Am J Cancer Res 2021;11(11):5319-5337
www.ajcr.us /ISSN:2156-6976/ajcr0139655

Original Article

ALDH2 is a prognostic biomarker and related with immune infiltrates in HCC

Senbang Yao[1,2*], Xiangxiang Yin[1,2*], Tingting Chen[1*], Wenjun Chen[2,3], He Zuo[1,2], Ziran Bi[1,2], Xiuqing Zhang[1,2], Yanyan Jing[1,2], Lulian Pang[1,2], Huaidong Cheng[1,2]

[1]Department of Oncology, The Second Affiliated Hospital of Anhui Medical University, Hefei 230601, Anhui, China; [2]Department of Oncology, Anhui Medical University, Hefei 230032, Anhui, China; [3]Department of Oncology, Anhui Chest Hospital, Hefei 230061, Anhui, China. [*]Equal contributors.

Received October 8, 2021; Accepted October 29, 2021; Epub November 15, 2021; Published November 30, 2021

Abstract: Hepatocellular carcinoma is a malignant type of carcinoma with complicated pathogenesis. For HCC patients, there is not only a lack of valuable therapeutic targets, but also a lack of prognostic biomarker. The protein encoded by Aldehyde Dehydrogenase 2 Family Member (ALDH2) is a critical member of the aldehyde dehydrogenase family. Many researchers have found that ALDH2 mutations play an important role in the activation of hepatocellular carcinoma carcinogenic pathways. However, the clinicopathological meaning of ALDH2 in HCC and its relation with immune infiltration is still indistinguishable. In this study, we explored the expression of ALDH2 in 41 HCC tissues by immunohistochemistry. The clinicopathological meaning and molecular function of ALDH2 were analyzed and evaluated through comprehensive bioinformatics. ALDH2 expression in HCC was validated in TCGA, GEO and Oncomine databases, and a survival of ALDH2 based on TCGA database was analysed. LinkedOmics was used to classify the co-expressed genes of ALDH2 and its regulatory factors. The relation between ALDH2 and immune infiltration in HCC was further explored by TIMER. IHC results showed decreased levels of ALDH2 in HCC tumor tissues compared with corresponding normal liver tissues. The pathological grade and prognosis of patients with low expression of the ALDH2 gene were worse. Bioinformatics analysis results showed that ALDH2 was considerably down-regulated in cancer tissues compared with corresponding normal liver tissues in 8 GEO series and TCGA profile (all $P<0.05$). A nomogram was designed using expression of ALDH2 and clinical factors. ALDH2 was correlated with dendritic cells and macrophages in immune infiltration. In conclusion, ALDH2 has significant prognostic value in hepatocellular carcinoma and they may play key roles in regulating tumor progression and the immune cells infiltration. Our results suggest that ALDH2 may be a new type of tumor biomarker, which can be used to judge the prognosis, targeted therapy and immunotherapy of patients with HCC.

Keywords: Hepatocellular carcinoma, ALDH2, prognostic markers, immunotherapy, immune infiltration

Introduction

Malignant tumors pose a serious threat to the health of people all around the globe, and many people's lives are taken away by Hepatocellular carcinoma (HCC) under the action of pathogenic factors such as hepatitis and alcohol [1]. HCC is the sixth widespread cancer and the third leading reason of tumor mortality today [2]. Although some patients with liver cancer are cured by local hepatectomy, the overall survival outcome of liver cancer is still poor. The bad prognosis of Hepatocellular carcinoma can be attributed to the reason that diagnosis is usually made at a late phase in cancer stage [3]. The main treatments for HCC include surgery, radiofrequency ablation and biotherapy. However, patients with liver cancer still face a poor 5-year survival time, so exploring new targets is an important research direction for liver cancer treatment. Prognostic markers are of great significance in the treatment and research of HCC patients.

Acetaldehyde dehydrogenase 2 (ALDH2) is a main enzyme for acetaldehyde metabolism during alcohol metabolism, ALDH2 is known for its alcohol oxidation among many aldehyde dehydrogenase genes, and about 30% to 40% of Asians have genetic defects in this enzyme.

Individual exposure to large amounts of the catalytic active form of acetaldehyde may also make it more susceptible to many types of cancer. The correlation between ALDH2 polymorphism and many cancers has been confirmed by scholars from many countries. The ALDH2 gene defect is affiliated with an increased risk of HCC in patients with hepatitis B cirrhosis due to excessive alcohol consumption [4]. The results of Jin S [5] show that ALDH2 plays the role of cancer suppressor in maintaining the balance of the liver genome, and the common human ALDH2 mutation may be an important risk factor for liver cancer. The mechanism of ALDH2 inhibiting HCC progression through ALDH2-acetaldehyde-redox-AMPK axis has been clearly confirmed by many scholars [5, 6], and ALDH2 has been confirmed to be an important hepatocellular carcinoma suppressor gene in vivo and in vitro. However, previous studies have paid little attention to the correlation between ALDH2 expression and clinicopathology of patients, and it is not sufficient for further transformation of ALDH2 to predict prognosis and immunotherapy in HCC patients.

In our study, clinical pathological samples of HCC and cases from TCGA were comprehensively analyzed, the correlation between expression level of ALDH2 and pathological features was analyzed based on abundant sample size. ALDH2 expression was significantly related to tumor grade. The prognostic value of ALDH2 in HCC has been confirmed, and the generated prognostic nomogram can be directly transformed and applied to the prognostic analysis of HCC patients. The correlation between ALDH2 and immune cell infiltration was confirmed by a series of bioinformatics analyses. Overall, this study could be an important step towards the application of ALDH2 transformation in the prediction of HCC clinical prognosis and the modulation of immunotherapy.

Materials and methods

Patients information

All studies obtained the informed consent of the participants. This research was authorized by the Shanghai Outdo Biotech Company. The tissue microarray (TMA) was commissioned and produced by Shanghai Outdo Biotech Company (ODCTDgLiv04001, Shanghai, China), which contained 41 HCC tissues and corresponding 41 non-HCC liver normal tissues. The gene expression and clinical data were downloaded from TCGA. And we also downloaded the data of Platform and Series matrix of 3574 patients in eight datasets (GSE22058, GSE36376, GSE14520, GSE10143, GSE46444, GSE54326, GSE63898 and GSE76427) of HCC samples from GEO. The RStudio (version 3.6.3) and ggplot2 (version 3.3.3) were applied to analysis TCGA and GEO datasets, including annotation, combine, the complement of missing values, and raw data standardization.

Bioinformatics tools

TIMER [7] is an online datasets server for complete exploration of immune cell in tumor-infiltrating. Timer algorithm could be used to assess the abundance of 6 immunoinfiltrates (CD8+ T cells, neutrophils, CD4+ T cells, dendritic cells, B cells, and macrophages). Researchers can input the parameters of specific functions in the Timer Web server, and obtain the results of dynamic display, which is greatly convenient for researchers and clinicians to access the tumor immunological, clinical and genomic characteristics.

TISIDB [8] integrates a variety of data resources of tumor immunology. Clinicians and biologists can analyze the tumor-immune interaction of target genes by high-throughput data and literature mining, and generate high-quality icon results. (http://cis.hku.hk/TISIDB).

GCBI (Gene Cloud of Biotechnology Information) is the web site which we used to research information about gene pathways. (https://www.gcbi.com.cn/gclib).

HCCDB [9] is a convenient tool to exploring gene expression in HCC. It provides the overall different expression chart of liver cancer by analyzing consistent differentially expressed genes in numerous data sets. HCCDB also offers links to third-party databases of proteomics, drugs and literature, and graphically displays several useful outcomes of calculation and analysis.

