Observational Study

Medicine

OPEN

Bioinformatic analysis the expression and clinical significance of CDRT15 in cholangiocarcinoma using TCGA database

Tianyang Yu, MM[a], Tiezhao Zhang, MM[a], Luwen Zhao, MM[b], Kefan Li, MM[a], Jian Li, MD[a], Aijun Yu, MM[a,*]

Abstract

Cholangiocarcinoma (CCA) is a common and lethal malignant tumor originating from bile duct epithelial cells. Various tumor biomarkers have been used for its clinical screening, such as carbohydrate antigen 19-9 and carcinoembryonic antigen. This study aimed to demonstrate the value of associated genes—CMT1A duplicated region transcript 15 (CDRT15) for prognosis of CCA by integrated bioinformatics analysis. We obtained CDRT15 expression data and clinical information on patients with CCA from The Cancer Genome Atlas database. Then, we processed the data by differentially expressed gene analysis, gene set enrichment analysis, statistical analysis, etc. Gene Ontology enrichment analysis was aimed to explore the function of gene-related proteins. Single-sample gene set enrichment analysis was used to analyze the correlation between CDRT15 and immune cells. Finally, we constructed the nomogram to predict the prognosis of patients with CCA. The analysis of data in The Cancer Genome Atlas database revealed that CDRT15 was overexpressed in CCA tissues. We performed the interrelation analysis of immune infiltration, showing that CDRT15 are mainly associated with the immune/inflammatory response. ROC curve showed that CDRT15 can be a diagnostic marker of CCA. Subsequently, the prognostic analysis showed that the high expression of CDRT15 was correlated with the poor OS, and patients with high CDRT15 expression may have a poor prognosis. CDRT15 is more highly expressed in CCA, thus we identified that CDRT15 could be an efficient biomarker for patients. CDRT15 expression was negatively correlated with prognosis of CCA. CDRT15 may be involved in the immune infiltration process of CCA.

Abbreviations: BPs = biological processes, CA19-9 = carbohydrate antigen 19-9, CCA = cholangiocarcinoma, CCs = cellular components, CDRT15 = CMT1A duplicated region transcript 15, DC = dendritic cells, DEG = differentially expressed gene, GO = Gene Ontology, GSEA = gene set enrichment analysis, NK = natural killer, OS = overall survival, ROC = receiver operating characteristic, ssGSEA = single-sample GSEA, TCGA = The Cancer Genome Atlas database, TPM = transcripts per million reads.

Keywords: bioinformatics analysis, CDRT15, cholangiocarcinoma, TCGA

1. Introduction

Cholangiocarcinoma (CCA) is a highly malignant tumor, originating from bile duct epithelial cells. According to statistics, CCA accounts for about 3 percent of all malignant gastrointestinal tumors collectively.[1] In recent decades, the morbidity and mortality of CCA have gradually increased worldwide.[2] At present, the main treatment methods of CCA are surgical resection of tumor lesions, liver transplantation and adjuvant therapy (radiotherapy, chemotherapy).[3–6] Among them, surgical resection is the only possible cure for CCA.[7] However, because the onset of the disease is hidden and early diagnosis is difficult, most patients are only found in the late stage or metastatic stage, losing the opportunity of operation.[8] Because CCA seldom shows any specific symptoms, it is generally considered to be difficult to diagnose during the early stage. Thus, the patients are under a poor prognosis in early stage. Although various biomarkers have been widely studied,[9] such as carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen,[10] their reliability is controversial and the diagnostic value remains limited. Therefore, there is an urgent need to find effective predictive indicators of CCA to provide novel indicators for treatment, diagnosis and prognostic evaluation.

CMT1A duplicated region transcript 15 (CDRT15) is a protein-coding gene, which was first identified by Inoue et

This work was supported in part by Natural Science Foundation of Hebei Province, No. H2021406047.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

There were no cells, tissue, or animal studies. No ethical requirements are involved.

[a] The First Department of General Surgery, Affiliated Hospital of Chengde Medical University, Chengde, PR China, [b] The First Department of Gynecology, Affiliated Hospital of Chengde Medical University, Chengde, PR China.

**Correspondence: Aijun Yu, The First Department of General Surgery, Affiliated Hospital of Chengde Medical University, Chengde, Hebei Province 067000, PR China (e-mail: ccw1979@126.com).*

How to cite this article: Yu T, Zhang T, Zhao L, Li K, Li J, Yu A. Bioinformatic analysis the expression and clinical significance of CDRT15 in cholangiocarcinoma using TCGA database. Medicine 2023;102:31(e34602).

Received: 24 December 2022 / Received in final form: 13 April 2023 / Accepted: 14 July 2023

http://dx.doi.org/10.1097/MD.0000000000034602

al[11] in 2001 in the study of the 1.4-Mb CMT1A Duplication/HNPP Deletion Genomic Region, and was hardly studies for the next 2 decades. Based on the available data, the genomic locations for CDRT15 gene are at chromosome 17:14235673-14236862(GRCh38/hg38) with 1190 bases. The recommended name for the protein encoded by CDRT15 gene is CDRT15 protein. CDRT15 protein is composed of 188 amino acids and has a molecular mass of 20651Da. It mainly expresses in fetal heart, kidney, liver, lung and spleen, and its predicted location is plasma membrane.[12]

To explore the correlation between CDRT15 and CCA, and clarify the prognostic role of CDRT15 in CCA patients, we analyzed the expression level of CDRT15 in CCA and the correlation between CDRT15 expression and clinicopathological indicators based on The Cancer Genome Atlas (TCGA). We analyzed the relationship between CDRT15 and prognostic values, constructed a multivariate Cox regression model and developed a nomogram to predict patient prognosis. In addition, biological pathways related to CDRT15 and CCA were detected by gene set enrichment analysis (GSEA) and Gene Ontology (GO).

