

RESEARCH

Open Access



Overexpression of *CDCA8* predicts poor prognosis and drug insensitivity in lung adenocarcinoma

Huiquan Gu¹, Xinzheng Gao², Wenlong Han¹, Fangyu Wang¹, Hanqiang Zhang³, Longyu Yao³, Weimin Chen^{1,4*} and Qiang Liu^{1,4*}

Abstract

Background Lung adenocarcinoma (LUAD) accounts for the highest proportion of lung cancers; however, specific biomarkers are lacking for diagnosis, treatment, and prognostic assessment. Cell division cycle-associated 8 (CDCA8) is a cell cycle regulator with elevated expression in various cancers. However, the association between CDCA8 expression and LUAD prognosis remains unclear.

Methods The association between CDCA8 and LUAD prognosis was evaluated based on the The Cancer Genome Atlas (TCGA) dataset, and CDCA8 related functions were determined using gene enrichment and gene ontology analyses. We also analyzed the association between CDCA8 expression and immune cell infiltration. Immunohistochemistry was used to determine the differential expression of CDCA8 in tumors and controls. Finally, we evaluated the differences in the sensitivity of different levels of CDCA8 to different anticancer drugs in LUAD.

Results CDCA8 expression was significantly higher in primary LUAD tumors than in normal tissues ($P < 0.001$). Moreover, Kaplan–Meier survival analysis demonstrated that high CDCA8 expression predicted poor survival in patients with LUAD ($P = 0.006$). The receiver operating characteristic (ROC) curves indicated that CDCA8 was an effective guide for the diagnosis of LUAD. Functional annotation indicated that CDCA8 might be involved in functions such as p53 stabilization, nucleotide metabolism, RNA-mediated gene silencing, and the G2/M phase checkpoint. Immune infiltration results suggested that CDCA8 was positively correlated with Th2 cells and Tgd and negatively correlated with Eosinophils and Mast cells ($P < 0.01$). In addition, elevated expression of CDCA8 may increase the sensitivity of patients to certain anticancer drugs.

Conclusions CDCA8 upregulation is significantly associated with poor survival and immune infiltration in patients with LUAD. Our study suggests that CDCA8 can be used as a biomarker for LUAD prognosis and a reference for personalized medication.

[†]Huiquan Gu and Xinzheng Gao contributed equally to this work.

*Correspondence:

Weimin Chen

470048098@qq.com

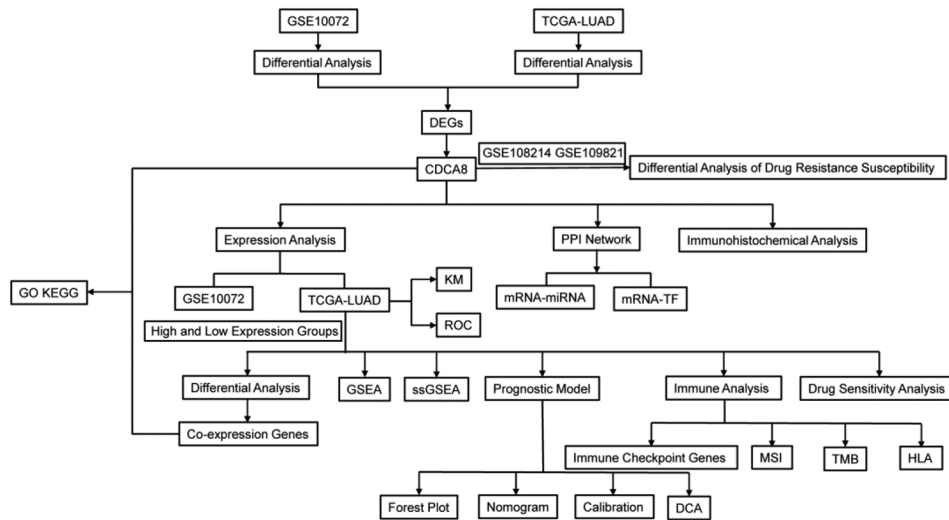
Qiang Liu

744526211@qq.com

Full list of author information is available at the end of the article



Graphical Abstract



Keywords LUAD, CDCA8, Biomarker, Prognosis, Sensitivity

Introduction

Lung cancer has become one of the most common malignant tumors worldwide in recent years, with a reported rate of approximately 2,206,700 new cases and 1,796,100 deaths worldwide in 2020; it accounts for 18% of all cancer deaths and is the leading cause of death from malignant tumors worldwide [1]. Lung adenocarcinoma (LUAD), the most common type, is advanced in many patients because of the lack of early symptoms, leading to poor treatment and prognosis of LUAD [2]. With the development of detection technologies, the emergence of molecularly targeted drugs has transformed the treatment of LUAD into standard first-line therapy [3]. However, not all patients benefit from these treatments, and many molecular targets have not yet been identified [4]. Therefore, there is an urgent need to screen novel biomarkers for the early diagnosis and subsequent treatment of patients with lung cancer.

CDCA8 is a member of the Cell division cycle associated protein (CDCA) family of genes and is associated with Aurora B, INCENP, and Survivin, which form an essential component of the chromosomal passenger complex (CPC) [5]. Structurally, CDCA8 binds directly to Survivin and INCENP and exhibits a triple helix-like structure in vitro [6]. In embryonic stem cells, CDCA8 can be localized to the central spindle and intermediates through the N-terminal 141 residues already interacting with Survivin. It regulates the stability of mitotic granules during mitosis. In addition, CDCA8 is expressed at low levels or is not expressed in normal tissues. CDCA8 is aberrantly expressed in malignant tumors such as hepatocellular carcinoma [7], prostate cancer [8], ovarian

cancer [9], and melanoma [10]. It is also associated with a poor clinical prognosis. Recent studies have revealed that CDCA8 may contribute to the development of endometrial cell carcinoma by mediating the cell cycle and the P53/Rb pathway [11]. CDCA8 silencing can promote tumor cell apoptosis and increase cell sensitivity to laparib and cisplatin by inhibiting the G2/M phase [12].

Although previous studies have identified CDCA8 overexpression in a variety of cancers, including LUAD, our study aimed to further extend this knowledge. By exploring the prognostic significance of CDCA8 and its potential role in drug resistance and immune cell infiltration, we performed a comprehensive integrated analysis. Employing bioinformatics tools, survival analysis, immune infiltration analysis, and drug susceptibility prediction, we provided a more comprehensive insight into CDCA8 in LUAD. This comprehensive analysis not only validated the overexpression of CDCA8, but also revealed its potential application as a multifunctional biomarker, providing a new scientific basis for future therapeutic strategies in LUAD patients.

Materials and methods

Data download

From the TCGA database [13], a total of 598 LUAD clinical data samples were obtained, including 539 LUAD patient tumor tissues and 59 LUAD patient para-cancer tissues, which were normalized in Fragments Per Kilo-baseper Million (FPKM) format. TCGA-LUAD counts, sequencing results, and corresponding FPKM-formatted data were normalized using the limma package [14]. The total baseline data of TCGA-LUAD and the baseline data

of the different expression level groups based on CDCA8 are summarized in Table 1.

The gene expression profile data of LUAD related datasets GSE10072 [15], GSE108214 [16], and GSE109821 was downloaded from GEO database through the R package GEOquery [17]. From the GSE10072 dataset, we chose to include 58 LUAD samples and 49 control samples for this study. The GSE108214 dataset was derived from non-small-cell lung cancer cells, including 15 cisplatin-resistant and 7 cisplatin-sensitive samples. All the above-mentioned samples were enrolled in this study. The dataset GSE109821 was obtained from *Homo sapiens*. The data platform was GPL16791 Illumina HiSeq 2500, the sample data for which the sequencing instrument was BCM was selected, and the sample source was adenocarcinoma of the lung. The count sequencing data of 5 resistant samples and 37 sensitive samples were included and standardized in the FPKM format.

Differential expression analysis and prognostic analysis of LUAD

According to the grouping of the TCGA-LUAD dataset, the samples were categorized as LUAD or paracancerous. The DEGs in the above two groups were analyzed using the R package limma [14]. $|\logFC| > 0.5$ and $\text{adj. } P < 0.05$ as the critical values of DEGs. The ANOVA results were used to plot a volcano map using the ggplot2 R package.

Table 1 Baseline Data table based on CDCA8 high-low expression grouping

characteristics	Low expression of CDCA8	High expression of CDCA8	p value	statistic	method
n	269	270			
Pathologic stage, n (%)			0.00994016	11.3578516	Chisq test
Stage I	165 (31.1%)	131 (24.7%)			
Stage II	56 (10.5%)	69 (13%)			
Stage III	32 (6%)	52 (9.8%)			
Stage IV	10 (1.9%)	16 (3%)			
Gender, n (%)			0.0107378	6.50820172	Chisq test
female	159 (29.5%)	130 (24.1%)			
Male	110 (20.4%)	140 (26%)			
Age, n (%)			0.00287116	8.88758793	Chisq test
<= 65	112 (21.5%)	145 (27.9%)			
> 65	149 (28.7%)	114 (21.9%)			

TCGA-LUAD data set the intersection by differences in genes, and GSE10072 map Wayne to display. The expression of CDCA8 in different groups of TCGA-LUAD and GSE10072 is shown in a group comparison plot.

For the prognostic analysis of CDCA8, we combined the clinical prognostic information of the LUAD group Overall Survival (OS) and OS time in TCGA-LUAD. We also plotted a Kaplan–Meier (KM) curve for the relationship between CDCA8 expression and patient survival and prognosis.

Analysis of different levels of CDCA8 differential gene

To clarify the differentially expressed genes and their potential mechanisms, related biological features, and pathways in LUAD in different level groups of CDCA8, related biological features, and pathways, we removed normal samples from the dataset TCGA-LUAD and bound it by the median CDCA8 expression. To obtain the genes co-expressed with CDCA8, we sorted the logFC after removing the normal samples from TCGA-LUAD, screened the top 15 saliently significant differentially expressed genes, and plotted a co-expression heat map.

Functional enrichment and pathway enrichment analysis via genomic enrichment analysis

We used the R package clusterProfiler to perform GO annotation analysis [18] and KEGG [19] on CDCA8; the top 15 significantly upregulated and downregulated genes and the top 15 significantly differentially expressed genes were subjected to GO annotation analysis and KEGG analysis using the clusterProfiler R package [20]. The screening guidelines were $\text{adj. } P < 0.05$ and $\text{FDR} < 0.05$. P -values were corrected using the Benjamini–Hochberg (BH) test. Finally, the associated pathway map visualization for KEGG enrichment analysis was demonstrated using the R package Pathview12 [21].

Gene Set Enrichment Analysis (GSEA)

We categorized the patients into high- and low-expression groups based on the median expression value of CDCA8. GSEA [22] was performed on all genes in the LUAD group of the TCGA-LUAD dataset based on logFC values using the R package clusterProfiler. The GSEA used in the set of parameters was as follows: the number of seeds was 2022, the number of calculations was 1000, and the number of genes included in the genome was set to a minimum of 10 and a maximum of 500. Gene set enrichment analysis (GSEA) was performed by obtaining gene set c2.cp.all.v2022.1.Hs.symbols.gmt [All Canonical Pathways] (3050) from the Molecular Signatures Database (MsigDB) [23]. The screening criteria for GSEA were $\text{adj. } P < 0.05$, $\text{FDR} < 0.05$, The P -values were corrected using BH.