Linkedomics [10] is a portal site containing multiple sets of data on 32 TCGA tumor types. It also includes spectrometry-based proteomic analysis data generated by CPTAC for TCGA

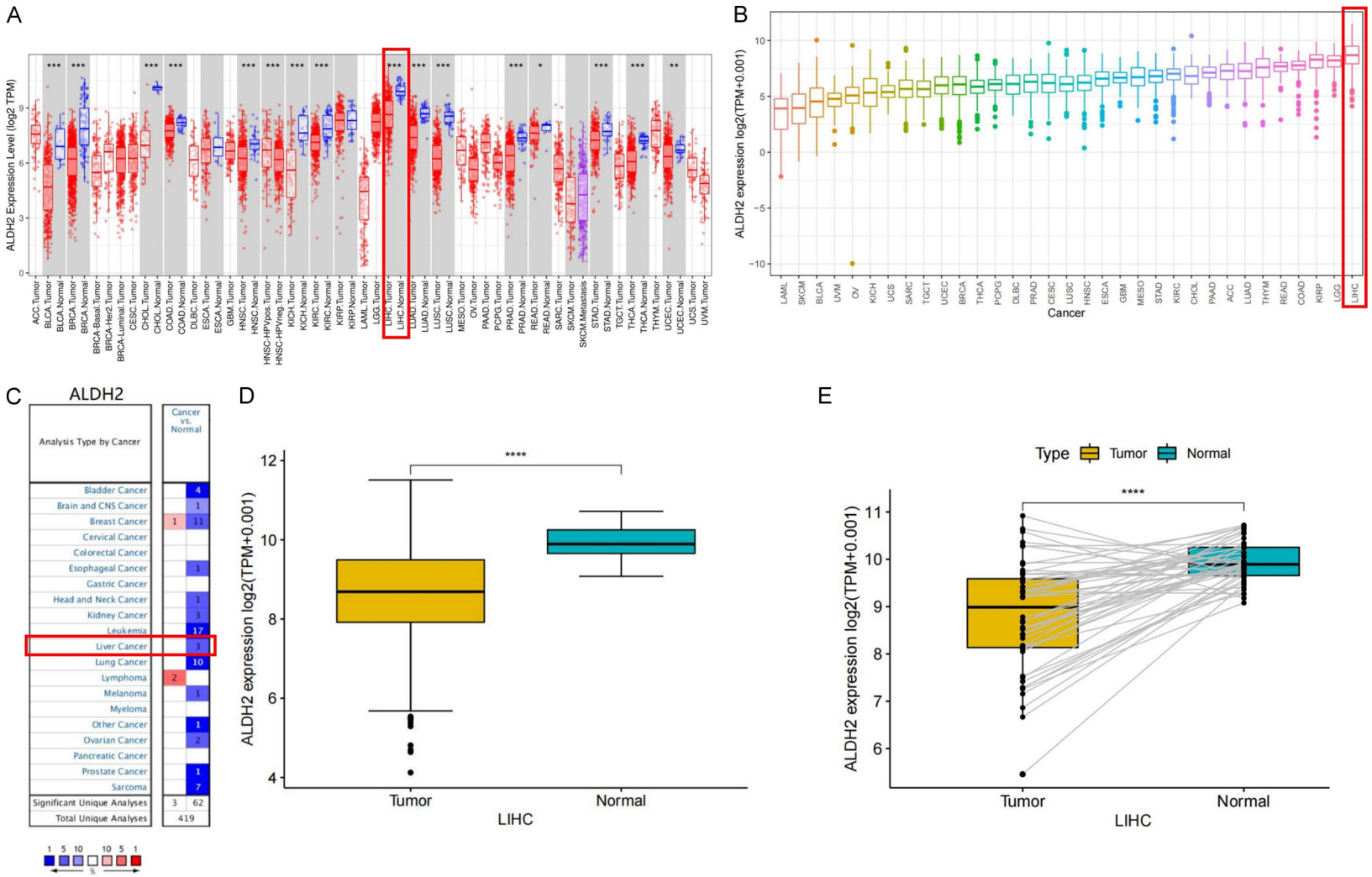


Figure 1. ALDH2 Expression Analysis of Pan-Cancer. A. ALDH2 expression in cancer and normal tissues in pan-cancer TCGA data. B. ALDH2 expression in cancer of TCGA. C. Compared with normal tissues in the Oncomine database, increased or decreased of ALDH2 expression in different cancer data cohorts. D. ALDH2 expression in HCC and unpaired normal tissues from TCGA-GTEx. E. ALDH2 expression HCC and paired normal tissues from TCGA.

Table 1. Detailed Information of involved datasets in this research

Datasets	Contributor(s)	Tumor	Non-tumor	Platform
GSE22058	Burchard J	100	97	Merck/Rosetta Human RSTA Custom Affymetrix 1.0 microarray
GSE36376	Cheol-Keun Park	240	193	Illumina HumanHT-12 V4.0 expression beadchip
GSE14520	Xin Wei Wang	225	220	Affymetrix Human Genome U133A 2.0 Array
GSE10143	Hoshida Y	80	82	Human 6k Transcriptionally Informative Gene Panel for DASL
GSE46444	Chen X	88	48	Illumina Human Whole-Genome DASL HT
GSE54236	Villa E	81	80	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F
GSE63898	Villanueva A	228	168	Affymetrix Human Genome U219 Array
GSE76427	Grinchuk OV	115	52	Illumina HumanHT-12 V4.0 expression beadchip
TCGA-HCC	TCGA	374	49	RNA-Seq

breast and ovarian tumors. In the "LinkFinder" unit of Linkedomics, pearson test was used to complete statistical analysis of ALDH2 co-expressed genes, and the outcomes were shown in the volcano map or the cluster heat map. "LinkInterpreter" unit of LinkedOmics was applied to conduct analyses of Gene Ontology, KEGG pathways, miRNA-target enrichment, kinase-target enrichment and transcription factor-target enrichment through GSEA. The rank criterion was FDR<0.05, and simulations were 500.

KAPLAN-Meier Plotter [11] database is commonly used for cancer prognosis analysis, including 5143 breast cancer samples, and 22277 genes can be used for prognosis analysis. We mainly use the mRNA gene chip module to analyze the survival of OS, RFS and so on.

Immunohistochemistry (IHC)

Immunohistochemistry was completed according to the manufacturer's instructions. After being rehydrated and deparaffinized, the TMA was put into a boiled 0.01 M citrate buffer (pH 6.0) to retrieve the antigen. 3% H_2O_2 was used to block internal peroxidase activity. The primary antibody against ALDH2 was an anti-human ALDH2 polyclonal antibody (15310-1-AP) (Proteintech, USA, dilution 1:200), which was incubated at 4°C overnight and then with HRP (horseradish peroxidase)-labeled secondary antibody at room temperature for 25 min. The last step involved the visualization of the HRP with 3,3'-diaminobenzidine (DAB), followed by dehydration, sealing, and evaluation with a bright-field microscope. Brown-yellow grime in the cytoplasm and/or nucleus indicated positive staining. Sections of negative control were incubated with phosphate-buffered saline during the primary antibody incubation [12].

The positive number and staining intensity of each slice are converted into corresponding values to achieve the purpose of semi-quantitative staining of tissues. H-Score (H-SCORE = $\sum$ (pi×i) = (percentage of weak intensity area ×1) + (percentage of moderate intensity area ×2) + (percentage of strong intensity area ×3)). In this formula, pi is the percentage of the pixel area of the positive signal, and i is the positive level. The higher the value and the stronger the overall positive intensity [13, 14].