2. Methods

2.1. Patients and samples

Gene expression data and clinical information of CCA patients were collected from TCGA Database (https://cancergenome.nih.gov/). HTSeq-counts data of TCGA-CHOL project was obtained from the Genomic Data Commons (https://gdc.cancer.gov/) using the R package "TCGA biolinks."[13] Get rid of the samples that do not have clinical information, the data of 36 patients with CCA were included in the study. The downloaded data format was level3 HTSeq-Fregments Per Kilobase per Million. Then, the data were transformed into TPM (transcripts per million reads) for the following analyses. Our study was in accordance with the publication guidelines provided by TCGA. This study does not contain any studies with human participants or animals and ethical approval was not required. The primary endpoint of this study was 5-year overall survival rate. Candidate variables for model inclusion were age, gender, weight, height, histological type, CA19-9 level, Child–Pugh grade, fibrosis ishak score, vascular invasion, perineural invasion, TP53 status, TNM stage, pathologic stage, histologic grade, residual tumor. The gene indicator examined in the analysis was CDRT15.

2.2. RNA-seq data preprocessing and differentially expressed gene analysis

RNAseq data in Fregments Per Kilobase per Million format were converted into TPM format for expression comparison between samples. The expression of CDRT15 in CHOL of TCGA was compared between adjacent samples and CCA samples.

All the raw RNA-seq data were divided into high and low expression groups according to the expression of CDRT15 in tumor samples (the median is cutoff point). The differentially expressed genes (DEGs) in CCA tissues and normal tissues were screened by DESeq2 package (version 3.8).[14] |Log(fold change)| (|logFC|) >2 and adjusted P value < .01 (P < .01) were set as threshold for the DEG. The results of DEG analysis were visualized by Volcano plot.

2.3. GO enrichment analysis

Metascape (http://metascape.org) is an open-access gene list analysis tool for analyzing the network of gene-related proteins using powerful computational analysis pipelines.[15] In this study, Metascape was used as a tool to analyze the GO enrichment of the entire list of DEG between CDRT15 high expression group and low expression group, including biological processes (BPs), molecular functions, and cellular components (CCs). BP is defined as biological objectives to which the gene or its product contributes. Molecular function refers to the biochemical activity of a gene product. CC is the place where a gene product is active in the cell.[16] It was considered that there was a significant statistical difference only if the following conditions were satisfied: P < .01, minimum count of 3, and enrichment factor > 1.5. R package clusterprofiler is an alternative to GO enrichment analysis and Kyoto Encyclopedia of Genes and Genomes enrichment analysis.[17]

2.4. Gene set enrichment analysis

GSEA method enriches the pathways related to gene expression to determine whether there was a significant difference between high and low expression state of CDRT15. The GSEA and visualization were carried out by R package clusterprofiler (3.8.0),[17,18] and the gene set permutations were performed 1000 times each analysis. The expression level of CDRT15 was used as a phenotype. The statistical significance index and threshold were set to adjusted P < .05, false discovery rate (FDR) q value < 0.25 and normalized enrichment score (|NES|) > 1.

2.5. Analysis of immune cell characteristics by single-sample GSEA

The immune infiltration quantitative analysis of CCA was done by single-sample GSEA (ssGSEA) method using R package GSVA.[19] The ssGSEA method can be used to enrich analysis a set of reference genes with common biological function, chromosome location and physiological regulation in a single sample. We quantified the relative tumor infiltration levels of 24 immune cell types,[20] the following immune cells were obtained: natural killer (NK) cells, effector memory T cells (Tem), macrophages, type-2 T helper cells (Th2), mast cells, eosinophils, neutrophils, T gamma delta cells (Tγδ), type-1 T helper cells (Th1), immature dendritic cells (DCs), CD56dim natural killer cells (CD56-NK), T central memory cells (Tcm), plasmacytoid DCs, B cells, DCs, CD56bright natural killer cells (CD56+NK), regulatory T cells (Tregs), T helper cells, follicular helper T cells (Tfh), CD8 + T cells, type-17 T helper cells (Th17), activated DCs, T cells and cytotoxic cells.

2.6. Statistical analysis

All statistical analyses were produced using R (Version 3.6.2). Wilcoxon rank sum test and Wilcoxon signed rank test were used to analyze the expression of CDRT15 in tumor and control samples. Kruskal–Wallis test, Wilcoxon rank sum test, Wilcoxon signed-rank test and Spearman correlation were used to evaluate the relationships between the grade of clinicopathological factors and CDRT15 expression. Direct correlation between the grade of clinicopathological factors and the grouping of CDRT15 using normal and corrected Pearson's χ^2 test, Fisher exact test, Univariate logistic regression and Multivariate Cox analysis were used to compare the effect of CDRT15 expression on survival rate. Wilcoxon rank sum test was used to analyze the relationship between high and low expression groups of CDRT15 or different clinicopathological factors and immune cell infiltration. The Kaplan–Meier method was used to construct the survival curve, and the differences between the survival curves were tested by log-rank test. All hypothesis tests were 2-tailed, and P < .05 is considered significant in all tests. We analyzed the expression of CDRT15 in patients with CCA and healthy people, and used pROC package[21] to draw receiver operating characteristic (ROC) curve to evaluate the diagnostic ability of CDRT15 in CCA.

According to explore the expression of CDRT15 and clinicopathological factors, multivariate Cox regression model was used to analyze the independent prognostic factors related to survival. R packages rms (https://CRAN.R-project.org/package=rms) was used to construct the nomogram and draw calibration plot to predict the prognosis of patients with CCA. Risk score of patients was calculated according to the formula of multivariate Cox regression model, and the distributions of risk scores, overall survival (OS) time, OS status and the expression level of CDRT15 were illustrated in the risk factor association map. Judging from the median risk score, patients were divided into low-risk group and high-risk group, and survival curve was drawn. Using risk score and survival status to do ROC analysis, get C-index to quantify, evaluate the prognostic value of tumor prognosis model.

3. Results

3.1. Clinical characteristics

The clinical characteristics of 36 patients including age, gender, race, tumor, node and metastasis stage, tumor location and other indexes were collected, including 20 female patients and 16 male patients, as shown in Table 1.

3.2. Identification of DEGs

According to the expression of CDRT15, TCGA CCA was divided into high expression group and low expression group. The DEGs were analyzed by DESeq2 package. There were 190 DEGs with |logFC|>1 and *P*-adjust (*P*-adj) < .05, 157 DEGs with |logFC|>1.5 and *P*-adj < .05, and 122 DEGs with |logFC|>2 and *P*-adj < .05.

The volcano map is used to show the results of the differential analysis. There were 32 DEGs with |logFC|>2 and *P*-adj < .01, and 19 DEGs with |logFC|< -2 and *P*-adj < .01 (Fig. 1).