Protein-protein interaction (PPI) network

We constructed a *CDCA8*-related PPI network based on *CDCA8* in the STRING database [24], with interaction scores >0.40. The GeneMANIA website [25] was used to predict the function of selected genes, similar genes, and their interacting proteins in the PPI network, as well as to construct the interaction network.

Construction of regulatory network

We mapped the miRNA network interacting with *CDCA8* by selecting data segments with a Target Score >60 using the MiRDB database [26]. We then retained the portion of TFs that were searched in the CHIPBase (version 3.0) [27] and HTFtarget databases [28] for binding with *CDCA8* and visualized them using Cytoscape software. The data are summarized in Supplemental Fig. 1.

Immune infiltration analysis

The enrichment scores calculated using ssGSEA in the R package represented the extent of infiltration of each immune cell type in each sample [29, 30]. Box and correlation Laplace plots were used to show the abundance of immune cell infiltration in tumor samples from the *CDCA8* differentially expressed group. Finally, we selected the two immune cells with the highest positive and negative correlations with the target gene *CDCA8* to plot the correlation scatter plots.

Construction of clinical prognostic model

Based on the univariate Cox regression analysis, we evaluated the clinical prognostic value of *CDCA8* in LUAD. After including variables with $P < 0.001$ in the multivariate Cox regression analysis, a multivariate Cox regression model was constructed. Nomograms were used to predict 1-, 3-, and 5-year survival in patients with LUAD. Calibration curves were used to assess the nomogram accuracy and resolution.

Immune checkpoint genes (ICG), microsatellite instability (MSI), TMB, HLA expression analysis

We screened 50 ICGs from the published literature (Table S1). We then analyzed the differences in ICG expression between subgroups with different expression levels of *CDCA8* in LUAD samples from TCGA-LUAD and plotted subgroup comparisons. We also calculated the Tumor Mutation Burden (TMB) of different *CDCA8* expression level groups in TCGA-LUAD samples using the U-test. Group differences in MSI and scores were also analyzed.

We searched the GeneCards genes with names beginning with HLA. A total of 21 HLA family genes were obtained and analyzed for differences in their expression between the high and low *CDCA8* groups in the

TCGA-LUAD samples, with comparative plots between groups (Table S2).

Drug sensitivity analysis

By searching the GDSC database (www.cancerRxgene.org) [31] and using the pRRophetic algorithm [32], based on the expression matrix of the TCGA-LUAD dataset in FPKM format, *CDCA8* was predicted from the TCGA-LUAD dataset by calculating the IC50 values of the sensitivity of the patients with LUAD to common anticancer drugs or small-molecule compounds. Additionally, the relationship between different expression levels of *CDCA8* and drug sensitivity in the TCGA-LUAD dataset was predicted. Results are presented in the form of subgroup comparison plots.

Immunohistochemical analysis

The expression of *CDCA8* in LUAD and normal lung gland tissues was analyzed via immunohistochemistry using the Human Protein Atlas (HPA) database [33]. IHC results for *CDCA8* in human cells from the database are displayed.

Comparison analysis between *CDCA8* resistant and sensitive groups

To assess changes in the *CDCA8* gene in the LUAD-resistant and LUAD-sensitive groups, we used the datasets GSE108214 and GSE109821. Intergroup comparison plots were used to show the differences between the target gene *CDCA8* in the resistant and sensitive groups and whether the trends were statistically significant.

Statistical analysis

Data processing was performed using R software (version 4.2.3). The Wilcoxon rank-sum test was performed to assess differences between the two groups. Kaplan-Meier survival curves showed differences between survival rates. Differences in survival time were assessed using the log-rank test. P -values were two-sided, and statistical significance was set at $P < 0.05$.

Results

Differentially expressed genes in LUAD

The data from the GSE10072 dataset were split into LUAD and control groups. To analyze the differences between the LUAD and para-carcinoma groups in the TCGA-LUAD and GSE10072 datasets, the R package limma was used to obtain DEGs for both groups. The results were as follows: TCGA LUAD - a total of 1669 data sets satisfied $|\log_{2}FC| > 0.5$ and $\text{adj. } P < 0.05$, the threshold of DEGs; a total of 704 genes were up-regulated; a total of 965 genes were down-regulated, according to the variance analysis results of the dataset map volcano (Fig. 1A). GSE10072 datasets, a total of 453 met |

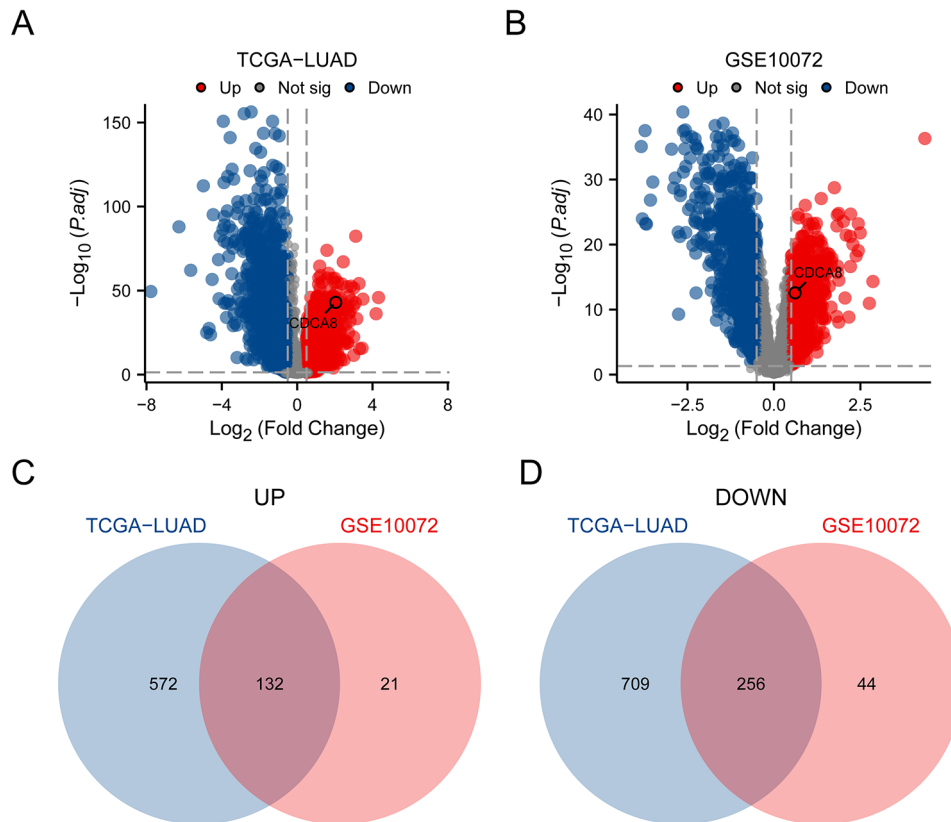


Fig. 1 Differential analysis of gene expression between TCGA-LUAD and GSE10072. **(A)** Volcano plot of differential genes between the LUAD and the paraneoplastic in TCGA-LUAD. **(B)** Volcano plot of differential genes between the LUAD and the control group in GSE10072. **(C)** TCGA-LUAD dataset with improved Wayne plots of differential genes in GSE10072. **(D)** Venn diagram of down-regulated genes in TCGA-LUAD and GSE10072 dataset. TCGA-LUAD dataset: $n=539$ (LUAD) and $n=59$ (Normal), GSE10072 dataset: $n=58$ (LUAD) and $n=49$ (Normal)

$\log_{2}FC > 0.5$ and adj. $P < 0.05$ threshold of DEGs, up-regulation of expressed genes at this threshold, a total of 153, there were 300 down-regulated genes, and a volcano map was drawn according to the difference analysis results of this dataset (Fig. 1B). To obtain the differential genes with the same expression changes in the TCGA-LUAD and GSE10072 datasets, the intersection of upregulated and downregulated differential genes in each of the two datasets was plotted as a Venn diagram (Fig. 1C-D). Among the upregulated genes in the two datasets, there were 132 common genes, and among the downregulated genes in the two datasets, there were 256 common genes.

Differential analysis of CDCA8 expression

To explore the difference in CDCA8 expression between TCGA-LUAD and GSE10072, we used group comparison plots in the TCGA-LUAD and GSE10072 datasets to determine whether the expression of the target gene CDCA8 in the LUAD and control groups was statistically significant (Fig. 2A-B). The expression of CDCA8 in the two datasets was significantly different ($P < 0.001$). According to the results in the TCGA datasets LUAD and GSE10072, CDCA8 expression in the cancer group

was significantly increased. Subsequently, a prognostic survival KM curve was drawn based on the expression of CDCA8 and the related prognostic data (Fig. 2C). Statistical significance was set at $P < 0.05$. The prognosis of the CDCA8 high expression group was worse. Finally, we plotted the ROC curves of CDCA8 in the TCGA-LUAD and GSE10072 (Fig. 2D-E) datasets, and the results showed that CDCA8 was highly accurate in assessing tumorigenesis.

Differences between groups with different expression levels of CDCA8

We first analyzed variance on the LUAD genes in the samples by using the R package, high and low expression group FPKM data to $|\log_{2}FC| > 0.5$ and adj. $P < 0.05$ standard screening gene as a difference. Volcano mapping revealed the localization of CDCA8 (Fig. 3A). We also selected the top 15 positively correlated differentially expressed genes found in the results of differential analysis by sorting them in ascending and descending logFC columns (Fig. 3B, positive correlation top 10: MAGEA4, DPPA2, MAGEA9B, HOXD13, GAGE2A, MAGEA10, MAGEB2, SP9, SLC6A15, CDH18, MAGEC1, PAGE1,

