologous recombination deficiency in cancer. *J. Natl. Cancer Inst.* **110**, 704–713. <https://doi.org/10.1093/jnci/djy085> (2018).
39. Mikhail, F. M. *et al.* Technical laboratory standards for interpretation and reporting of acquired copy-number abnormalities and copy-neutral loss of heterozygosity in neoplastic disorders: a joint consensus recommendation from the american college of medical genetics and genomics (ACMG) and the cancer genomics consortium (CGC). *Genet. Med.* **21**, 1903–1916. <https://doi.org/10.1038/s41436-019-0545-7> (2019).
40. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. <https://doi.org/10.1038/ncomms3612> (2013).
41. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795. <https://doi.org/10.1016/j.immuni.2013.10.003> (2013).
42. Yuan, H. *et al.* CancerSEA: A cancer single-cell state atlas. *Nucleic Acids Res.* **47**, D900–D908. <https://doi.org/10.1093/nar/gky939> (2019).
43. Oughtred, R. *et al.* The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200. <https://doi.org/10.1002/pro.3978> (2021).
44. 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. <https://doi.org/10.1093/nar/gkz430> (2019).
45. Li, T. *et al.* TIMER20 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, 509–514. <https://doi.org/10.1093/nar/gkaa407> (2020).
46. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omic* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
47. G C Data from: The igraph software package for complex network research
48. T P Data from: ggraph: An Implementation of Grammar of Graphics for Graphs and Networks
49. Ru, B. *et al.* TISIDB: An integrated repository portal for tumor-immune system interactions. *Bioinformatics* **35**, 4200–4202. <https://doi.org/10.1093/bioinformatics/btz210> (2019).
50. Vickers, A. J. & Elkin, E. B. Decision curve analysis: A novel method for evaluating prediction models. *Med. Decis. Making* **26**, 565–574. <https://doi.org/10.1177/0272989X06295361> (2006).
51. Bray, F., Laversanne, M., Weiderpass, E. & Soerjomataram, I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer* **127**, 3029–3030. <https://doi.org/10.1002/cncr.33587> (2021).
52. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca Cancer J. Clin.* **71**, 209–249. <https://doi.org/10.3322/caac.21660> (2021).
53. Scharping, N. E. *et al.* The Tumor Microenvironment represses T cell mitochondrial biogenesis to drive intratumoral T cell metabolic insufficiency and dysfunction. *Immunity* **45**, 374–388. <https://doi.org/10.1016/j.immuni.2016.07.009> (2016).
54. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* <https://doi.org/10.1126/science.aal3321> (2017).
55. Zoller, E. *et al.* The intermembrane space protein Mix23 is a novel stress-induced mitochondrial import factor. *J. Biol. Chem.* **295**, 14686–14697. <https://doi.org/10.1074/jbc.RA120.014247> (2020).
56. Vogtle, F. N. *et al.* Intermembrane space proteome of yeast mitochondria. *Mol. Cell Proteom.* **11**, 1840–1852. <https://doi.org/10.1074/mcp.M112.021105> (2012).
57. Von Ohlen, T., Luce-Fedrow, A., Ortega, M. T., Ganta, R. R. & Chapes, S. K. Identification of critical host mitochondrion-associated genes during Ehrlichia chaffeensis infections. *Infect Immun.* **80**, 3576–3586. <https://doi.org/10.1128/IAI.00670-12> (2012).
58. Kustatscher, G. *et al.* Co-regulation map of the human proteome enables identification of protein functions. *Nat. Biotechnol.* **37**, 1361–1371. <https://doi.org/10.1038/s41587-019-0298-5> (2019).
59. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489. <https://doi.org/10.1126/science.aab4082> (2015).
60. Gkountela, S. *et al.* Circulating tumor cell clustering shapes dna methylation to enable metastasis seeding. *Cell* **176**, 98–112. <https://doi.org/10.1016/j.cell.2018.11.046> (2019).
61. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38. <https://doi.org/10.1038/npp.2012.112> (2013).
62. Chan, T. A. *et al.* Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann. Oncol.* **30**, 44–56. <https://doi.org/10.1093/annonc/mdy495> (2019).
63. Hugo, W. *et al.* Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44. <https://doi.org/10.1016/j.cell.2016.02.065> (2016).
64. Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer—the stable evidence. *Nat. Rev. Clin. Oncol.* **7**, 153–162. <https://doi.org/10.1038/nrclinonc.2009.237> (2010).
65. Zheng, C. H. *et al.* Mutant allele frequency-based intra-tumoral genetic heterogeneity related to the tumor shrinkage mode after neoadjuvant chemotherapy in breast cancer patients. *Front. Med. (Lausanne)* **8**, 651904. <https://doi.org/10.3389/fmed.2021.651904> (2021).
66. Wu, X., Song, P., Guo, L., Ying, J. & Li, W. Mutant-allele tumor heterogeneity, a favorable biomarker to assess intra-tumor heterogeneity, in advanced lung adenocarcinoma. *Front Oncol* **12**, 888951. <https://doi.org/10.3389/fonc.2022.888951> (2022).
67. Hinshaw, D. C. & Shevde, L. A. The tumor microenvironment innately modulates cancer progression. *Cancer Res* **79**, 4557–4566. <https://doi.org/10.1158/0008-5472.CAN-18-3962> (2019).
68. Xiao, Y. & Yu, D. Tumor microenvironment as a therapeutic target in cancer. *Pharmacol. Ther.* **221**, 107753. <https://doi.org/10.1016/j.pharmthera.2020.107753> (2021).
69. Liu, S. *et al.* NK cell-based cancer immunotherapy: From basic biology to clinical development. *J. Hematol. Oncol.* **14**, 7. <https://doi.org/10.1186/s13045-020-01014-w> (2021).
70. Sakamoto, N. *et al.* Phase I clinical trial of autologous NK cell therapy using novel expansion method in patients with advanced digestive cancer. *J. Transl. Med.* **13**, 277. <https://doi.org/10.1186/s12967-015-0632-8> (2015).
71. Miller, J. S. *et al.* Successful adoptive transfer and in vivo expansion of human haploidentical NK cells in patients with cancer. *Blood* **105**, 3051–3057. <https://doi.org/10.1182/blood-2004-07-2974> (2005).
72. Schreiber, S. *et al.* Metabolic interdependency of Th2 cell-mediated type 2 immunity and the tumor microenvironment. *Front. Immunol.* **12**, 632581. <https://doi.org/10.3389/fimmu.2021.632581> (2021).
73. Shang, Q. *et al.* Polysaccharides regulate Th1/Th2 balance: A new strategy for tumor immunotherapy. *Biomed. Pharmacother.* **170**, 115976. <https://doi.org/10.1016/j.biopha.2023.115976> (2024).
74. Sun, Y., Liu, Y., Ma, X. & Hu, H. The influence of cell cycle regulation on chemotherapy. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms22136923> (2021).

75. Matthews, H. K., Bertoli, C. & de Bruin, R. Cell cycle control in cancer. *Nat. Rev. Mol. Cell Biol.* **23**, 74–88. <https://doi.org/10.1038/s41580-021-00404-3> (2022).

Author contributions

R.Z., H.Q., Y.C., K.Y., and Y.M. were involved in the study design and data interpretation. K.Y. and Y.M. contributed to finishing the manuscript. W.C. and L.L. collected the data. K.Y., Z.Y., X.C., and J.S. carried out the statistical analyses and interpreted the results. C.H., N.Z., and X.L. prepared Figures and Tables. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-59154-9>.

Correspondence and requests for materials should be addressed to Y.C., H.Q. or R.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024