Statistical analysis

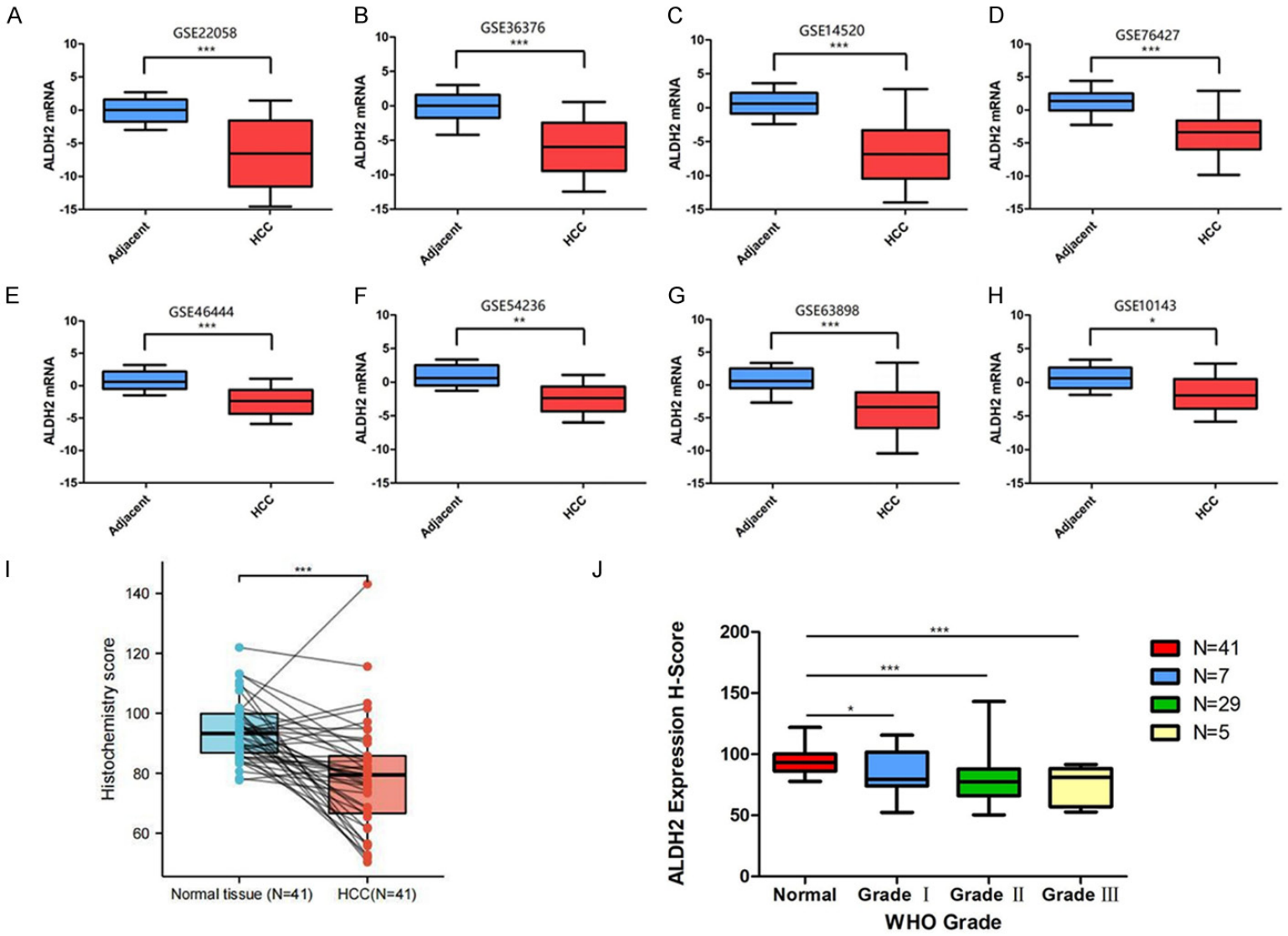
The results of this study were examined by GraphPad Prism5 and SPSS21.0 software. The mean ± standard deviation was used to measure the data. The differences between groups were analyzed by the T test and Chi-square test. The relationship between prognosis and ALDH2 expression was analyzed by Kaplan-Meier survival curve. $P \leq 0.05$ is considered statistically significant.

Results

ALDH2 expression analysis of pan-cancer

We initially explored ALDH2 expression in TCGA pan-cancer database. The results showed that ALDH2 lower expression can be found in 21 kinds of tumor compared with normal tissues (**Figure 1A**). The highest ALDH2 expression was found in HCC compared with other kind of cancer (**Figure 1B**).

Similarly, Oncomine database analysis results showed that compared with normal liver tissue, expression in HCC is highly down-regulat-


A
GSE22058
***
B
GSE36376
***
C
GSE14520
***
D
GSE76427
***
E
GSE46444
***
F
GSE54236
**
G
GSE63898
***
H
GSE10143
*
ALDH2 mRNA
Adjacent
HCC
I
***
Histochemistry score
Normal tissue (N=41)
HCC(N=41)
J
ALDH2 Expression H-Score
***
***
*
N=41
N=7
N=29
N=5
Normal
Grade Ⅰ
Grade Ⅱ
Grade Ⅲ
WHO Grade

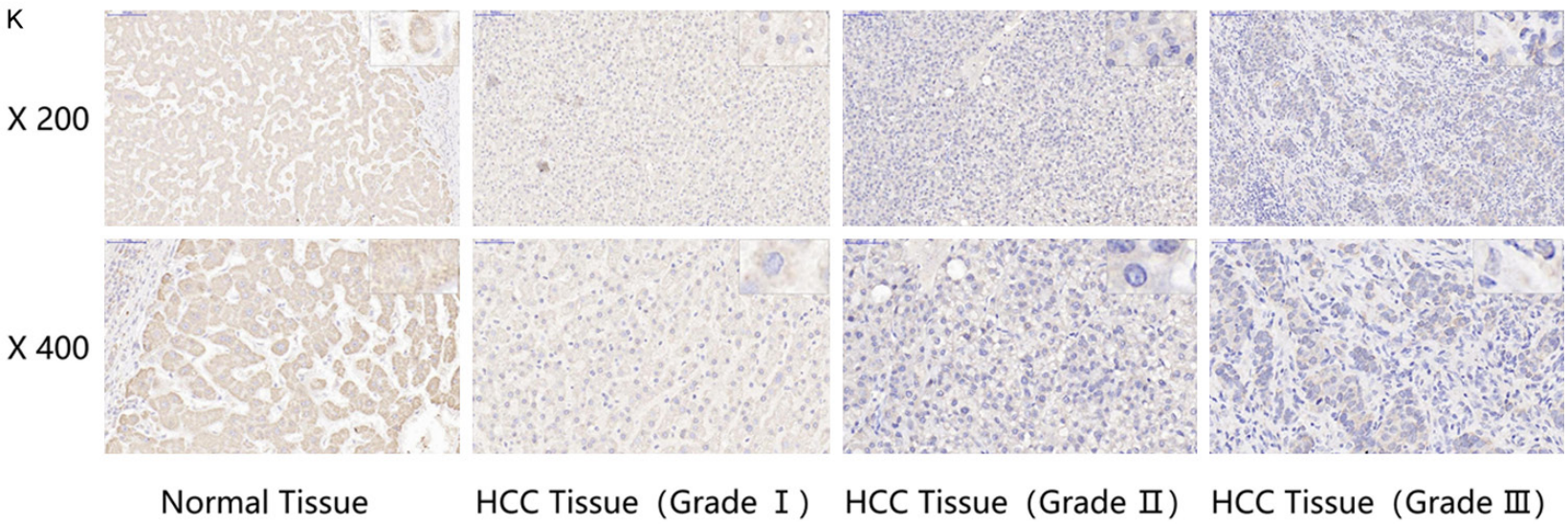


Figure 2. ALDH2 expression in HCC and adjacent normal tissues was verifed by GEO database and IHC. A-H. The expression of ALDH2 was significantly lower in 1157 cases of HCC than 940 adjacent normal tissues in the eight datasets (GSE22058, GSE36376, GSE14520, GSE10143, GSE46444, GSE54236, GSE63898 and GSE76427). The expression of ALDH2 was explored using an unpaired t-test. I-K. ALDH2 immunohistochemistry in adjacent normal tissues and HCC cancer tissues.