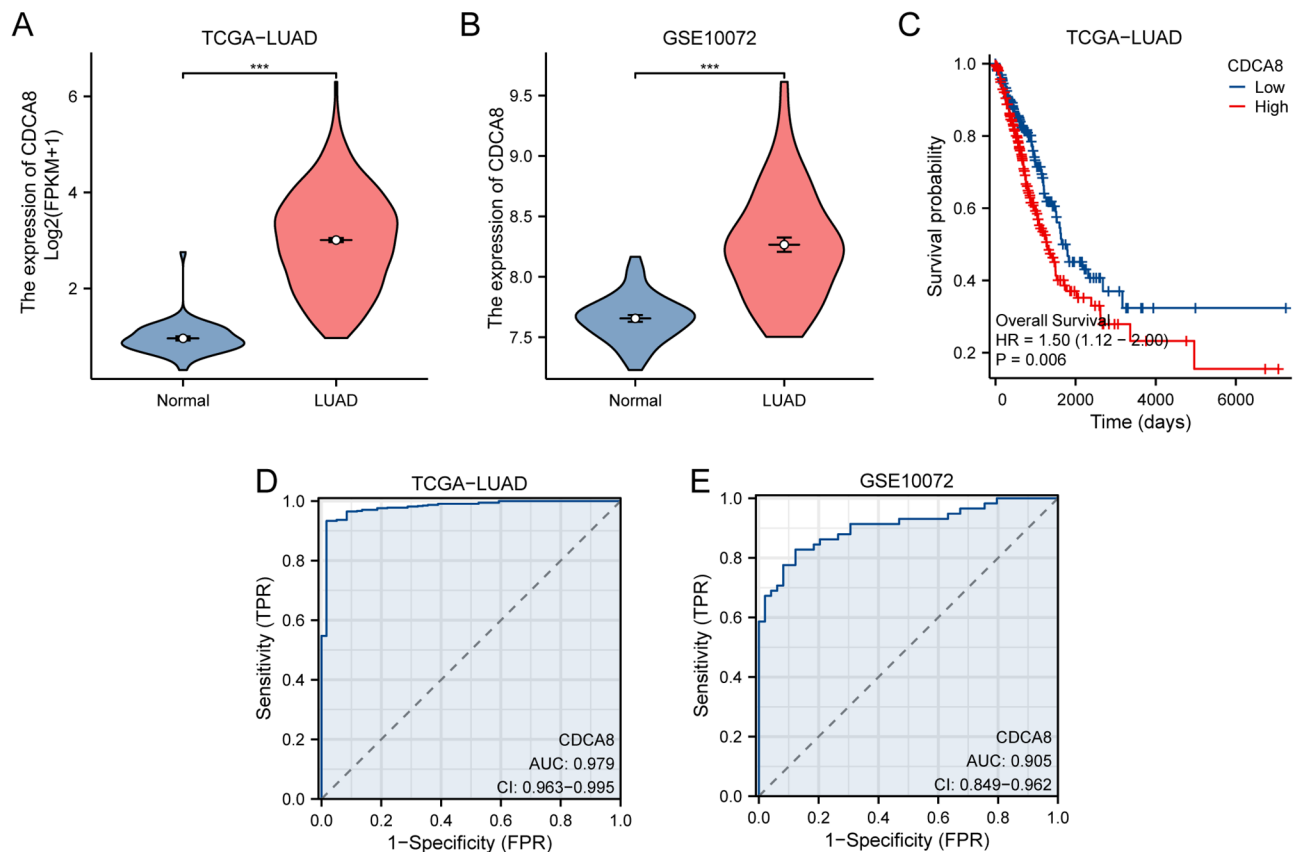


Fig. 2 Differential expression analysis of CDCA8. **(A)** Comparison of CDCA8 expression groups in the TCGA-LUAD. **(B)** Comparison of differential expression groups of CDCA8 in GSE10072 dataset. **(C)** Prognostic KM curves between CDCA8 high and low groups and overall survival of LUAD samples. **(D)** ROC curve of CDCA8 in the TCGA-LUAD. **(E)** ROC curve of CDCA8 in the GSE10072 dataset. TCGA-LUAD dataset: $n = 539$ (LUAD) and $n = 59$ (Normal), GSE10072 dataset: $n = 58$ (LUAD) and $n = 49$ (Normal)

SPANXB1, CT45A1, CASP14) and the top 15 negatively correlated differentially expressed genes (Fig. 3C, negative correlation top 10: PGC, PCSK2, GKN2, MAB21L2, SLC10A2, REG1A, H1-1, H4C6, SCGB3A2, H4C13, AMELX, H2BC3, H4C3, AL138752.2, SULT1C3) as other molecules, *CDCA8* was used as the target molecule to further analyze the correlation between them, and the results were displayed by single gene co-expression heat map (Fig. 3B-C).

Functional enrichment and pathway enrichment analyses of CDCA8 and its co-expressed genes

Functional enrichment analysis (GO) was used to further explore the relationship between *CDCA8*, 30 co-expressed genes, and LUAD. *CDCA8* and 30 co-expressed genes were used for GO and KEGG analyses (Table 2), and the results were visualized using a bar chart (Fig. 4A). The results showed that *CDCA8* and 30 co-expressed genes were mainly enriched in nucleosome organization, chromatin assembly, and other biological processes (Fig. 4B); nucleosome, CENP-A-containing nucleosome, CENP-A-containing chromatin,

chromosome, centromeric core domain, DNA packaging complex, and other cellular components (Fig. 4C); histone deacetylase binding, organic acid, and sodium symporter activity; and molecular functions such as protein heterodimerization activity (Fig. 4D). The enriched KEGG pathways in LUAD included systemic lupus erythematosus, alcoholism, viral carcinogenesis (Fig. 4E).

We also analyzed the results of KEGG pathway enrichment in *CDCA8* and co-expressed genes for viral carcinogenesis, alcoholism, and systemic lupus erythematosus (Fig. 4F-I).

Gene set enrichment analysis

To determine the effect of the differential expression of *CDCA8* in TCGA-LUAD, we performed a genomic enrichment analysis to investigate the involvement and related functions of all genes in the LUAD group (Fig. 5A). The results are listed in Table 3. The enrichment results indicated that the DEGs in TCGA-LUAD samples were highly enriched in pyrimidine metabolism (Fig. 5B), stabilization of p53 (Fig. 5C), metabolism of nucleotides (Fig. 5D), metabolic reprogramming in colon

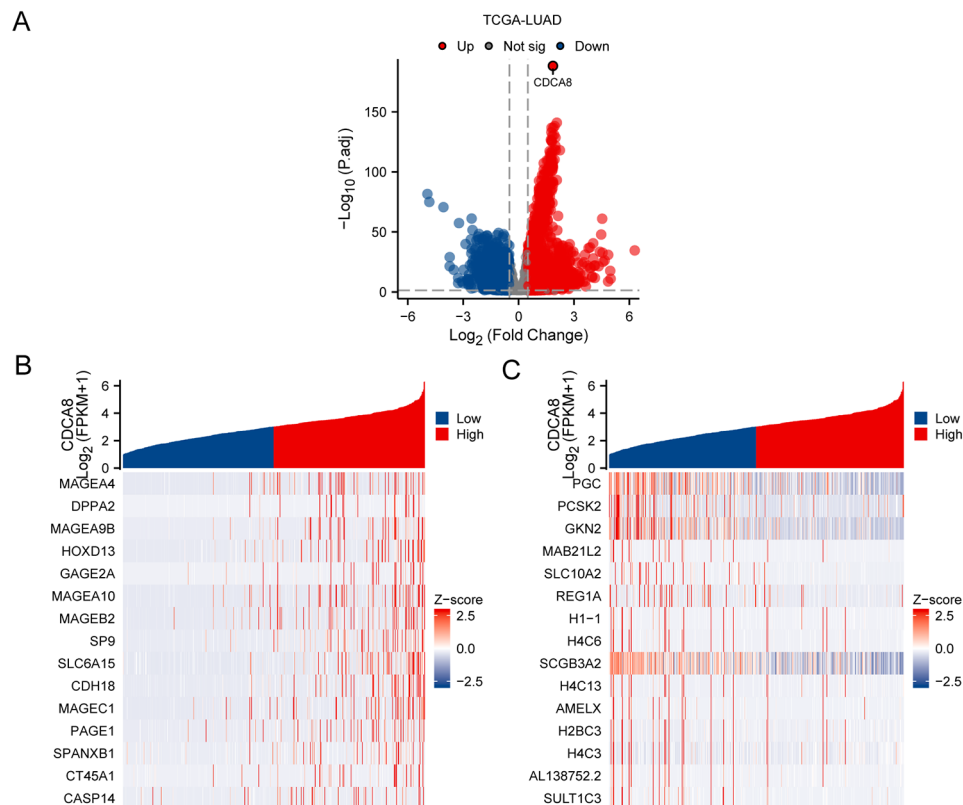


Fig. 3 Differential analysis of groups with different expression levels of *CDCA8*. **(A)** Volcano map of *CDCA8* differential expression. **(B-C)** Single gene co-expression heat map of gene *CDCA8*. TCGA-LUAD dataset: $n = 539$ (LUAD) and $n = 59$ (Normal)

cancer (Fig. 5E), pyrimidine metabolism (Fig. 5F), metabolism of polyamines (Fig. 5G), gene silencing by RNA (Fig. 5H) and other biologically related functions and signaling pathways.

PPI network

PPI analysis of *CDCA8* was performed using the STRING database with a minimum requirement of medium confidence (0.400), and a set of 10 *CDCA8*-related genes was constructed, namely *ATP5F1A*, *AURKB*, *BIRC5*, *BUB1B*, *CCNB1*, *CDC20*, *CDK1*, *INCENP*, *KIF20A*, and *SGO1* (Fig. 6A). Subsequently, the interaction network of the 11 genes was predicted and constructed using the GeneMANIA website (Fig. 6B) to observe co-expression and other related information.

LUAD dataset immune infiltration analysis

The ssGSEA algorithm was used to count 24 types of immune cells in the *CDCA8* differentially expressed group of TCGA-LUAD, and the Wilcoxon test algorithm was used to compare differences in infiltration levels. The results showed that the difference in the infiltration levels of 19 immune cells between the two groups was significant ($P < 0.05$) (Fig. 7A), in which CD8 T cells, dendritic cells, eosinophils, immature dendritic cells, mast cells,

NK CD56dim cells, NK cells, central memory T cells, follicular helper T cells, γ δ T cells, T helper type 17 cells, and T helper type 2 cells in the *CDCA8* high and low expression groups were highly statistically significant ($P < 0.001$). The expression of aDC and pDC significantly differed between groups ($P < 0.01$). The expression levels of B cells, macrophages, T cells, central memory CD8+ T cells, and regulatory T cells differed significantly between the groups ($P < 0.05$).

Subsequently, we calculated the correlations between the 19 immune cells and *CDCA8* and visualized them with a Laplace plot (Fig. 7B). We selected the two most positively correlated immune cell types, Th2 cells and Tgd, and the two most negatively correlated immune cell types, mast cells, and eosinophils, for correlation scatter plot visualization (Fig. 7C-F).

Construction of a prognostic risk model for LUAD

To determine the prognostic value of *CDCA8* in the TCGA-LUAD dataset, we first counted the LUAD samples obtained from the TCGA-LUAD replicated dataset and statistically analyzed the clinical information of the patients. We then performed a univariate Cox regression analysis based on *CDCA8* levels combined with clinical variables (stage, age, and sex), and a multivariate Cox

Table 2 Result of GO and KEGG Enrichment Analysis for CDCA8 and coexpressed genes