3.3. Functional enrichment analysis of CDRT15

We carried out GO enrichment analysis in Metascape to predict the functional enrichment information of CDRT15 interacting genes. There were 106 results that are statistically significant, including 9 BPs and CCs. The results showed that CDRT15 and its related DEGs were mainly located in developmental maturation, sex differentiation, gonad development, regulation of branching involved in ureteric bud morphogenesis, positive regulation of mesonephros development and positive regulation of branching involved in ureteric bud morphogenesis (Table 2, Fig. 2).

Table 1

Clinical characteristic.

Characters	Level	Low expression of CDRT15	High expression of CDRT15	*P*	Test
n		18	18	–	–
T stage (%)	T1	11 (61.1%)	8 (44.4%)	.519	Exact
	T2	4 (22.2%)	8 (44.4%)		
	T3	3 (16.7%)	2 (11.1%)		
N stage (%)	N0	16 (94.1%)	10 (71.4%)	.148	Exact
	N1	1 (5.9%)	4 (28.6%)		
M stage (%)	M0	15 (88.2%)	13 (81.2%)	.656	Exact
	M1	2 (11.8%)	3 (18.8%)		
Pathologic stage (%)	Stage I	11 (61.1%)	8 (44.4%)	.446	Exact
	Stage II	4 (22.2%)	5 (27.8%)		
	Stage III	1 (5.6%)	0 (0.0%)		
	Stage IV	2 (11.1%)	5 (27.8%)		
Histologic grade (%)	G1	1 (5.6%)	0 (0.0%)	.862	Exact
	G2	8 (44.4%)	7 (38.9%)		
	G3	8 (44.4%)	10 (55.6%)		
	G4	1 (5.6%)	1 (5.6%)		
Residual tumor (%)	R0	15 (88.2%)	13 (81.2%)	.656	Exact
	R1	2 (11.8%)	3 (18.8%)		
Gender (%)	Female	13 (72.2%)	7 (38.9%)	.094	–
	Male	5 (27.8%)	11 (61.1%)		
Histological type (%)	Distal	2 (11.1%)	0 (0.0%)	.539	Exact
	Hilar/perihilar	2 (11.1%)	2 (11.1%)		
	Intrahepatic	14 (77.8%)	16 (88.9%)		
Child–Pugh grade (%)	A	12 (92.3%)	7 (87.5%)	1.000	Exact
	B	1 (7.7%)	1 (12.5%)		
Fibrosis ishak score (%)	0	7 (50.0%)	9 (69.2%)	.695	Exact
	1/2	6 (42.9%)	3 (23.1%)		
	3/4	1 (7.1%)	1 (7.7%)		
Vascular invasion (%)	No	16 (88.9%)	13 (81.2%)	.648	Exact
	Yes	2 (11.1%)	3 (18.8%)		
Perineural invasion (%)	No	15 (83.3%)	11 (73.3%)	.674	Exact
	Yes	3 (16.7%)	4 (26.7%)		
TP53 status (%)	Mut	2 (11.8%)	1 (5.6%)	.603	Exact
	WT	15 (88.2%)	17 (94.4%)		
Age (median [IQR])		62.00 [55.50, 71.75]	67.50 [58.50, 71.75]	.601	Nonnorm
Height (median [IQR])		162.00 [159.25, 169.50]	169.00 [165.00, 183.00]	.105	Nonnorm
Weight (median [IQR])		72.50 [67.25, 80.00]	86.50 [63.25, 99.50]	.229	Nonnorm
CA19-9 value (median [IQR])		41.60 [26.00, 67.50]	91.00 [18.15, 247.00]	.487	Nonnorm

CDRT15 = CMT1A duplicated region transcript 15, IQR = interquartile range.

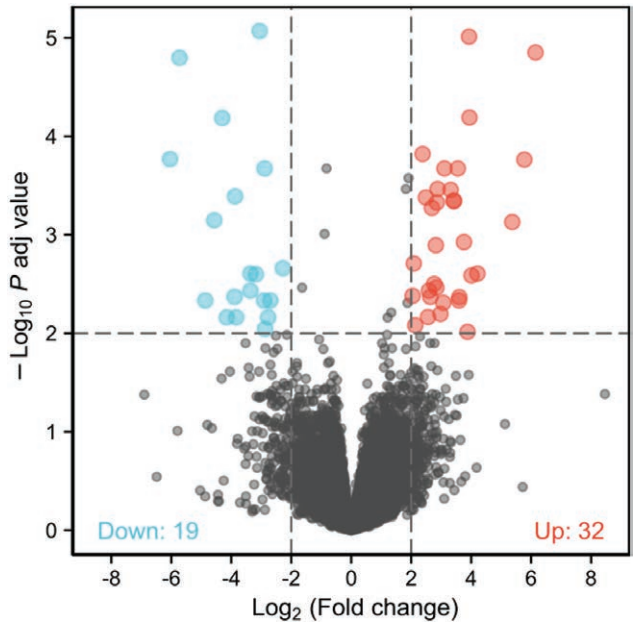

Figure 1. The volcano map of the differential analysis. There are 32 up-regulated genes and 19 down-regulated genes.

Table 2

Functional enrichment analyses of CDRT15.

Ontology	ID	Description	*P* value
BP	GO:0021700	Developmental maturation	<.001
BP	GO:0090190	Positive regulation of branching involved in ureteric bud morphogenesis	<.001
BP	GO:0061213	Positive regulation of mesonephros development	<.001
BP	GO:0007548	Sex differentiation	<.001
BP	GO:0008406	Gonad development	<.001
BP	GO:0090189	Regulation of branching involved in ureteric bud morphogenesis	<.001
BP	GO:0045137	Development of primary sexual characteristics	<.001
BP	GO:0061217	Regulation of mesonephros development	<.001
CC	GO:0098793	Presynapse	.001086
CC	GO:0098978	Glutamatergic synapse	.001652
CC	GO:0044306	Neuron projection terminus	.001807
CC	GO:0030658	Transport vesicle membrane	.002041
CC	GO:0072687	Meiotic spindle	.002079
CC	GO:0005581	Collagen trimer	.002218
CC	GO:0005891	Voltage-gated calcium channel complex	.002494
CC	GO:0005788	Endoplasmic reticulum lumen	.003429
CC	GO:0043195	Terminal bouton	.004057

BP = biological process, CC = cellular component, CDRT15 = CMT1A duplicated region transcript 15.