Table 2. Information of TCGA-HCC patients in this study

Characteristic	Low expression of ALDH2	High expression of ALDH2	*P*
n	187 n (%)	187 n (%)	
Age			0.055
≤60	99 (26.3%)	79 (21.2%)	
>60	88 (23.6%)	108 (29%)	
Gender			0.047*
Female	70 (18.7%)	51 (13.6%)	
Male	117 (31.3%)	136 (36.4%)	
T stage			0.048*
T1	79 (21.3%)	104 (28%)	
T2	56 (15.1%)	39 (10.5%)	
T3	44 (11.9%)	36 (9.7%)	
T4	8 (2.2%)	5 (1.3%)	
Histologic grade			<0.001***
G1	16 (4.3%)	39 (10.6%)	
G2	80 (21.7%)	98 (26.6%)	
G3	81 (22%)	43 (11.7%)	
G4	8 (2.2%)	4 (1.1%)	
AFP (ng/ml)			<0.001***
≤400	87 (31.1%)	128 (45.7%)	
>400	50 (17.9%)	15 (5.4%)	
Pathologic stage			0.040*
Stage I	76 (21.7%)	97 (27.7%)	
Stage II	50 (14.3%)	37 (10.6%)	
Stage III	49 (14%)	36 (10.3%)	
Stage IV	4 (1.1%)	1 (0.3%)	

Note: *P<0.05, ***P<0.001.

ed (**Figure 1C**). Especially, low ALDH2 expression was detected in HCC of TCGA data compared with paired or unpaired normal adjacent tissues (**Figure 1D**, **1E**). Detailed information of involved datasets in this research was shown in **Table 1**.

ALDH2 expression verification

There were 8 datasets that met the criterion, as follows: GSE22058, GSE36376, GSE14520, GSE10143, GSE46444, GSE54236, GSE63898 and GSE76427. There were 1157 cases of HCC and 940 adjacent normal controls. The analysis result showed that ALDH2 expression was significantly lower in 1157 cases of HCC than 940 adjacent normal controls (**Figure 2A-H**). Then we analyzed 41 pairs of HCC tissues and the corresponding normal tissues to verify ALDH2 expression (**Figure 2I**). IHC result showed that ALDH2 was significantly down-regulated in HCC. In general, like the result of TCGA based on RNA high-throughput sequencing, ALDH2 was also down-regulated in the collected specimen in HCC tissues, which can be used as a prognosis biomarker for HCC. The expression of ALDH2 decreased with increasing grade, being lowest in high-grade HCC (Grade III) (**Figure 2J**, **2K**).

Association between clinico-pathological variables and ALDH2 expression

We evaluated ALDH2 expression data across all patient characteristics derived from TCGA. Information of TCGA-HCC patients in this study were shown in **Table 2**. ALDH2 transcription level was meaningfully lower in HCC patients compared with normal people in subgroup analyses classified by stage of T, stage of N, stage of M, Pathologic stage, Histologic grade and AFP (**Figure 3**). Thus, ALDH2 may be a diagnostic marker for HCC.

Survival prognostic role of ALDH2 in HCC

The correlation between mRNA expression of ALDH2 and survival prognosis in HCC were explored by using the data from GEO and TCGA. Especially, the analysis revealed that ALDH2 expression was significantly in connection with prognosis in HCC (**Figure 4A**, **4B**). Interestingly, the cohort from Kaplan-Meier Plotter including 364 samples showed that ALDH2 high expression is correlated with better prognosis of HCC (OS: HR=0.42, 95% CI 0.29-0.6, *P*=1.3e-06; PFS: HR=0.53, 95% CI 0.38-0.74, *P*=0.00014; RFS: HR=0.62, 95% CI 0.43-0.90, *P*=0.012; DSS: HR=0.43, 95% CI 0.27-0.68, *P*=0.00026) (**Figure 4C**). The prognosis of the high expression of ALDH2 was also shown to be better than ALDH2 with low expression (*P*=0.0044) by the TCGA analysis (**Figure 4D**). The prognostic capacity of ALDH2 was evaluated by using the AUC of a ROC curve. The ROC curve analysis showed that the area which under the curve (AUC) was 0.925 (**Figure 4E**). A nomogram was designed using expression of ALDH2 and clinical factors (**Figure 5**).

Enrichment analysis of ALDH2 functional networks in HCC

In order to clarify the biological function of ALDH2 in HCC, we used the functional module