ONTOLOGY	ID	GeneRatio	BgRatio	p value	adj.p	qvalue
BP	GO:0045653	2023/3/28	18/18,800	2.38 e-06	0.000234	0.00018
BP	GO:0034728	2023/5/28	159/18,800	3.41 e-06	0.000234	0.00018
BP	GO:0065004	2023/5/28	203/18,800	1.12 e-05	0.000393	0.000301
BP	GO:0031497	2023/5/28	205/18,800	1.18 e-05	0.000393	0.000301
BP	GO:0006335	2023/3/28	32/18,800	1.43 e-05	0.000393	0.000301
BP	GO:0034723	2023/3/28	32/18,800	1.43 e-05	0.000393	0.000301
BP	GO:0006336	2023/3/28	33/18,800	1.57 e-05	0.000393	0.000301
BP	GO:0034724	2023/3/28	34/18,800	1.72 e-05	0.000393	0.000301
BP	GO:0045652	2023/3/28	36/18,800	2.04 e-05	0.000421	0.000323
BP	GO:0071824	2023/5/28	237/18,800	2.37 e-05	0.000443	0.00034
BP	GO:0006338	2023/5/28	266/18,800	4.11 e-05	0.000706	0.000541
BP	GO:0030219	2023/3/28	57/18,800	8.20 e-05	0.0013	0.000996
BP	GO:0045638	2023/3/28	91/18,800	0.000329	0.004845	0.003713
BP	GO:0006352	2023/3/28	134/18,800	0.001018	0.013982	0.010717
BP	GO:0032200	2023/3/28	162/18,800	0.001756	0.022614	0.017333
BP	GO:0045637	2023/3/28	208/18,800	0.003566	0.043216	0.033124
CC	GO:0000786	2023/5/29	129/19,594	1.20 e-06	2.64 e-05	1.76 e-05
CC	GO:0043505	2023/3/29	18/19,594	2.34 e-06	2.64 e-05	1.76 e-05
CC	GO:0061638	2023/3/29	18/19,594	2.34 e-06	2.64 e-05	1.76 e-05
CC	GO:0034506	2023/3/29	19/19,594	2.78 e-06	2.64 e-05	1.76 e-05
CC	GO:0044815	2023/5/29	198/19,594	9.77 e-06	7.42 e-05	4.93 e-05
CC	GO:0032993	2023/5/29	220/19,594	1.63 e-05	0.000103	6.84 e-05
CC	GO:0000781	2023/4/29	166/19,594	0.0001	0.000543	0.000361
CC	GO:0098687	2023/5/29	366/19,594	0.000182	0.000863	0.000574
CC	GO:0000775	2023/4/29	227/19,594	0.000332	0.001401	0.000932
CC	GO:0000228	2023/3/29	228/19,594	0.004545	0.017271	0.011482
MF	GO:0042826	2023/3/29	126/18,410	0.001004	0.025859	0.014945
MF	GO:0005343	2/29	30/18,410	0.001014	0.025859	0.014945
MF	GO:0046982	2023/4/29	332/18,410	0.001728	0.029371	0.016974
KEGG	hsa05322	2023/4/5	136/8164	3.64 e-07	1.82 e-06	3.83 e-07
KEGG	hsa05034	2023/4/5	187/8164	1.31 e-06	2.32 e-06	4.89 e-07
KEGG	hsa04613	2023/4/5	190/8164	1.40 e-06	2.32 e-06	4.89 e-07
KEGG	hsa05203	2023/4/5	204/8164	1.86 e-06	2.32 e-06	4.89 e-07

GO, Gene Ontology; BP, Biological Process; CC, Cellular Component; MF, Molecular Function

prognostic model was constructed by including variables with $P < 0.001$ (Table 4). We then present the results of the univariate Cox regression in the form of a forest plot (Fig. 8A). We will obtain the model of the risk score multivariable Cox RiskScore for TCGA datasets with the median value - LUAD sample of high- and low-risk groups.

We then determined the prognostic power of the model by analyzing the nomograms (Fig. 8B). In addition, we performed 1-,3-,5-year prognostic calibration analyses and plotted calibration curves for the column line plots of the multifactorial Cox prognostic model (Fig. 8C-E).

We then used DCA to evaluate and present the results of the constructed multivariate Cox model in terms of clinical utility at 1-,3-,5-year (Fig. 8F-H). The multivariate Cox model we constructed was more accurate for clinical prediction at the 3-year and 5-year periods than at the 1-year. Subsequently, we built a CDCA8 Cox prognostic

model of gene expression for the prognosis of the Cox model samples for visualization (Fig. 8I). We combined the prognostic information of patients with LUAD and plotted a time-dependent ROC curve (Fig. 8J) to demonstrate the effect of risk scores from the multivariate Cox prognostic model on survival outcomes.

ICG, MSI, TMB, HLA analysis

We analyzed the differences in MSI and TMB between CDCA8 differential expression groups in the LUAD group based on TCGA-LUAD. There was no statistically significant difference in MSI in the CDCA8 differential expression group ($P > 0.05$; Fig. 9A). However, the TMB of the LUAD group was remarkably different from that of the CDCA8 differentially expressed groups ($P < 0.001$; Fig. 9B).

We also obtained information on ICGs and HLA family genes from published literature, the GeneCards database,

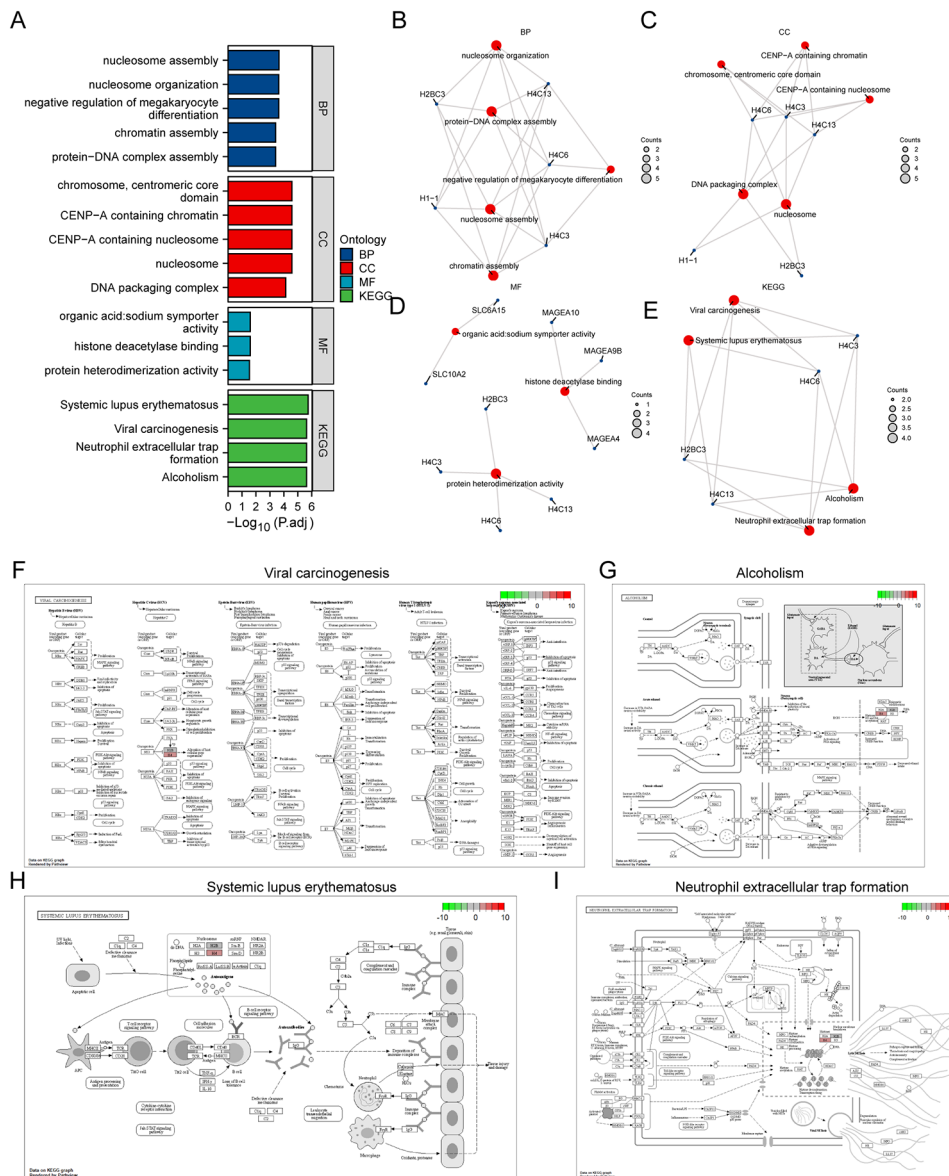


Fig. 4 Enrichment analysis of the gene CDCA8. **(A)** CDCA8 and enrichment of expressed genes function analysis and pathway enrichment analysis histogram analysis results show. **(B-E)** Mesh plot of the results of functional enrichment analysis and KEGG analysis of CDCA8 with co-expressed genes. **(F-I)** KEGG analysis of CDCA8 and co-expressed genes

and other sources. After crossing with TCGA-LUAD genes, a matrix consisting of 30 ICGs and their expression levels was obtained, as listed in Table S5. A matrix of 19 HLA family genes and their corresponding expression levels was obtained, as listed in Table S6.

Finally, we combine TCGA - LUAD dataset CDCA8 grouping situation of high and low expression group use the Mann - Whitney U test to explore immune checkpoint genes expressed in CDCA8 statistical differences between groups (Fig. 9C). The results showed that the immune checkpoint genes *BTLA*, *CD28*, *CD27*, *CD40LG*, *CD48*, *BTN2A2*, *BTNL9*, *CD96*, and *TDO2* were significantly different between the *CDCA8* differential

expression groups ($P < 0.001$). *HHLA2* expression was significantly different between the *CDCA8* differential expression groups ($P < 0.01$). *IDO1* and *BTN3A1* were statistically significant between *CDCA8* differential expression groups ($P < 0.05$).

Finally, we combined the TCGA-LUAD dataset *CDCA8* grouping situation of high and low expression groups using the Mann-Whitney U test to explore the family of HLA genes expressed in *CDCA8* statistical differences between groups (Fig. 9D). The results showed that HLA family genes, such as *HLA-DMA*, *HLA-DQA1*, and *HLA-DRB5*, were statistically significant ($P < 0.001$) between the *CDCA8* differentially expressed groups in