3.4. Significant genes and pathways obtained by GSEA

GSEA was performed on CDRT15 low-expression and high-expression data sets based on the co-expression gene analysis results of CDRT15, and based on the TCGA CCA expression matrix, through clusterPorfiler package. There are significant differences (FDR < 0.05, adjusted $P < .05$) in the enriched MSigDB set (C2.all.v6.2.symbols). There are 13 data sets with FDR < 0.25 and P-adj < .05.

3.5. The correlation between CDRT15 expression and immune infiltration

Spearman correlation analysis was used to show the correlation between the expression level of CDRT15 and immune cell enrichment level (generated by ssGSEA) in the tumor microenvironment of CCA. The CDRT15 expression was negatively correlated with the Treg, DCs, and Cytotoxic cells (Fig. 3, $P < .05$).

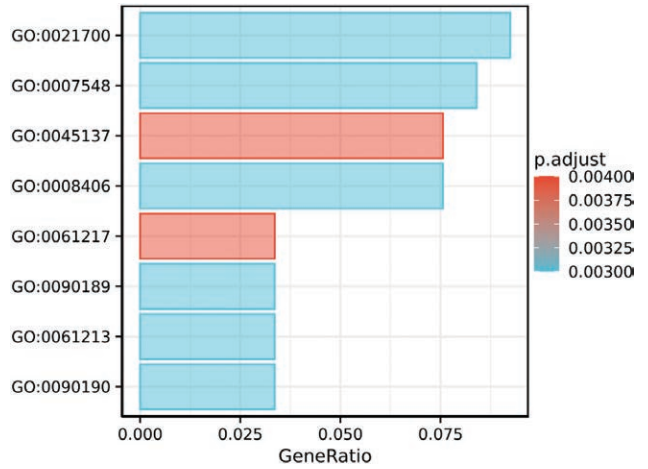

Figure 2. Functional enrichment analysis of CDRT15. CDRT15 = CMT1A duplicated region transcript 15, GO = Gene Ontology.

3.6. Difference of CDRT15 expression between tumor and normal tissue

Wilcoxon rank sum test was used to compare the expression of CDRT15 in 9 adjacent samples and 36 CCA samples in TCGA. Finally, CDRT15 was significantly overexpressed in CCA samples, and the result was statistically significant ($P = .001$) (Fig. 4A).

Wilcoxon signed rank test was used to compare the expression of CDRT15 in 9 CCA samples of TCGA CCA and the corresponding paired adjacent samples. Finally, CDRT15 was found to be significantly overexpressed in samples of CCA ($P = .004$) (Fig. 4B).

We analyzed the efficacy of distinguishing tumor from non-tumor tissue by CDRT15 expression level using ROC curve. The area under the curve is 0.855, which indicates that the expression of CDRT15 has a strong ability to distinguish tumor from normal tissue (Fig. 4C).

3.7. Clinical correlation analysis of CDRT15 expression in tumor patients

Kruskal–Wallis rank sum test was used to analyze the relationship between CDRT15 expression and clinical features. The highly expressed CDRT15 expression is correlated with the N stage ($P = .013$) and pathologic stage ($P = .034$) (Fig. 5).

3.8. Difference of CDRT15 expression between cancer tissues and control samples in pan-cancerous tumors

We download RNAseq data in TPM format of TCGA and GTEx from the UCSCXENA database (https://xenabrowser.net/datapages/), which has been uniformly processed by the Toil process[22] shows that the expression of CDRT15 in normal samples in GTEx and TCGA is compared with that in tumor samples corresponding to TCGA by Wilcoxon rank sum test. In the end, it was concluded that CDRT15 was significantly expressed in CCA ($P < .05$) (Fig. 6).

3.9. Role of CDRT15 in CCA patient survival

Kaplan–Meier survival analysis showed that CCA with CDRT15-high had a worse prognosis than that with CDRT15-low ($P = .013$).