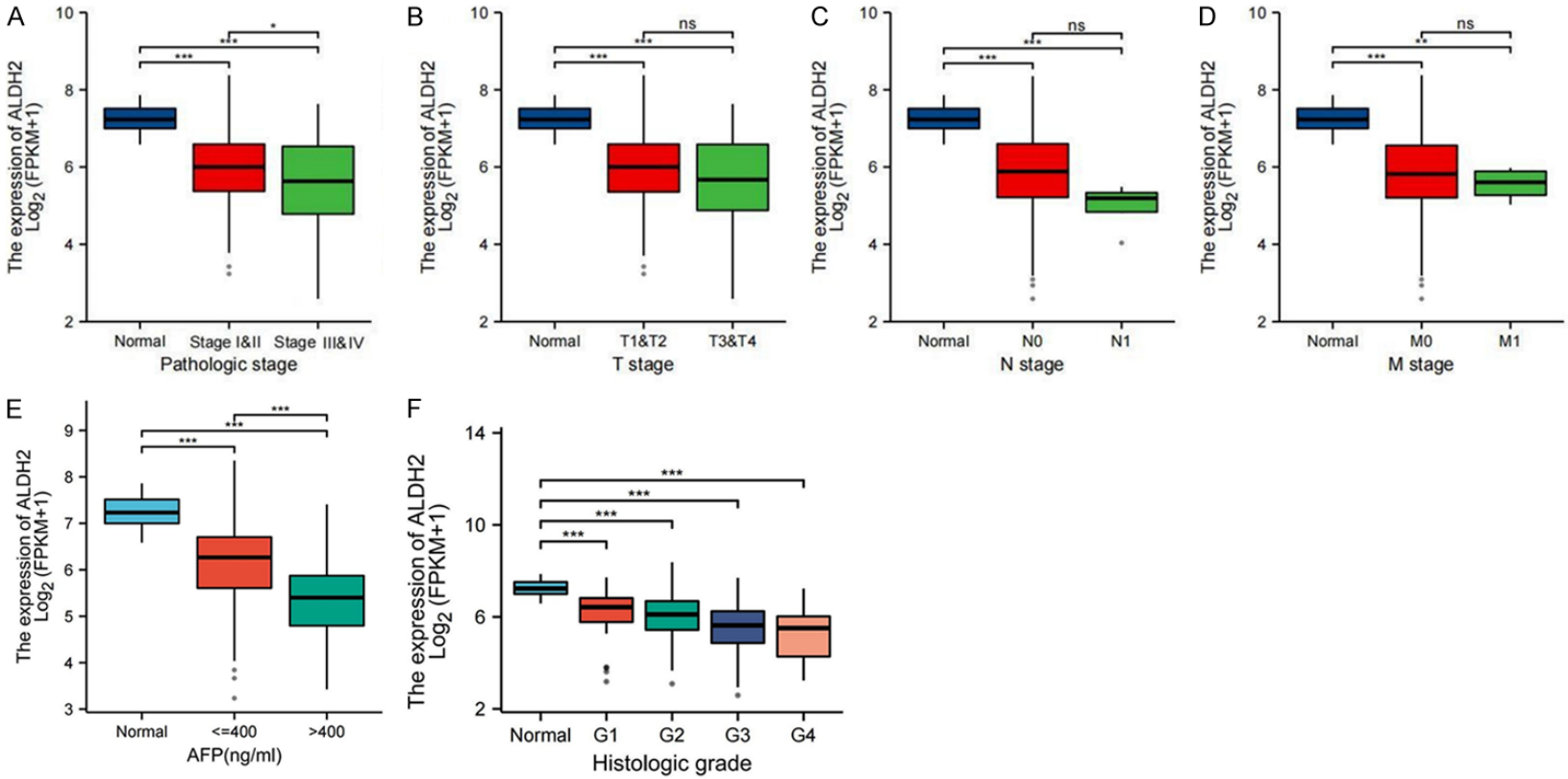
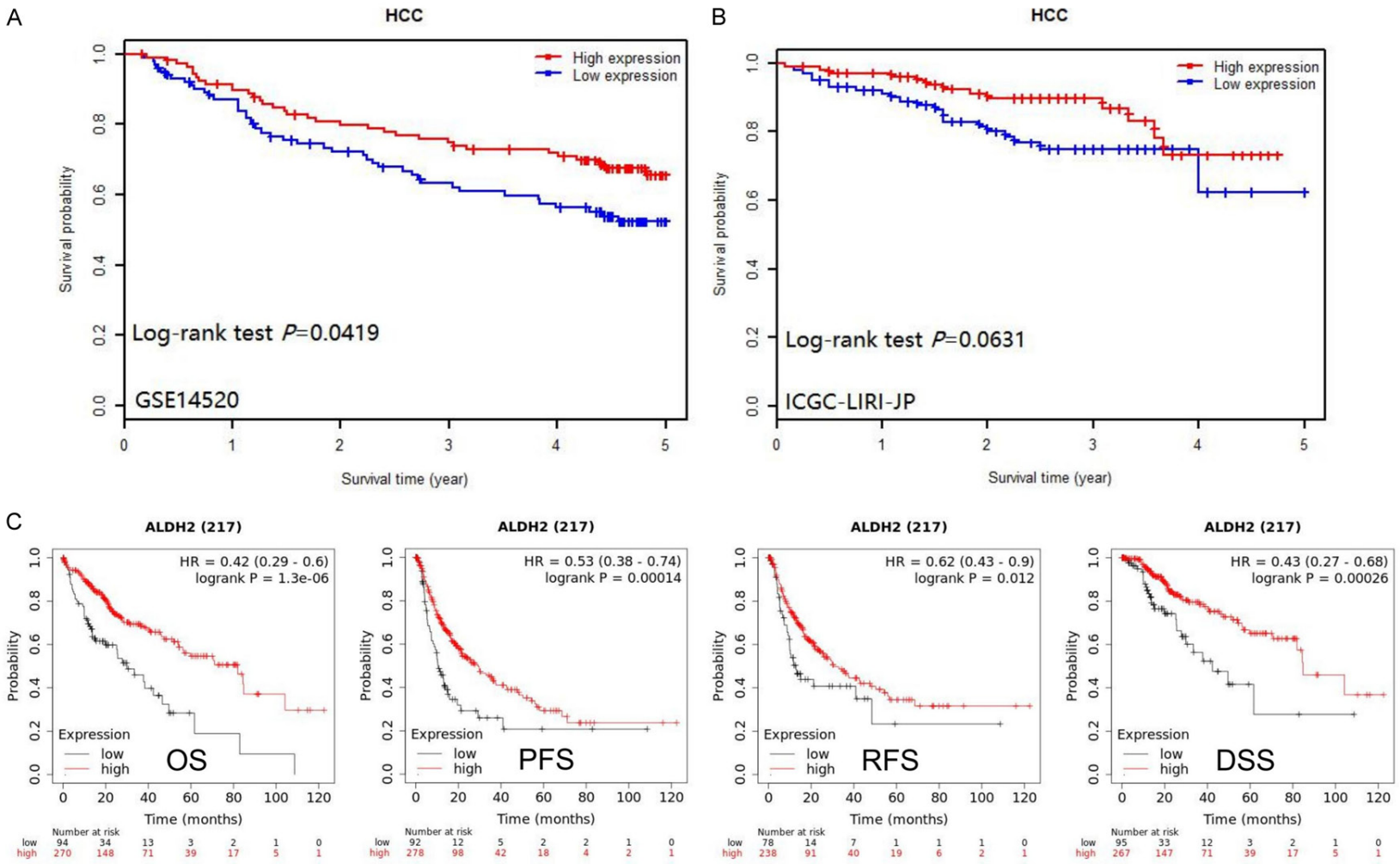


Figure 3. The transcription level of ALDH2 in the subgroup of HCC patients, and classified according to criteria such as stage, grade, lymph node metastasis status, etc. A. Boxplot showing expression of ALDH2 in HCC patients in Pathologic stages 1&2, 3&4. B. The box plot showing that ALDH2 is expressed in HCC patients with T stages 1&2, 3&4. C. The Box plot shows ALDH2 expression in corresponding normal tissues and HCC patients with N stage 0 or 1. D. The Box plot shows ALDH2 expression in corresponding normal tissues and HCC patients with M stage 0 or 1. E. ALDH2 expression in HCC patient with different AFP expression quantity. F. ALDH2 expression in HCC patient with different histological grades.


A
HCC
High expression
Low expression
Survival probability
Log-rank test P=0.0419
GSE14520
Survival time (year)
B
HCC
High expression
Low expression
Survival probability
Log-rank test P=0.0631
ICGC-LIRI-JP
Survival time (year)
C
ALDH2 (217)
HR = 0.42 (0.29 - 0.6)
logrank P = 1.3e-06
Probability
Expression
low
high
OS
Time (months)
Number at risk
low 94 34 13 3 2 1 0
high 270 148 71 39 17 5 1
ALDH2 (217)
HR = 0.53 (0.38 - 0.74)
logrank P = 0.00014
Probability
Expression
low
high
PFS
Time (months)
Number at risk
low 92 12 5 2 2 1 0
high 278 98 42 18 4 2 1
ALDH2 (217)
HR = 0.62 (0.43 - 0.9)
logrank P = 0.012
Probability
Expression
low
high
RFS
Time (months)
Number at risk
low 78 14 7 1 1 1 0
high 238 91 40 19 6 2 1
ALDH2 (217)
HR = 0.43 (0.27 - 0.68)
logrank P = 0.00026
Probability
Expression
low
high
DSS
Time (months)
Number at risk
low 95 33 12 3 2 1 0
high 267 147 71 39 17 5 1