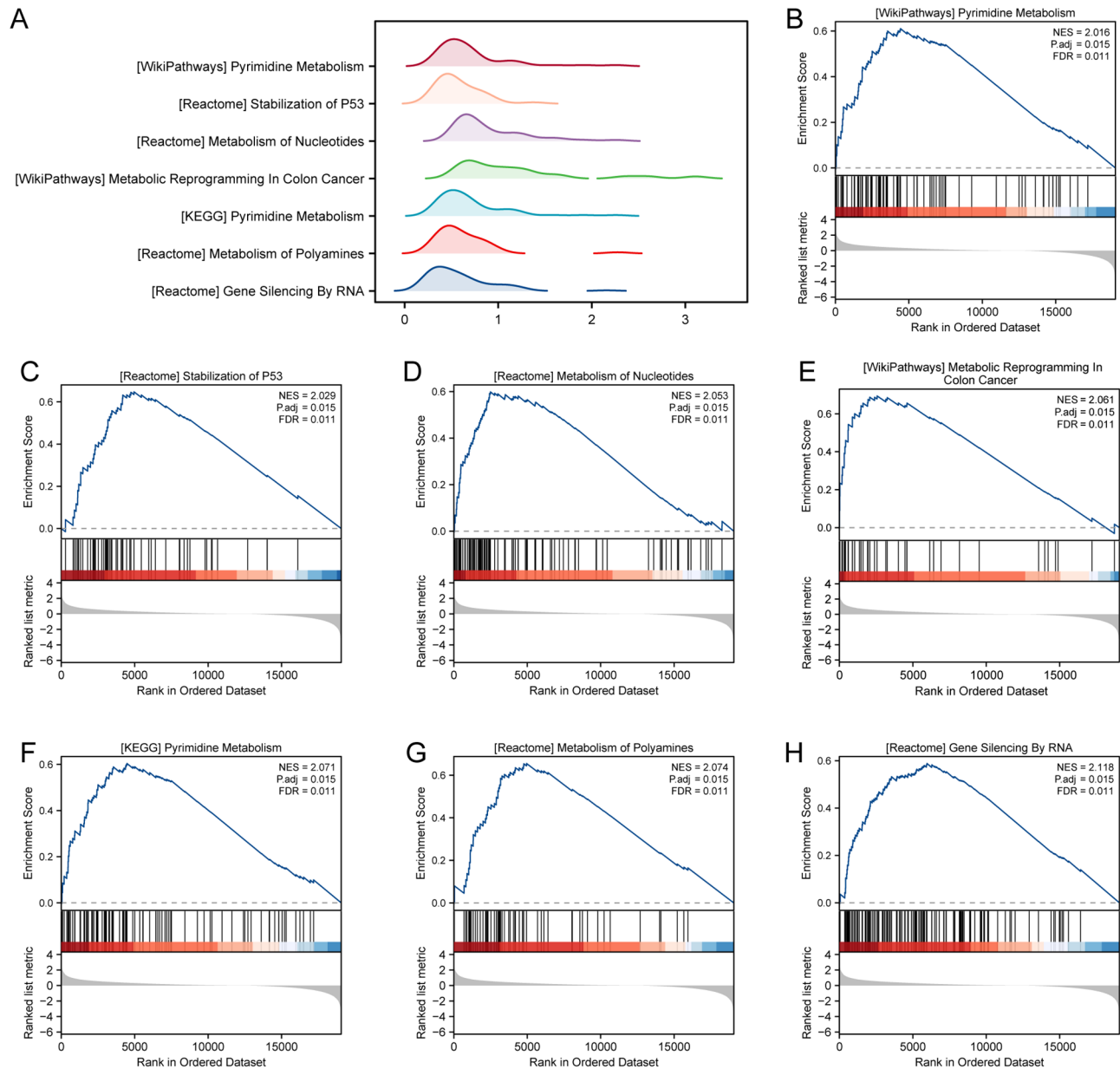


Fig. 5 GSEA enrichment analysis of LUAD samples in the TCGA-LUAD. **(A)** GSEA seven mountains figure display biology function. **(B)** WP_pyrimidine metabolism. **(C)** REACTOME_stabilization of p53. **(D)** REACTOME_metabolism of nucleotides. **(E)** WP_metabolic reprogramming in colon cancer. **(F)** KEGG_pyrimidine metabolism. **(G)** REACTOME_metabolism of polyamines. **(H)** REACTOME_gene silencing by RNA. TCGA-LUAD dataset: $n = 539$ (LUAD) and $n = 59$ (Normal)

the TCGA LUAD dataset; *HLA - DQA2* was statistically significant in the CDCA8 differential expression group of the TCGA LUAD dataset ($P < 0.05$).

Drug sensitivity analysis of CDCA8 differential expression groups

To explore suitable therapeutic strategies for mRNA vaccination in patients with CDCA8 differential expression, we used drug sensitivity data from the GDSC database as a training set to predict the sensitivity of samples in the CDCA8 differential expression groups

to common anticancer drugs in TCGA-LUAD. We then used the Mann–Whitney U test to evaluate the TCGA LUAD dataset CDCA8 in the LUAD group in the differentially expressed groups LUAD sensitivity to different anticancer drugs. We kept CDCA8 high and low expression groups with relatively large differences in the top 20 drugs: CCT007093, Nutlin.3a, PD.0332991, MK.2206, AS601245, Bicalutamide, FH535, Roscovitine, VX.702, Erlotinib, PF.02341066, Chr.99,021, and BMS. 754,807, LFM A13, AZD6244, JNK. 9 l, GDC0941, DMOG, PD. 0325901, and AZD8055, and the results are

Table 3 Results of GSEA for TCGA-LUAD

ID	setSize	enrichmentScore	NES	P value	adj.p	q value
REACTOME_GENE_SILENCING_BY_RNA	136	0.588189	2.117929	0.002247	0.015083	0.010503
REACTOME_METABOLISM_OF_POLYAMINES	59	0.655156	2.074453	0.002008	0.015083	0.010503
KEGG_PYRIMIDINE_METABOLISM	97	0.604081	2.070613	0.002075	0.015083	0.010503
WP_METABOLIC_REPROGRAMMING_IN_COLON_CANCER	42	0.694271	2.060654	0.001988	0.015083	0.010503
REACTOME_METABOLISM_OF_NUCLEOTIDES	97	0.599061	2.053406	0.002075	0.015083	0.010503
REACTOME_STABILIZATION_OF_P53	57	0.647422	2.029098	0.002058	0.015083	0.010503
WP_PYRIMIDINE_METABOLISM	82	0.610423	2.016468	0.002041	0.015083	0.010503
REACTOME_AUF1_HNRNP_D0_BINDS_AND_DESTABILIZES_MRNA	55	0.644924	2.007902	0.002037	0.015083	0.010503
REACTOME_PRC2_METHYLATES_HISTONES_AND_DNA	70	0.615519	1.979317	0.002053	0.015083	0.010503
REACTOME_REGULATION_OF_TP53_ACTIVITY_THROUGH_PHOSPHORYLATION	92	0.582505	1.976902	0.002066	0.015083	0.010503
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_TP53	359	0.487299	1.945361	0.002353	0.015083	0.010503
REACTOME_HDACS_DEACETYLATE_HISTONES	92	0.571373	1.939123	0.002066	0.015083	0.010503
WP_AEROBIC_GLYCOLYSIS	12	0.896771	1.930784	0.001934	0.015083	0.010503
REACTOME_DNA_DAMAGE_TELOMERE_STRESS_INDUCED_SENESCENCE	79	0.58548	1.926922	0.002058	0.015083	0.010503
REACTOME_HATS_ACETYLATE_HISTONES	140	0.52996	1.911469	0.002278	0.015083	0.010503
REACTOME_REGULATION_OF_TP53_ACTIVITY	160	0.498416	1.841011	0.002288	0.015083	0.010503
REACTOME_RMTS_METHYLATE_HISTONE_ARGININES	77	0.554635	1.821045	0.002041	0.015083	0.010503
REACTOME_TP53_REGULATES_TRANSCRIPTION_OF_GENES_INVOLVED_IN_G2_CELL_CYCLE_ARREST	18	0.733009	1.801823	0.003817	0.021713	0.015119
REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES	61	0.56237	1.792109	0.002008	0.015083	0.010503
REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES	373	0.446735	1.789221	0.002404	0.015083	0.010503
REACTOME_NEGATIVE_REGULATION_OF_NOTCH4_SIGNALING	54	0.576921	1.787628	0.002049	0.015083	0.010503
WP_PURINE_METABOLISM_AND_RELATED_DISORDERS	22	0.687951	1.77983	0.001876	0.015083	0.010503
REACTOME_REGULATION_OF_MRNA_STABILITY_BY_PROTEINS_THAT_BIND_AU_RICH_ELEMENTS	87	0.529096	1.770848	0.002058	0.015083	0.010503
REACTOME_TP53_REGULATES_TRANSCRIPTION_OF_DNA_REPAIR_GENES	62	0.549345	1.751743	0.004024	0.021713	0.015119
REACTOME_GLYCOLYSIS	72	0.539235	1.744881	0.002041	0.015083	0.010503
REACTOME_GLUCOSE_METABOLISM	91	0.515221	1.735192	0.00211	0.015083	0.010503

TCGA, The Cancer Genome Atlas, LUAD, Lung adenocarcinoma GSEA, Gene Set Enrichment Analysis

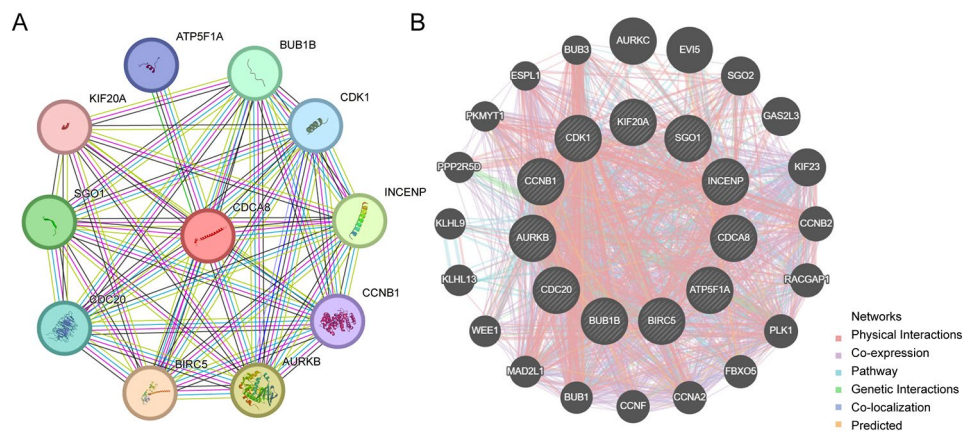


Fig. 6 Construction of PPI network. (A) CDCA8 PPI Network. (B) Functionally similar gene interaction network of 11 genes predicted by GeneMANIA website

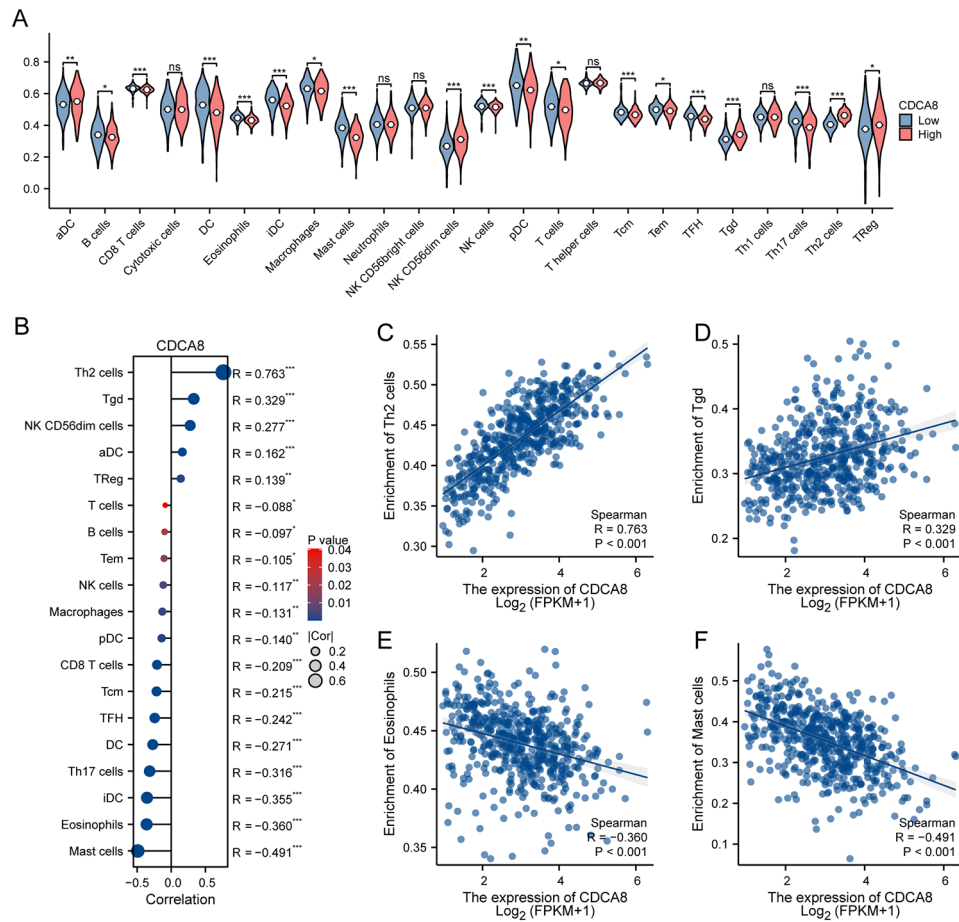


Fig. 7 Differential analysis of ssGSEA immune characteristics between CDCA8 differential expression groups. **(A)** There are 24 immune cells in the TCGA - LUAD group that are significantly different in the grouping comparison plot of the CDCA8 differential expression groups. **(B)** Lollipop plot of correlation between CDCA8 and 19 significantly different immune cells. **(C-F)** Scatter plot of the association of CDCA8. TCGA-LUAD dataset: $n = 539$ (LUAD) and $n = 59$ (Normal)