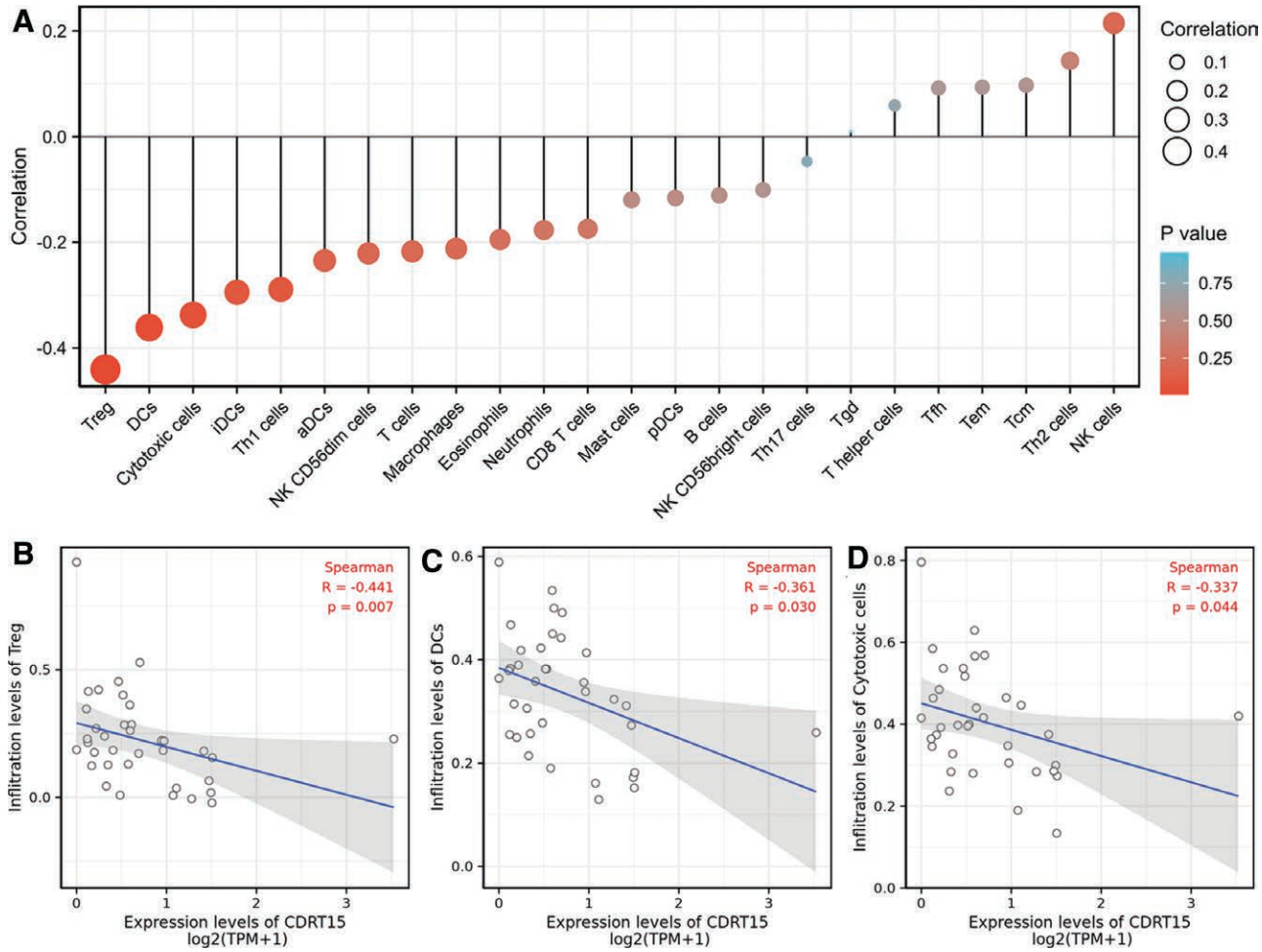

Figure 3. (A)The correlation between CDRT15 expression and immune infiltration. The correlation between the expression levels of CDRT15 and infiltration levels of Treg (B), DCs (C), cytotoxic cells (D). aDCs = activated dendritic cells, CD = cluster of differentiation, CDRT15 = CMT1A duplicated region transcript 15, DCs = dendritic cells, iDCs = immature dendritic cells, NK CD56 bright = CD56bright natural killer cells, NK CD56dim = CD56dim natural killer cells, pDCs = plasmacytoid dendritic cells, Tcm = T central memory cells, Tem = effector memory T cells, Tfh = follicular helper T cells, Tgd = T gamma delta cells, Th1 Cells = type-1 T helper cells, Th17 cells = type-17 T helper cells, Th2 cells = type-2 T helper cells, TPM = transcripts per million reads.

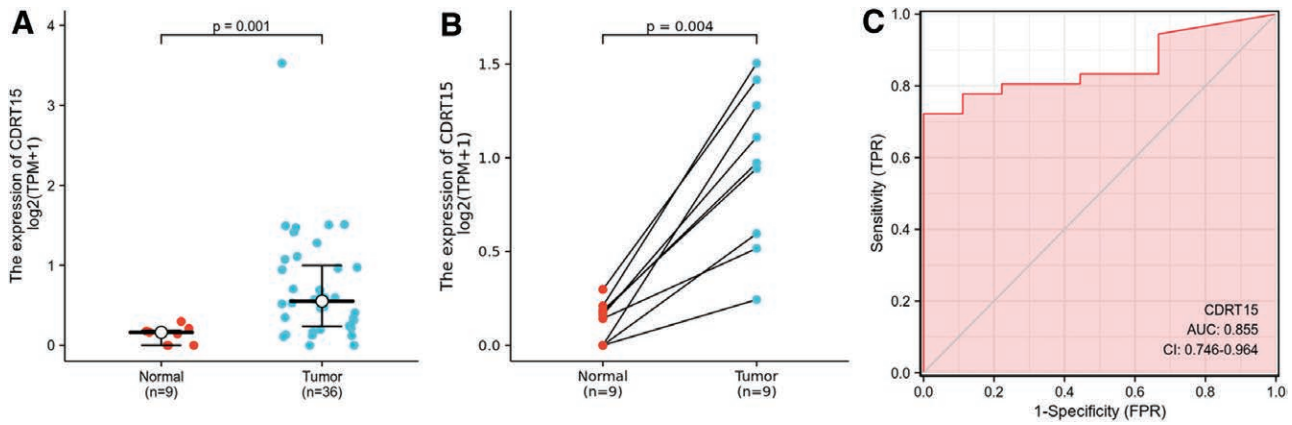

Figure 4. (A) Difference of CDRT15 expression between tumor and normal tissue. (B) Difference of CDRT15 expression between tumor and corresponding paired adjacent sample. (C) ROC curve of the efficacy of distinguishing tumor from non-tumor tissue of CDRT15. AUC = area under curve, CDRT15 = CMT1A duplicated region transcript 15, CI = confidence interval, FPR = false positive rate, TPM = transcripts per million reads, TPR = true positive rate.

The univariate analysis showed that the high expression of CDRT15 was significantly correlated with the poor OS (P = .029). Other clinicopathological variables associated with poor OS include perineural invasion (P = .014). We then use Cox regression model for multivariate analysis. The results show perineural invasion (P = .011) and the high expression of CDRT15 (P = .049) were independently correlated with OS (Fig. 7, Table 3).

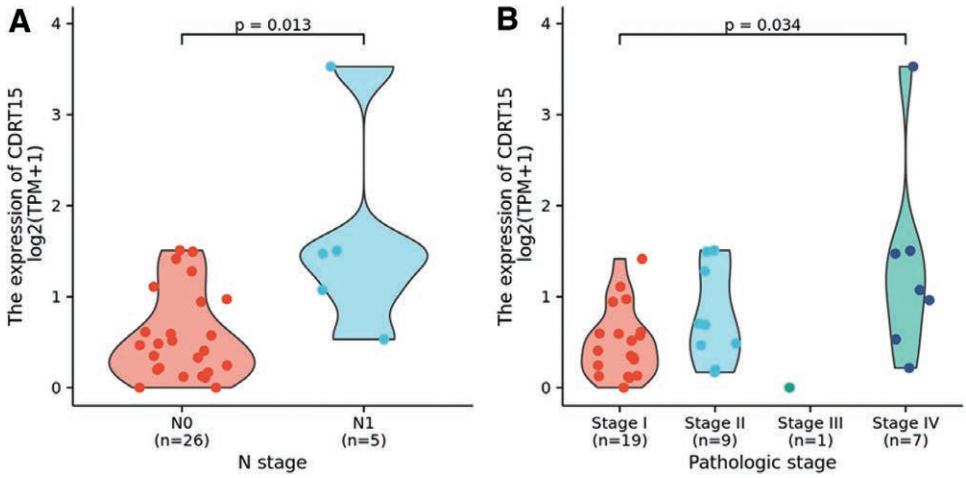

Figure 5. Clinical correlation between CDRT15 expression, the N stage (A) and the pathologic stage (B). CDRT15 = CMT1A duplicated region transcript 15, TPM = transcripts per million reads.