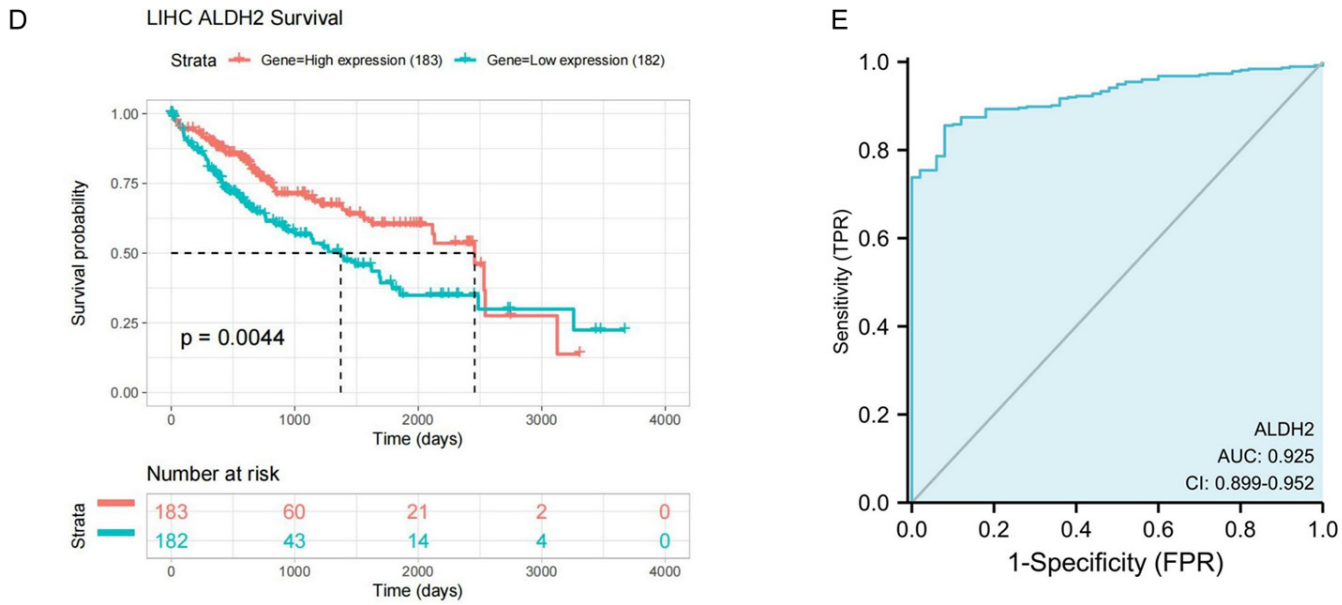


Figure 4. Comparison of Kaplan-Meier survival between high and low expression of ALDH2 in HCC. A. Survival curve for overall survival (OS) in GEO cohort (GSE14520, n=225). B. High ALDH2 expression was in connection with better OS in ICGC cohort (ICGC-LIRI-JP, n=212). C. ALDH2 expression was correlated with better prognosis of HCC (OS, PFS, RFS, DSS). D. High ALDH2 expression was in connection with better OS in TCGA cohort (n=365). E. The prognostic capacity of ALDH2 was evaluated by using the AUC of a ROC curve.

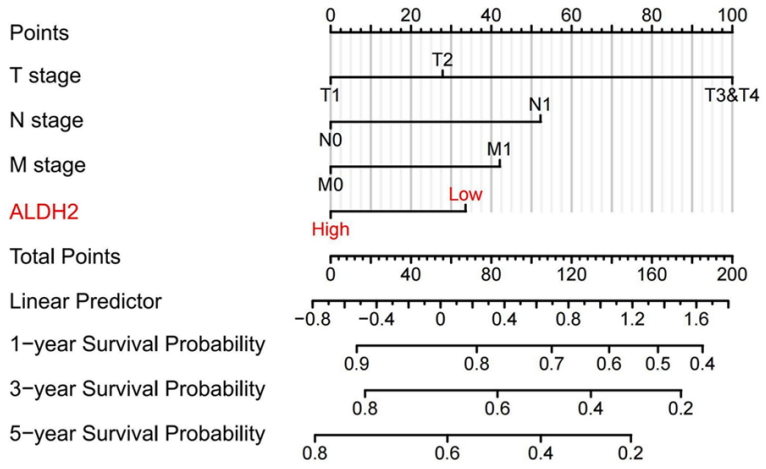


Figure 5. A nomogram for predicting HCC prognosis by ALDH2 expression.

of LinkedOmics to explore the co-expression pattern of ALDH2 in the HCC cohort. As shown in **Figure 6A**, genes of red dots were shown significant positive correlations with ALDH2, whereas genes of green dots were shown significant negative correlations (FDR<0.01). The heat map was used to show the top 50 significant gene positively and negatively related with ALDH2 (**Figure 6B**). Kinases Targets, miRNAs Targets and Transcription factor network of ALDH2 in HCC was shown in **Table 3**. In order to explore the molecular mechanisms of correlated genes in HCC, transcription factors were predicted and a regulatory network of mRNA, microRNA and lncRNA was constructed by GCBI platform (**Figure 6C**). To determine the relationship among this core genes, we constructed GeneMANIA network analysis (**Figure 6D**).

The R software clusterProfiler software package was used to explore the potential functional pathways based on the top 300 genes. GO functional enrichment analysis exposed that ALDH2 was associated mostly with multiple metabolic processes (**Figure 7A-C**). GSEA was conducted to search KEGG and Reactome pathway databases. The KEGG results showed that Valine, leucine and isoleucine degradation were enriched significantly (**Figure 7D**). The Reactome analysis discovered significant enrichment of biological oxidations, protein localization, and diseases of metabolism pathways (**Figure 7E**). These outcomes suggest that ALDH2 expression is connected with the hyperactivation of numerous oncogenic pathways in HCC, especially signaling which control oxidemetabolism.

Relation between immune cells infiltration and ALDH2 expression

In this section, the scores of immune cells infiltration in HCC patients from TCGA database were investigated. The infiltration scores of M1 macrophage were higher in ALDH2 high expression group contrasted with ALDH2 low expression group (Figure S1). Cox analysis of the relation between six immune cells and ALDH2 expression in HCC patients prognosis was visualized in **Table 4**. The expression of ALDH2 was negatively correlated with the infiltration level of macrophages (cor= -0.315, *P*=2.75e-09), dendritic cell (cor= -0.248, *P*=3.59e-06), B cells (cor=-0.241, *P*= 6.25e-06), CD4+ T cells (cor=-0.230, *P*= 1.60e-05), CD8+ T cells (cor=-0.207, *P*= 1.13e-04), and neutrophils (cor=-0.151, *P*= 4.85e-03) in HCC tissues (**Figure 8A**). There was a certain correlation between ALDH2 copy number in dendritic cell and infiltration level (**Figure 8B**).

ALDH2 expression was meaningfully correlated with the Treg cells and M1 macrophage infiltration (**Figure 8C**), indicating that high ALDH2 expression weakens the intratumoral accumulation of T cells and macrophages, especially M1-like macrophage. Such results suggest that ALDH2 high expression is closely correlated to the immune-activated status of HCC. Further correlation analysis showed that ALDH2 was negatively correlated with PD-1 and CTLA4 in HCC (**Figure 8D**).

Discussion

Mutations in the ALDH2 gene have previously been associated with a variety of diseases, it has been widely studied in the past as a therapeutic target for cardiovascular diseases [15], Studies on its association with liver related non-malignant tumor diseases are also common [16], a study using cell and animal experiments suggested that ALDH2-acetaldehyde-redox-AMPK axis may play a modulatory function in the malignant progress of HCC. However, its actual expression and clinical significance in cancer, especially in liver cancer,