Table 4 Result of Cox Analysis

Characteristics	Total(N)	HR(95% CI) Univariate analysis	p value Univariate analysis	HR(95% CI) Multivariate analysis	p value Multivariate analysis
Pathologic stage	522				
Stage I	292	Reference		Reference	
Stage II	123	2.341 (1.638–3.346)	<0.001	2.237 (1.562–3.203)	<0.001
Stage III	81	3.576 (2.459–5.200)	<0.001	3.343 (2.291–4.879)	<0.001
Stage IV	26	3.819 (2.211–6.599)	<0.001	3.592 (2.070–6.235)	<0.001
Gender	530				
female	283	Reference			
Male	247	1.087 (0.816–1.448)	0.569		
Age	520				
<= 65	257	Reference			
> 65	263	1.216 (0.910–1.625)	0.186		
CDCA8	530	1.229 (1.090–1.386)	<0.001	1.153 (1.018–1.306)	0.025

HR, Hazard thewire, general HR > 1 shows variable is the risk factor, HR < 1 is the protection factors. Univariate p value < 0.001 was included in the analysis

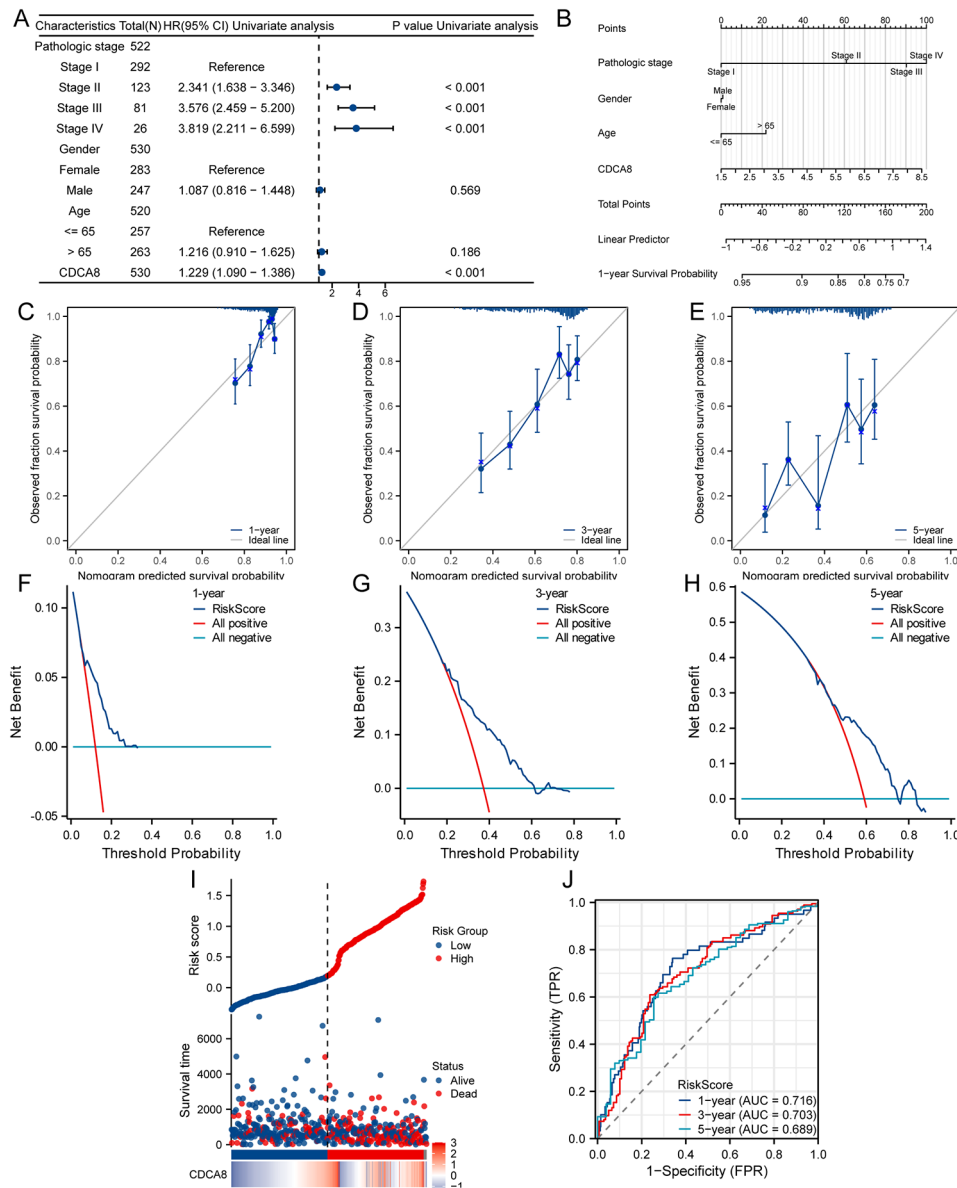


Fig. 8 TCGA - LUAD dataset multivariable Cox regression model building. **(A)** TCGA - LUAD dataset forest picture of single factor Cox regression model. **(B)** nomogram of multi-factors Cox regression model. **(C-E)** Calibration curves at 1-,3-,5-year for multivariate Cox regression model nomogram analysis. **(F-H)** DCA plots at 1-,3-,5-year of the multivariate Cox regression model. **(I)** Cox prognosis model of risk factors. **(J)** The ROC results of Cox prognostic modeling with OS survival outcomes in LUAD patients. TCGA-LUAD dataset: $n = 539$ (LUAD) and $n = 59$ (Normal)

shown (Fig. 10A-T). We found that among the 20 drugs with significant differences, the CDCA8 low-expression group generally showed higher drug sensitivity than the CDCA8 high-expression group (Fig. 10A-T). Based on these results, it is speculated that patients with low CDCA8 expression may have a higher sensitivity to these drugs, which further emphasizes the importance of individualized treatment for patients with tumors.

Immunohistochemical analysis of CDCA8 and LUAD

The immunohistochemical analysis results showed that the expression level of CDCA8 was higher in lung

adenocarcinoma (LUAD) tissue (Fig. 11A) compared to normal lung glandular tissue (Fig. 11B).

Difference analysis of CDCA8 resistance and susceptibility groups

We respectively in GSE108214 and GSE109821 data sets, using grouping comparison chart, shows CDCA8 gene in drug-resistant and Sensitive group of expression (Fig. 12A-B). In the GSE109821 dataset, the CDCA8 levels were higher in the resistant group than in the sensitive group, but the difference was not statistically significant.

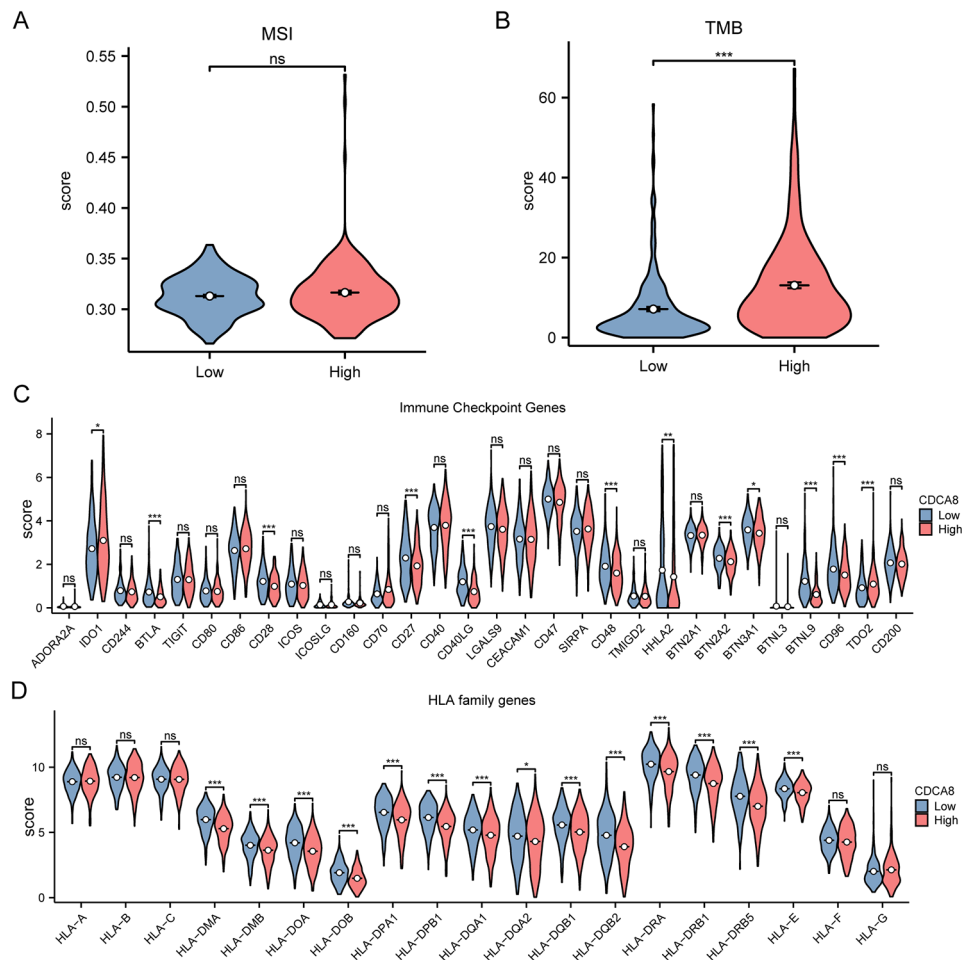


Fig. 9 Differential analysis of CDCA8 gene with MSI, TMB, Immune Checkpoint and HLA family genes in high and low risk groups. (A) TMB score. (B) immune checkpoint gene. (C) the family of the HLA gene. (D) the grouping comparison chart

Discussion

LUAD is the most common histological subtype of NSCLC, and the overall survival rate of patients with intermediate and advanced stages of the disease is less than 15% because of the lack of effective early diagnostic methods. Therefore, screening for additional biomarkers related to tumor staging and prognosis is extremely important for early diagnosis, prognostic evaluation, and treatment. Uncontrolled cell proliferation caused by abnormalities in cell cycle-related proteins endows tumor cells with an enhanced ability to invade, metastasize, and become drug resistant. Therefore, dysregulation of cell cycle progression is also considered a common feature of cancer [34, 35]. CDCA8, a cell cycle regulatory protein located on human chromosome 1p34.2, is primarily expressed in embryonic stem cells [36]. An increasing number of studies have confirmed that CDCA8 overexpression is linked to the occurrence of various malignant tumors, such as bladder cancer [37], rectal cancer [38], and breast cancer [39]. However, its clinical relevance

as a biomarker for LUAD has not yet been thoroughly investigated.