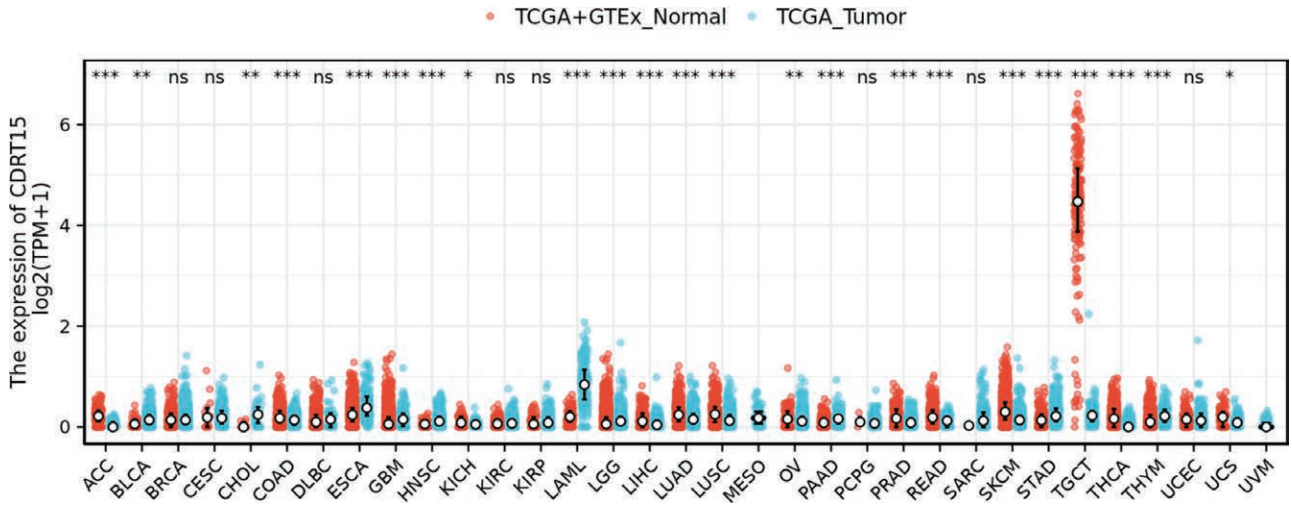

Figure 6. Difference of CDRT15 expression between cancer tissues and pan-cancerous tumors. ACC = adrenocortical carcinoma, BLCA = bladder urothelial carcinoma, BRCA = breast invasive carcinoma, CDRT15 = CMT1A duplicated region transcript 15, CESC = cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL = cholangiocarcinoma, COAD = colon adenocarcinoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, ESCA = esophageal carcinoma, GBM = glioblastoma multiforme, GTEx = Genotype-Tissue Expression, HNSC = head and neck squamous cell carcinoma, KICH = kidney chromophobe, KIRC = kidney renal clear cell carcinoma, KIRP = kidney renal papillary cell carcinoma, LAML = acute myeloid leukemia, LGG = brain lower grade glioma, LIHC = liver hepatocellular carcinoma, LUAD = lung adenocarcinoma, LUSC = lung squamous cell carcinoma, MESO = mesothelioma, OV = ovarian serous cystadenocarcinoma, PAAD = pancreatic adenocarcinoma, PCPG = pheochromocytoma and paraganglioma, PRAD = prostate adenocarcinoma, READ = rectum adenocarcinoma, SARC = sarcoma, SKCM = skin cutaneous melanoma, STAD = stomach adenocarcinoma, TCGA = the Cancer Genome Atlas, TGCT = testicular germ cell tumors, THCA = thyroid carcinoma, THYM = thymoma, TPM = transcripts per million reads, UCEC = uterine corpus endometrial carcinoma, UCS = uterine carcinosarcoma, UVM = uveal melanoma.

3.10. Construction of a nomogram based on the CDRT15

In order to provide clinicians with a quantitative method for predicting the prognosis of CCA patients, we have constructed a nomogram to show the prognostic model, and can also be used as a scoring tool to evaluate the risk probability. Two independent prognostic variables, perineural invasion and CDRT15, were included in the model of the nomogram, and each variable was marked on the corresponding line segment to represent the range of acceptable values of the variable, while the length of the line segment reflected the contribution of this factor to the prognostic outcome events. Each variable corresponds to its single score in different values (the value of this variable corresponds to the first row of Points is the single score). The corresponding single score of all variables is summed up to get the Total Points (the fifth row from the bottom); Total Points corresponds to the survival probability on different time nodes (1, 3, and 5 years) (Fig. 8). Sampling tests were carried out by Bootstrap method, and the C-index of the model was 0.677 (95% CI: 0.595–0.760). These findings suggest that the nomogram is a better model than a single prognostic factor for predicting survival in patients with CCA.

4. Discussion

CCA is a common malignant tumor of the bile tract system. The symptoms of early CCA are insidious. Resection or transplantation is only suitable for patients with early CCA and the risk of postoperative recurrence is high. Drug therapy is also not very effective for advanced CCA or metastatic CCA. Overall, the prognosis of cancer patients is not ideal.[23] To study the