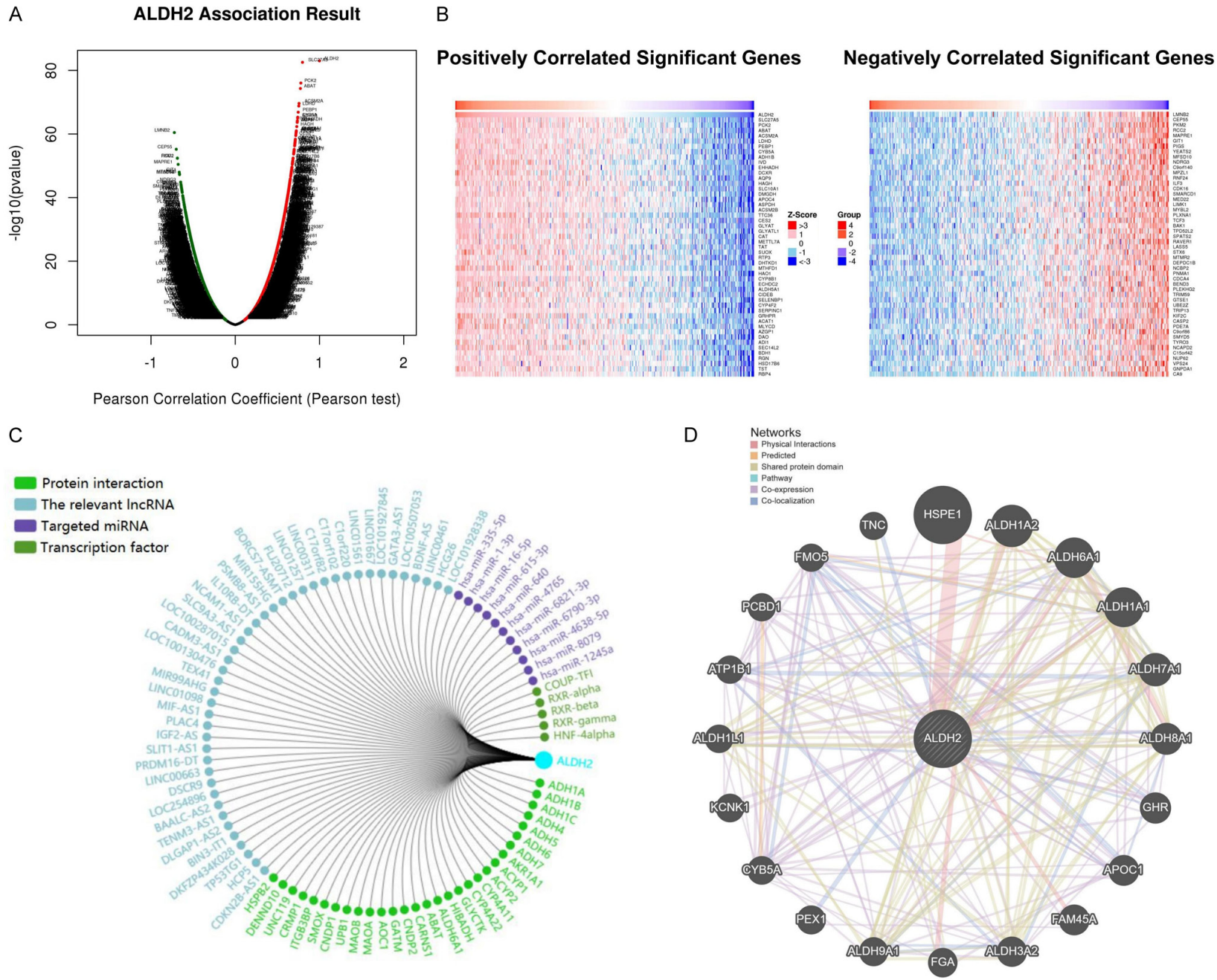

A
ALDH2 Association Result
-log10(pvalue)
Pearson Correlation Coefficient (Pearson test)
B
Positively Correlated Significant Genes
Negatively Correlated Significant Genes
Z-Score
Group
C
Protein interaction
The relevant lncRNA
Targeted miRNA
Transcription factor
ALDH2
D
Networks
Physical Interactions
Predicted
Shared protein domain
Pathway
Co-expression
Co-localization
HSPE1
TNC
ALDH1A2
FMO5
ALDH6A1
PCBD1
ALDH1A1
ATP1B1
ALDH7A1
ALDH1L1
ALDH2
ALDH8A1
KCNK1
GHR
CYB5A
APOC1
PEX1
FAM45A
ALDH9A1
FGA
ALDH3A2

Figure 6. ALDH2 co-expressed genes in HCC. A. Volcano figure of ALDH2 highly correlated genes identified by Pearson test in HCC cohort. B. Heat maps showing top 50 genes negatively and positively correlated with ALDH2 in HCC. C. The GCBI database was used to display the hub genes network. D. The GeneMANIA database was used to display the hub genes network.

Table 3. Kinases Targets, miRNAs Targets and Transcription factor network of ALDH2 in HCC

Enriched Categories	Genesets	LeadingEdgeNum	FDR
Kinase Targets	Kinase_ATM	45	0.00E+00
	Kinase_CDK1	105	0.00E+00
	Kinase_PRKD1	14	3.51E-4
	Kinase_MAPKAPK2	10	4.17E-3
	Kinase_PRKD3	3	1.72E-2
miRNA Targets	GACAGGG, MIR-339	17	3.50E-2
	CTCAGGG, MIR-125B, MIR-125A	93	3.99E-2
	ACATTCC, MIR-1, MIR-206	58	4.38E-2
	AGGGCCA, MIR-328	26	7.07E-2
	GACTGTT, MIR-212, MIR-132	55	9.28E-2
Transcription Factors	V$E2F_Q4_01	86	0.00E+00
	V$E2F4DP2_01	87	0.00E+00
	V$AP2GAMMA_01	76	1.37E-3
	V$HNF1_Q6	52	1.05E-1
	V$HNF4_DR1_Q3	55	1.27E-1

have not been explored deeply. We have explored the public data set through bioinformatics analysis in order to offer a direction for the research of the potential biological mechanism of ALDH2 in HCC.

Our analysis of the data confirmed that ALDH2 mRNA in HCC was meaningfully lower than that in liver normal tissues (**Figure 7**). In addition, ALDH2 low expression is significantly correlated with poor prognosis survival and pathological grading in multiple cohorts. This is consistent with the results of Zahid KR [17] and other scholars, that is, patients with high expression of ALDH2 and ADH1A genes tend to have better prognostic survival and lower state of disease invasion. Therefore, our results show that down-regulation of ALDH2 appears in many HCC cases and as a possible prognostic and diagnostic biomarker, it is worthy expecting of further clinical exploration.

By exploring the regulatory factors that may affect the abnormal regulation of ALDH2, we found that ALDH2 in HCC is correlated with kinases including ATM, CDK1, PRKD1, MAPKAPK2 and PRKD3. Such kinases control the cell cycle, gene transcription, regulation of cell shape and showed different expression and survival prognosis in HCC. In fact, there is a certain correlation between the protein encoded by the ATM gene and the PI3/PI4 kinase gene family. The ATM is an important phosphorylated checkpoint kinase of cell cycle, and targeting ATM may be a promising strategy for cancer treatment. Targeted CDK1 inhibitors are being widely developed, and some have been used in clinical trials (phase I and phase II), showing significant effect on various solid tumors and hematologic malignancies [18]. In HCC, ALDH2 plays a vital anti-metastatic role in HCC and can be used as a predictor of HCC prognoses [6].