We performed bioinformatics analysis of RNA-seq data of patient tissue samples obtained from the TCGA database to assess the prognostic value of CDCA8 in LUAD. We found higher levels of CDCA8 in LUAD tissues than in the controls. Subsequently, we plotted prognostic survival curves and predicted that patients with higher levels of CDCA8 had a poorer prognosis. This is consistent with a previous report of CDCA8 expression in hepatocellular carcinoma [7]. Therefore, we hypothesized that CDCA8 could serve as a biomarker of LUAD.

In this study, ssGSEA analysis revealed a significant relationship between CDCA8 expression levels and the infiltration abundance of 24 immune cells in LUAD. The results showed that a total of 19 immune cells showed significant differences in the infiltration levels between high and low CDCA8 expression groups (p value < 0.05), including CD8 T cells, dendritic cells, eosinophils and iDCs (p value < 0.001). These results suggest that the high expression of CDCA8 may affect the immune

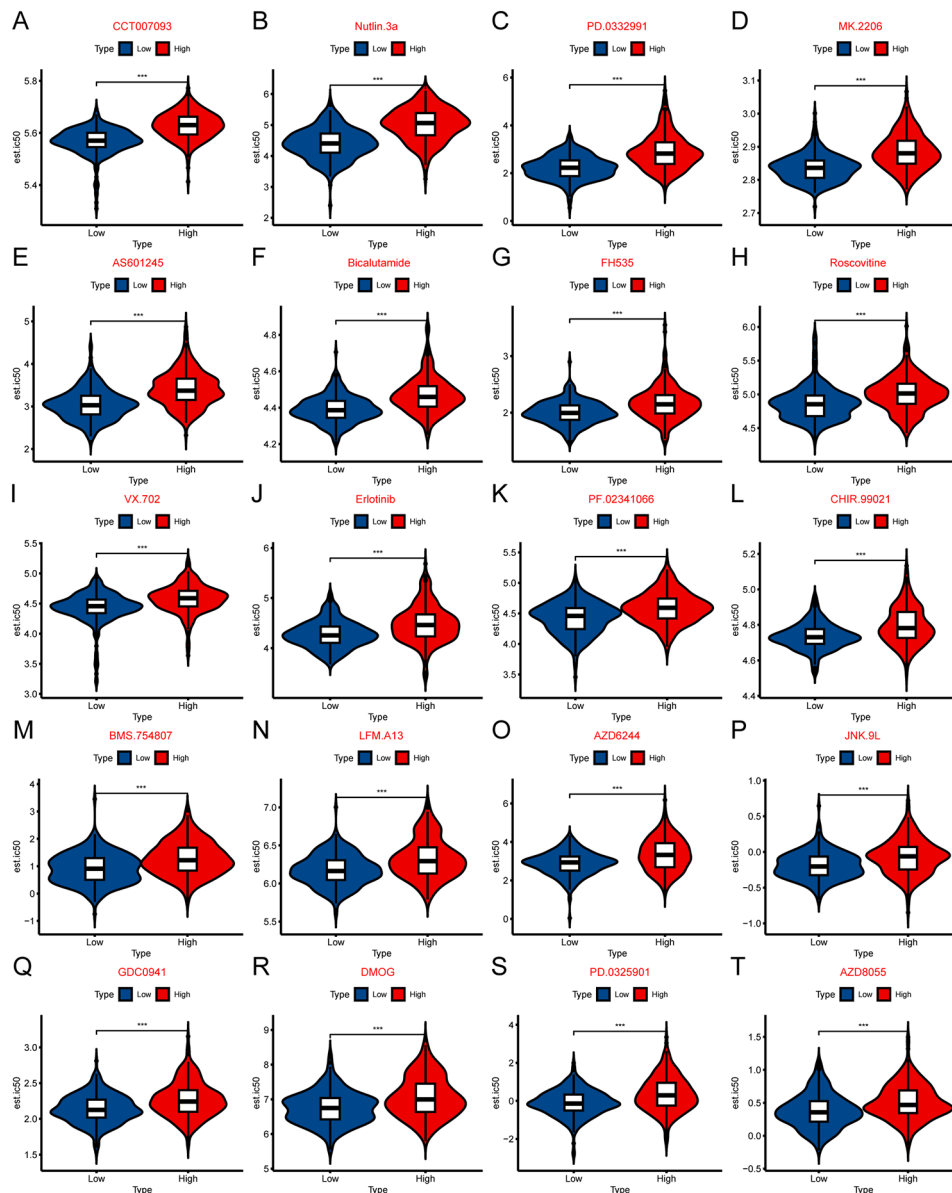


Fig. 10 genes CDCA8 drug sensitivity analysis. **(A)** The results of the sensitivity analysis for the drug CCT007093. **(B)** Nutlin.3a. **(C)** PD.0332991. **(D)** MK.2206. **(E)** AS601245. **(F)** Bicalutamide. **(G)** FH535. **(H)** Roscovitine. **(I)** VX. 702. **(J)** Erlotinib. **(K)** PF. 02341066. **(L)** CHIR. 99,021. **(M)** BMS. 754,807. **(N)** LFM. A13. **(O)** AZD6244. **(P)** JNK. 9L. **(Q)** GDC0941. **(R)** DMOG. **(S)** PD. 0325901. **(T)** AZD8055. TCGA-LUAD dataset: $n = 539$ (LUAD) and $n = 59$ (Normal)

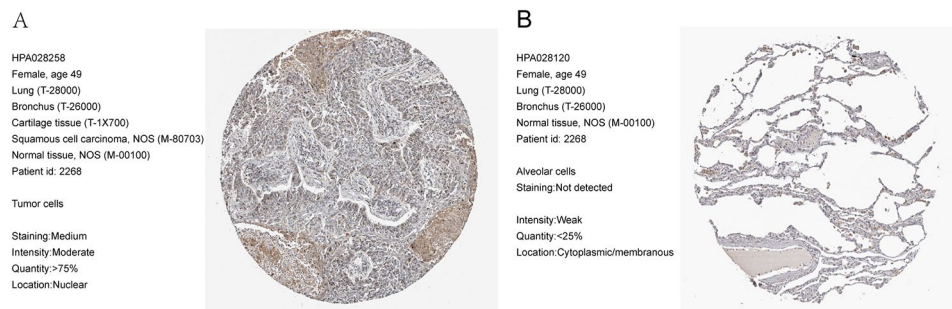


Fig. 11 Immunohistochemical analysis of CDCA8. **(A)** CDCA8 genes in Normal tissue. **(B)** CDCA8 genes in LUAD immunohistochemical analysis. Data are obtained from the HPA database

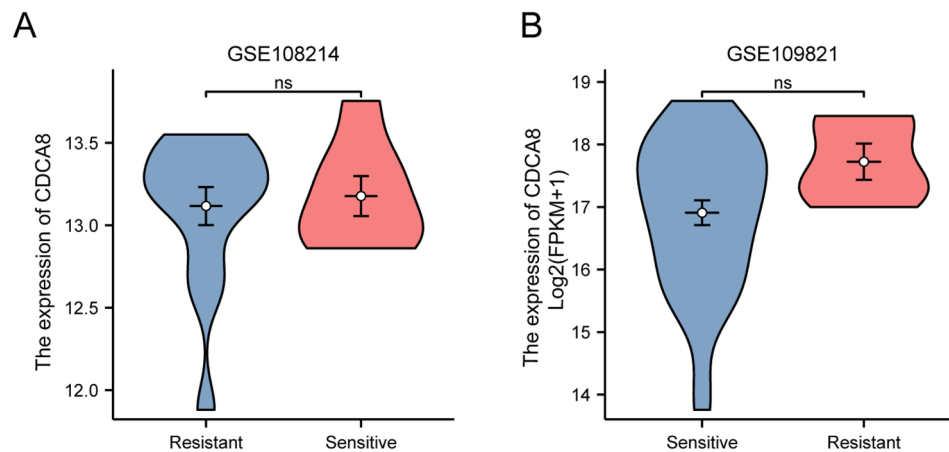


Fig. 12 CDCA8 expression differences in sensitive resistance groups. **(A)** Comparison of differential expression groups of CDCA8 in GSE108214 dataset. **(B)** CDCA8 differentially expressed in GSE109821 data set grouping comparison chart, but not statistically significant ($P \geq 0.05$). GSE108214 dataset: $n = 15$ (resistant samples) and $n = 7$ (sensitive samples), GSE109821 dataset: $n = 5$ (resistant samples) and $n = 37$ (sensitive samples)

microenvironment and tumor progression in LUAD by modulating the infiltration of immune cells. In particular, the significant changes in the anti-tumor immune responses of CD8 T cells and DCs suggested that CDCA8 might play an important role in regulating the functions of these key immune cells. Further correlation analysis showed that CDCA8 was positively correlated with Th2 cells and Tgd cells and negatively correlated with mast cells and eosinophils [40–42]. These findings imply that CDCA8 may affect the tumor immune microenvironment through different mechanisms, thereby regulating tumor growth and patient prognosis. Taken together, the present study reveals the critical role of CDCA8 in immune cell infiltration in LUAD, providing new evidence for its role as a potential immunotherapeutic target. Future experimental studies will further validate these results and explore the specific mechanisms by which CDCA8 regulates immune cell function.

Dysregulation of cell cycle-associated proteins is the most prominent feature of malignant tumor proliferation [34], and cell cycle-associated proteins can regulate drug resistance in tumor cells in a variety of ways, e.g., regulating cell cycle progression, increasing DNA damage repair, and regulating stem cell self-regeneration [43–45]. Previous studies have found that CDCA8 overexpression promotes cancer progression and enhanced drug resistance, and that drug resistance in cancer cells can be reversed and apoptosis induced by targeting CDCA8 inhibition [12, 46, 47]. In this study, the GDSC database was used to predict the sensitivity of CDCA8 to anticancer drugs and 20 drugs with significant differences were selected. The results showed that CDCA8 may be involved in cellular drug resistance through multiple mechanisms. Such as cell cycle-associated proteins: PD.0332991 (CDK4/CDK6 inhibitor), Roscovitine (CDKs inhibitor), LFM.A13 (PLK3 inhibitor), Nutlin.3a (inhibits MDM2-p53 interactions).

PI3K-mTOR signaling pathway: CCT007093 (inhibits mTORC1 pathway), AZD8055 (ATP-competitive mTOR inhibitor), MK.2206 (AKT inhibitor), GDC0941 (PI3K α/δ inhibitor). MAPK-MEK signaling pathway: VX.702 (MAPK inhibitor), AZD6244 (non-ATP competitive MEK1/2 inhibitor), PD.0325901 (selective and non-ATP competitive MEK inhibitor). These results suggest that high levels of CDCA8 lead to insensitivity to cell cycle-related inhibitors and resistance to inhibitors of cell proliferation-related pathways. Clinical selection of chemotherapeutic agents may be beneficial by evaluating CDCA8 expression levels, and development of combination therapy with CDCA8-targeted inhibitors and chemotherapeutic agents may be effective as a therapeutic option for the treatment of cancer.