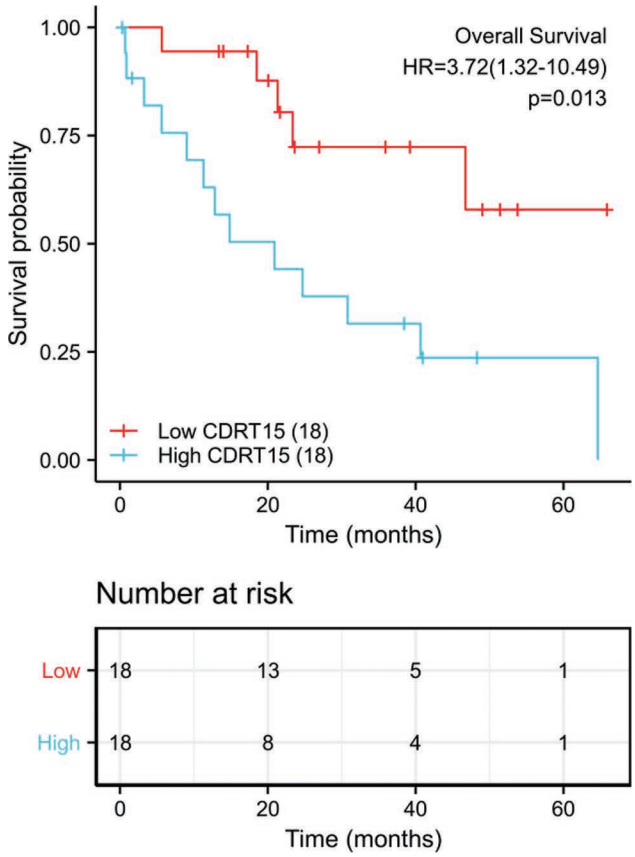

Figure 7. Overall survival of CCA patients. High expression of CDRT15 was significantly associated with reduced overall survival. CCA = cholangiocarcinoma, CDRT15 = CMT1A duplicated region transcript 15, HR = hazard ratio.

Table 3

Cox regression model for multivariate analysis.

Characteristics	Total (N)	Univariate analysis HR (95% CI)	Univariate analysis *P* value	Multivariate analysis HR (95% CI)	Multivariate analysis *P* value
T stage (T2&T3&T4 vs T1)	35	1.619 (0.583–4.492)	.355	–	–
N stage (N1 vs N0)	30	1.656 (0.350–7.836)	.525	–	–
M stage (M1 vs M0)	32	1.167 (0.259–5.258)	.84	–	–
Pathologic stage (Stage II&Stage III&Stage IV vs Stage I)	35	1.619 (0.583–4.492)	.355	–	–
Histologic grade (G3&G4 vs G1&G2)	35	0.596 (0.212–1.676)	.326	–	–
Residual tumor (R1 vs R0)	33	1.573 (0.438–5.654)	.488	–	–
Age (>65 vs ≤65)	35	1.028 (0.380–2.781)	.956	–	–
Gender (male vs female)	35	1.404 (0.520–3.792)	.503	–	–
Weight (>70 vs ≤70)	35	1.074 (0.371–3.109)	.895	–	–
Height (≥170 vs <170)	34	1.271 (0.450–3.589)	.651	–	–
Histological type (intrahepatic vs distal&hilar/perihilar)	35	1.076 (0.242–4.773)	.924	–	–
CA19-9 level (normal vs abnormal)	29	0.956 (0.306–2.986)	.938	–	–
Child–Pugh grade (B vs A)	20	2.771 (0.306–25.107)	.365	–	–
Fibrosis ishak score (1/2&3/4&5/6 vs 0)	26	0.222 (0.028–1.763)	.155	–	–
Vascular invasion (Yes vs No)	33	1.400 (0.305–6.430)	.666	–	–
Perineural invasion (Yes vs No)	32	5.262 (1.395–19.846)	.014	6.562 (1.553–27.728)	.011
TP53 status (Mut vs WT)	34	0.422 (0.052–3.418)	.419	–	–
CDRT15 (high vs low)	35	3.262 (1.127–9.437)	.029	3.100 (1.003–9.584)	.049

CA19-9 = carbohydrate antigen 19-9, CDRT15 = CMT1A duplicated region transcript 15, CI = confidence interval, HR = hazard ratio, TP53 = tumor protein p53.

genetic factors of CCA is of great significance for promoting the diagnosis and treatment of CCA at the molecular biological level.

CDRT15 is a protein-coding gene, and an important paralog of it is CDRT15L2. The cDNA sequence is divided into 3 exons, encoding CDRT15 protein, an 188 amino acid protein whose protein symbol is Q96T59-CDRT.[12] In Ken Inoue's study, at least 3 cDNA clones contain open reading frames with possible exon/intron structures. Some of them have insertion/deletion mutations that lead to frame-shifts of the open reading frame, and encode different proteins; others have insertions/deletions that result in early termination.[11] An associated disease of CDRT15 is Charcot-Marie-Tooth Disease.[24]

Obviously, whether CDRT15 plays a role in the development of CCA remains unclear. To obtain a reliable conclusion, we analyzed the data in TCGA on bioinformatics. There was no significant difference in gender, age, Child–Pugh grade, TNM stage, CA19-9 value and other baseline data between the high expression group and low expression group of CDRT15. In this study, 51 genes positively correlated with CDRT15 expression were screened from TCGA database, including 32 up-regulated genes and 19 down-regulated genes, were statistically significant. Spearman correlation analysis performed between CDRT15 and other genes showed that 28 genes were statistically significant, such as SAGE1, SOST, TUBA4B, etc. In PubMed and other databases, we found that there was a lack of CDRT15 studies.[25,26]

We analyzed the correlation between expression levels of CDRT15 and 24 types of immune cell infiltration in CCA, and the result showed that the CDRT15 expression was negatively correlated with the Treg, DCs, and Cytotoxic cells. We considered that CDRT15 may be involved in the immune infiltration process of CCA and play a certain role.

In our study, the ROC curve analysis showed that CDRT15 had a good diagnostic value for CCA. Laboratory examinations related to CDRT15 have the potential to be used as a diagnostic basis. Meanwhile, different expression levels of CDRT15 significantly affect the prognosis of patients with CCA. The high expression group had worse overall survival. CDRT15 has potential effect in predicting overall survival and disease-specific survival of patients with CCA at different time points (1, 3 or 5 years).

However, this study also has some limitations. The number of patients with CCA in TCGA and other databases is small, which may produce a large bias. Meanwhile, this study did not collect tissues from specific clinical cases for experiments. Further experimental verification is needed to elucidate the biological functions and mechanism of CDRT15 in CCA. Moreover, we would like to perform more in vivo or in vitro experiments to examine the diagnosis and prognosis effect of CDRT15 in our further research.

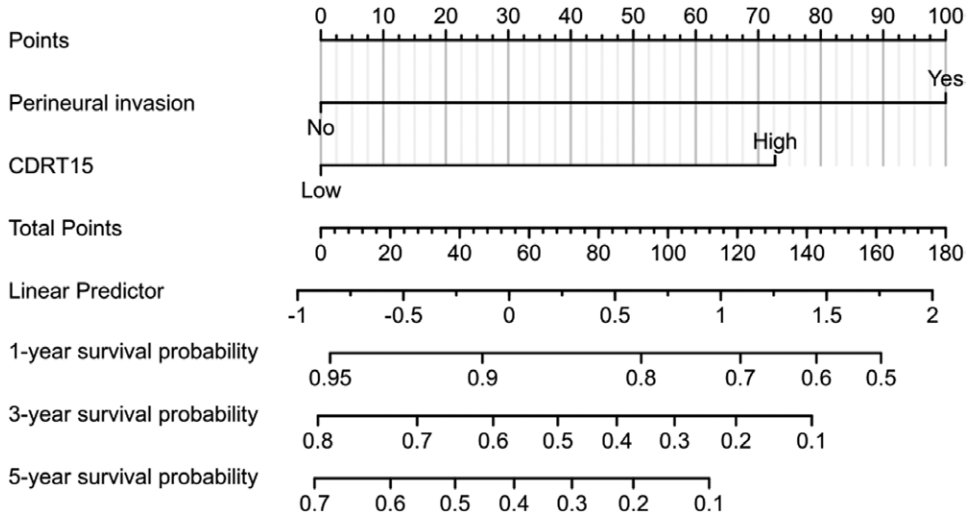

Figure 8. Nomogram of the prognostic model. CDRT15 = CMT1A duplicated region transcript 15.

5. Conclusions

In our research, CDRT15 is highly expressed in CCA tissues compared to normal tissues. According to ROC curve, CDRT15 has the ability to differentiate between normal tissue and tumors. Meanwhile, the CDRT15 expression was negatively correlated with the Treg, DCs, and Cytotoxic cells, which demonstrated that CDRT15 may be involved in the immune infiltration process of CCA and play a certain role. To sum up, CDRT15 may contribute to examination and diagnosis of CCA.

Author contributions

Conceptualization: Tianyang Yu, Luwen Zhao, Aijun Yu.
Data curation: Tianyang Yu, Tiezhao Zhang, Kefan Li.
Formal analysis: Tianyang Yu, Tiezhao Zhang, Kefan Li.
Funding acquisition: Aijun Yu.
Methodology: Tianyang Yu, Luwen Zhao, Jian Li.
Project administration: Luwen Zhao, Jian Li, Aijun Yu.
Supervision: Luwen Zhao, Aijun Yu.
Validation: Tianyang Yu, Tiezhao Zhang, Luwen Zhao.
Visualization: Tianyang Yu, Aijun Yu.
Writing – original draft: Tianyang Yu.
Writing – review & editing: Tianyang Yu, Aijun Yu.

References

[1] Khan SA, Tavolari S, Brandi G. Cholangiocarcinoma: epidemiology and risk factors. Liver Int. 2019;39:19–31.
[2] Mastoraki A, Schizas D, Charalampakis N, et al. Contribution of histone deacetylases in prognosis and therapeutic management of cholangiocarcinoma. Mol Diagn Ther. 2020;24:175–84.
[3] Sha M, Jeong S, Xia Q. Analysis of liver resection versus liver transplantation on outcome of small intrahepatic cholangiocarcinoma and combined hepatocellular-cholangiocarcinoma in the setting of cirrhosis. Liver Transpl. 2020;26:1202–3.
[4] Zhang Y, Esmail A, Mazzaferro V, et al. Newest therapies for cholangiocarcinoma: an updated overview of approved treatments with transplant oncology vision. Cancers (Basel). 2022;14:5074.
[5] Abdelrahim M, Esmail A, Xu J, et al. Gemcitabine plus cisplatin versus non-gemcitabine and cisplatin regimens as neoadjuvant treatment for cholangiocarcinoma patients prior to liver transplantation: an institution experience. Front Oncol. 2022;12:908687.
[6] Borakati A, Froghi F, Bhogal RH, et al. Stereotactic radiotherapy for intrahepatic cholangiocarcinoma. World J Gastrointest Oncol. 2022;14:1478–89.
[7] Valle JW, Kelley RK, Nervi B, et al. Biliary tract cancer. Lancet. 2021;397:428–44.
[8] Nagino M, Hirano S, Yoshitomi H, et al. Clinical practice guidelines for the management of biliary tract cancers 2019: the 3rd English edition. J Hepatobiliary Pancreat Sci. 2021;28:26–54.
[9] Razumilava N, Gores GJ. Cholangiocarcinoma. Lancet. 2014;383:2168–79.
[10] Li Y, Li DJ, Chen J, et al. Application of joint detection of AFP, CA19-9, CA125 and CEA in identification and diagnosis of cholangiocarcinoma. Asian Pac J Cancer Prev. 2015;16:3451–5.
[11] Inoue K, Dewar K, Katsanis N, et al. The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. Genome Res. 2001;11:1018–33.
[12] The Human Gene Database. c1996-2022 [cited 2022 OCT 30]. CDRT15 Gene - CMT1A Duplicated Region Transcript 15. Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CDRT15.
[13] Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44:e71.
[14] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
[15] Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10:1523.
[16] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
[17] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.
[18] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–50.
[19] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. 2013;14:7.
[20] Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity. 2013;39:782–95.
[21] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. 2011;12:77.
[22] Vivian J, Rao AA, Nothaft FA, et al. Toil enables reproducible, open source, big biomedical data analyses. Nat Biotechnol. 2017;35:314–6.
[23] Squadroni M, Tondulli L, Gatta G, et al. Cholangiocarcinoma. Crit Rev Oncol Hematol. 2017;116:11–31.
[24] Pal CV, Eble TN, Burnside RD, et al. Variable levels of tissue mosaicism can confound the interpretation of chromosomal microarray results from peripheral blood. Eur J Med Genet. 2014;57:264–6.
[25] Guo J, Li Y, Duan H, et al. LncRNA TUBA4B functions as a competitive endogenous RNA to inhibit gastric cancer progression by elevating PTEN via sponging miR-214 and miR-216a/b. Cancer Cell Int. 2019;19:156.
[26] Yang Y, Qu Y, Li Z, et al. Identification of novel characteristics in TP53-mutant hepatocellular carcinoma using bioinformatics. Front Genet. 2022;13:874805.