To further understand the link between CDCA8 and drug resistance, we found that patients with low levels of CDCA8 had higher sensitivity to drugs compared to patients with high CDCA8 levels by evaluating their resistance to chemotherapeutic drugs. This suggests that combining with a targeted inhibitor against CDCA8 could increase the sensitivity of patients to the drug and improve its efficacy.

Although our study provides new insights into the correlation between CDCA8 expression and LUAD, it has certain limitations. First, the evaluated dataset was small, and the analysis results may have been biased by the interference of some samples. Therefore, the sample size should be increased to improve the reliability of the results. Second, some samples were analyzed without considering the actual clinical situation. Finally, to verify the authenticity of these results, more in-depth experiments are required to validate the biological functions of CDCA8 in vitro and in vivo.

Overall, our study revealed for the first time the prognostic value of CDCA8 in LUAD. Our findings suggest

that CDCA8 can potentially serve as a novel biomarker and target for improving drug sensitivity. Although this study revealed the potential role of CDCA8 in LUAD through multiple independent datasets and comprehensive bioinformatics analysis, the lack of experimental validation is a limitation. Future studies need to validate the specific role of CDCA8 in immune cell infiltration and tumor progression through in vivo and in vitro experiments to further confirm the preliminary findings of this study and explore its feasibility as a therapeutic target.

Abbreviations

LUAD	Lung adenocarcinoma
CDCA8	Cell division cycle-associated 8
TCGA	The Cancer Genome Atlas
ROC	Receiver operating characteristic
CDCA	Cell division cycle associated protein
GEO	Gene Expression Omnibus
SsGSEA	Single-sample genome set enrichment analysis
FPKM	Fragments Per Kilobaseper Million
OS	Overall Survival
KM	Kaplan–Meier
BH	Benjamini–Hochberg
KEGG	Kyoto Encyclopedia of Genes and Genomes
GSEA	Gene Set Enrichment Analysis
PPI	Protein-protein interaction
ICG	Immune checkpoint gene
MSI	Microsatellite Instability
TMB	Tumor Mutation Burden
HPA	Human Protein Atlas
PLK1	Polo-like Kinase 1
FOXM1	Forkhead box M1
CCNB1	Cyclin B1
USP24	Ubiquitin-specific peptidase 24
CDK1	Cyclin dependent kinase 1

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-024-02019-x>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7

Acknowledgements

This research was supported by the China Scholarship Council (NO. 202308460036) and Science Experiment Center of Hainan Medical University.

Author contributions

Qiang Liu and Weimin Chen designed this research. Huiquan Gu and Xinzhen Gao carried out the concrete analysis and wrote the first draw manuscript. Wenlong Han and Fangyu Wang edited the figures and tables. Hanqiang Zhang and Longyu Yao edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China with grant number (NO.82060851); the Hainan Medical University

Innovative Experimental Projects (NO.HYYS2021A35); the Hainan Medical College Student Innovation and Entrepreneurship Training Program Project (NO.202311810019). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors declare that they have no competing interests.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Pharmacology, Hainan Medical University, No. 3, Longhua Road, Haikou, Hainan 571199, China

²Department of Biology, Hainan Medical University, Haikou, Hainan, China

³Department of First Clinical College, Hainan Medical University, Haikou, Hainan, China

⁴Key Laboratory of Tropical Translational Medicine of Ministry of Education, School of Basic Medicine and Life Sciences, Hainan Medical University, No. 3, Longhua Road, Haikou, Hainan 571199, China

Received: 17 November 2023 / Accepted: 24 September 2024

Published online: 08 November 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. *Cancer J Clin.* 2021;71(3):209–49.
2. Musika W, Kamsa-Ard S, Jirapornkul C, Santong C, Phunmanee A. Lung Cancer Survival with current therapies and new targeted treatments: a Comprehensive Update from the Srinagarind Hospital-Based Cancer Registry from (2013 to 2017). *West Asia Organization for Cancer Prevention (WAOCP), APOCP's West Asia Chap.* 2021(8).
3. Arbour KC, Riely GJ. Systemic therapy for locally Advanced and Metastatic Non-small Cell Lung Cancer: a review. *JAMA.* 2019;322(8):764–74.
4. Rodriguez-Canales J, Parra-Cuentas E, Wistuba II. Diagnosis and molecular classification of Lung Cancer. *Cancer Treat Res.* 2016;170:25–46.
5. Chen C, Chen S, Pang L, Yan H, Luo M, Zhao Q, et al. Analysis of the expression of Cell Division Cycle-Associated genes and its prognostic significance in human lung carcinoma: a review of the literature databases. *Biomed Res Int.* 2020;2020:6412593.
6. Bonner MK, Haase J, Saunders H, Gupta H, Li BI, Kelly AE. The Borealin dimerization domain interacts with Sgo1 to drive Aurora B-mediated spindle assembly. *Mol Biol Cell.* 2020;31(20):2207–18.
7. Wu H, Liu S, Wu D, Zhou H, Sui G, Wu G. Cell division cycle-associated 8 is a prognostic biomarker related to immune invasion in hepatocellular carcinoma. *Cancer Med.* 2023;12(8):10138–55.
8. Gu P, Yang D, Zhu J, Zhang M, He X. Bioinformatics analysis of the clinical relevance of CDCA gene family in prostate cancer. *Med (Baltim).* 2022;101(5):e28788.
9. Dong C, Tian X, He F, Zhang J, Cui X, He Q, et al. Integrative analysis of key candidate genes and signaling pathways in ovarian cancer by bioinformatics. *J Ovarian Res.* 2021;14(1):92.
10. Ci C, Tang B, Lyu D, Liu W, Qiang D, Ji X, et al. Overexpression of CDCA8 promotes the malignant progression of cutaneous melanoma and leads to poor prognosis. *Int J Mol Med.* 2019;43(1):404–12.
11. Li W, Qin Y, Chen X, Wang X. Cell division cycle associated 8 promotes the growth and inhibits the apoptosis of endometrial cancer cells by regulating cell cycle and P53/Rb signaling pathway. *Am J Transl Res.* 2023;15(6):3864–81.

12. Qi G, Zhang C, Ma H, Li Y, Peng J, Chen J, et al. CDCA8, targeted by MYBL2, promotes malignant progression and olaparib insensitivity in ovarian cancer. *Am J Cancer Res.* 2021;11(2):389–415.
13. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGA-biolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016;44(8):e71.
14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
15. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE.* 2008;3(2):e1651.
16. Sarin N, Engel F, Rothweiler F, Cinatl J, Michaelis M, Frötschl R et al. Key players of Cisplatin Resistance: towards a systems Pharmacology Approach. *Int J Mol Sci.* 2018;19(3).
17. Davis S, Meltzer PS. GEOquery: a bridge between the Gene expression Omnibus (GEO) and BioConductor. *Bioinformatics.* 2007;23(14):1846–7.
18. Gene Ontology Consortium. Going forward. *Nucleic Acids Res.* 2015;43(Database issue):D1049–56.
19. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
20. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
21. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics.* 2013;29(14):1830–1.
22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102(43):15545–50.
23. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417–25.
24. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–13.
25. Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA update 2018. *Nucleic Acids Res.* 2018;46(W1):W60–4.
26. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 2020;48(D1):D127–31.
27. Zhou KR, Liu S, Sun WJ, Zheng LL, Zhou H, Yang JH, et al. ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.* 2017;45(D1):D43–50.
28. Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, et al. hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. *Genom Proteom Bioinform.* 2020;18(2):120–8.
29. Xiao B, Liu L, Li A, Xiang C, Wang P, Li H, et al. Identification and Verification of Immune-related gene prognostic signature based on ssGSEA for Osteosarcoma. *Front Oncol.* 2020;10:607622.
30. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity.* 2013;39(4):782–95.
31. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 2013;41(Database issue):D955–61.
32. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS ONE.* 2014;9(9):e107468.
33. Colwill K, Gräslund S. A roadmap to generate renewable protein binders to the human proteome. *Nat Methods.* 2011;8(7):551–8.
34. Liu K, Zheng M, Lu R, Du J, Zhao Q, Li Z, et al. The role of CDC25C in cell cycle regulation and clinical cancer therapy: a systematic review. *Cancer Cell Int.* 2020;20:213.
35. Yadav P, Subbarayalu P, Medina D, Nirzhor S, Timilsina S, Rajamanickam S, et al. M6A RNA methylation regulates histone ubiquitination to Support Cancer Growth and Progression. *Cancer Res.* 2022;82(10):1872–89.
36. Zhang Q, Lin G, Gu Y, Peng J, Nie Z, Huang Y, et al. Borealin is differentially expressed in ES cells and is essential for the early development of embryonic cells. *Mol Biol Rep.* 2009;36(3):603–9.
37. Gao X, Wen X, He H, Zheng L, Yang Y, Yang J, et al. Knockdown of CDCA8 inhibits the proliferation and enhances the apoptosis of bladder cancer cells. *PeerJ.* 2020;8:e9078.
38. Wang Y, Zhao Z, Bao X, Fang Y, Ni P, Chen Q, et al. Borealin/Dasra B is overexpressed in colorectal cancers and contributes to proliferation of cancer cells. *Med Oncol.* 2014;31(11):248.
39. Jiao DC, Lu ZD, Qiao JH, Yan M, Cui SD, Liu ZZ. Expression of CDCA8 correlates closely with FOXM1 in breast cancer: public microarray data analysis and immunohistochemical study. *Neoplasma.* 2015;62(3):464–9.
40. Park SY, Seo D, Jeon EH, Park JY, Jang BC, Kim JI et al. RPL27 contributes to colorectal cancer proliferation and stemness via PLK1 signaling. *Int J Oncol.* 2023;63(2).
41. Zeng L, Liang L, Fang X, Xiang S, Dai C, Zheng T, et al. Glycolysis induces Th2 cell infiltration and significantly affects prognosis and immunotherapy response to lung adenocarcinoma. *Funct Integr Genomics.* 2023;23(3):221.
42. Fan Z, Wu S, Sang H, Li Q, Cheng S, Zhu H. Identification of GPD1L as a potential prognosis Biomarker and Associated with Immune infiltrates in Lung Adenocarcinoma. *Mediators Inflamm.* 2023;2023:9162249.
43. Ling VY, Straube J, Godfrey W, Haldar R, Janardhanan Y, Cooper L, et al. Targeting cell cycle and apoptosis to overcome chemotherapy resistance in acute myeloid leukemia. *Leukemia.* 2023;37(1):143–53.
44. Gupta N, Huang TT, Horibata S, Lee JM. Cell cycle checkpoints and beyond: exploiting the ATR/CHK1/WEE1 pathway for the treatment of PARP inhibitor-resistant cancer. *Pharmacol Res.* 2022;178:106162.
45. Liu M, Zhang H, Li Y, Wang R, Li Y, Zhang H, et al. HOTAIR, a long noncoding RNA, is a marker of abnormal cell cycle regulation in lung cancer. *Cancer Sci.* 2018;109(9):2717–33.
46. Wen X, Liu S, Cui M. Effect of BRCA1 on the concurrent Chemoradiotherapy resistance of cervical squamous cell Carcinoma based on Transcriptome Sequencing Analysis. *Biomed Res Int.* 2020;2020:3598417.
47. Yu D, Shi L, Bu Y, Li W. Cell Division Cycle Associated 8 is a Key Regulator of tamoxifen resistance in breast Cancer. *J Breast cancer.* 2019;22(2):237–47.

Publisher's note

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