Subsequently, we found that the key transcription factor for the abnormal regulation of ALDH2 is the E2F family protein. E2F1 participates in the regulation of key links in the cell cycle network. Abnormal activation of E2F oncogenic signals can often be found in the progression of HCC, and researches have confirmed that the increase of E2F1 and E2F3 copy number drives the progression of HCC [6]. Based on our analysis results, it can be predicted that ALDH2 has an important regulatory effect on E2F1, and ALDH2 may regulate the proliferation ability and cell cycle of liver cancer

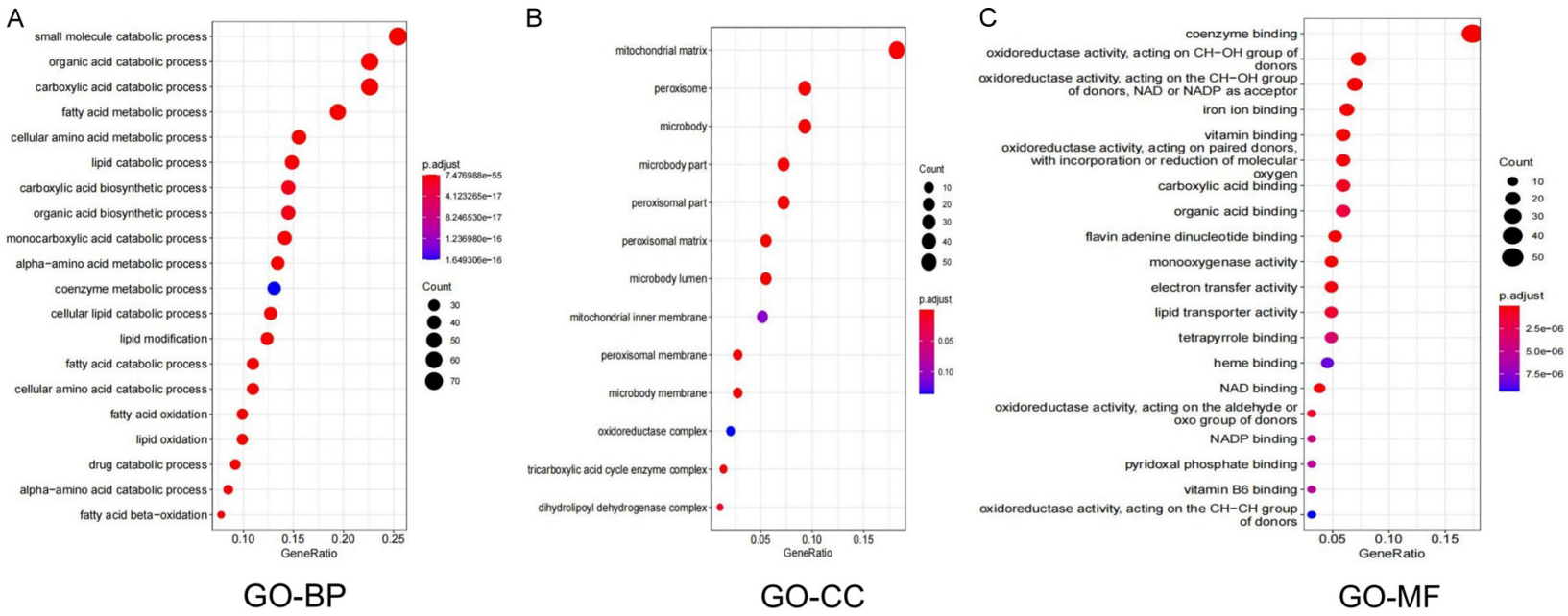

A
small molecule catabolic process
organic acid catabolic process
carboxylic acid catabolic process
fatty acid metabolic process
cellular amino acid metabolic process
lipid catabolic process
carboxylic acid biosynthetic process
organic acid biosynthetic process
monocarboxylic acid catabolic process
alpha−amino acid metabolic process
coenzyme metabolic process
cellular lipid catabolic process
lipid modification
fatty acid catabolic process
cellular amino acid catabolic process
fatty acid oxidation
lipid oxidation
drug catabolic process
alpha−amino acid catabolic process
fatty acid beta−oxidation
0.10 0.15 0.20 0.25
GeneRatio
p.adjust
7.476988e−55
4.123265e−17
8.246530e−17
1.236980e−16
1.649306e−16
Count
30 40 50 60 70
GO-BP
B
mitochondrial matrix
peroxisome
microbody
microbody part
peroxisomal part
peroxisomal matrix
microbody lumen
mitochondrial inner membrane
peroxisomal membrane
microbody membrane
oxidoreductase complex
tricarboxylic acid cycle enzyme complex
dihydrolipoyl dehydrogenase complex
0.05 0.10 0.15
GeneRatio
Count
10 20 30 40 50
p.adjust
0.05
0.10
GO-CC
C
coenzyme binding
oxidoreductase activity, acting on CH−OH group of donors
oxidoreductase activity, acting on the CH−OH group of donors, NAD or NADP as acceptor
iron ion binding
vitamin binding
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
carboxylic acid binding
organic acid binding
flavin adenine dinucleotide binding
monooxygenase activity
electron transfer activity
lipid transporter activity
tetrapyrrole binding
heme binding
NAD binding
oxidoreductase activity, acting on the aldehyde or oxo group of donors
NADP binding
pyridoxal phosphate binding
vitamin B6 binding
oxidoreductase activity, acting on the CH−CH group of donors
0.05 0.10 0.15
GeneRatio
Count
10 20 30 40 50
p.adjust
2.5e−06
5.0e−06
7.5e−06
GO-MF
D
Valine, leucine and isoleucine degradation
Peroxisome
Drug metabolism − cytochrome P450
Retinol metabolism
Metabolism of xenobiotics by cytochrome P450
Chemical carcinogenesis
PPAR signaling pathway
Complement and coagulation cascades
Fatty acid metabolism
Carbon metabolism
Steroid hormone biosynthesis
Bile secretion
Adipocytokine signaling pathway
Biosynthesis of amino acids
Glycolysis / Gluconeogenesis
Insulin resistance
Drug metabolism − other enzymes
AMPK signaling pathway
FoxO signaling pathway
Glucagon signaling pathway
0.3 0.6 0.9
p.adjust
0.01006648
GSEA-KEGG
E
Fatty acid metabolism
Biological oxidations
Phase I − Functionalization of compounds
Peroxisomal protein import
Metabolism of steroids
Phase II − Conjugation of compounds
Cytochrome P450 − arranged by substrate type
Protein localization
Metabolism of lipids
Diseases of metabolism
Metabolism of vitamins and cofactors
Metabolism of water−soluble vitamins and cofactors
Plasma lipoprotein assembly, remodeling, and clearance
Regulation of lipid metabolism by Peroxisome proliferator−activated receptor alpha (PPARalpha)
PPARA activates gene expression
Circadian Clock
Metabolism of amino acids and derivatives
SLC−mediated transmembrane transport
Transport of small molecules
Metabolism of carbohydrates
0.3 0.6 0.9
p.adjust
0.02629949
GSEA-Reactome

Figure 7. ALDH2 enrichment analysis in HCC. A. Gene Ontology-biological processes. B. Gene Ontology-cell component. C. Gene Ontology-molecular function. D. GSEA-KEGG pathways. E. GSEA-Reactome pathways.

Table 4. Cox analysis of the relation between six immune cells and ALDH2 expression in HCC patients prognosis

Variable	Coef	HR	95% CI_l	95% CI_u	*P*.value	Sig
Age	0.013	1.013	0.996	1.030	0.144	
gendermale	-0.022	0.978	0.608	1.573	0.926	
raceBlack	1.383	3.989	1.418	11.221	0.009	**
raceWhite	0.019	1.019	0.617	1.681	0.942	
stage 2	0.108	1.114	0.652	1.902	0.693	
stage 3	0.672	1.958	1.197	3.200	0.007	**
stage 4	1.235	3.438	0.991	11.931	0.052	
Purity	0.958	2.608	0.839	8.100	0.097	
B_cell	-8.447	0.000	0.000	0.394	0.028	*
CD8_Tcell	-6.238	0.002	0.000	0.292	0.015	*
CD4_Tcell	-7.735	0.000	0.000	0.737	0.041	*
Macrophage	6.520	678.438	1.997	230470.259	0.028	*
Neutrophil	-1.158	0.314	0.000	31935.325	0.844	
Dendritic	5.687	295.009	6.992	12446.737	0.003	**
ALDH2	-0.326	0.722	0.578	0.901	0.004	**

Note: *P<0.05, **P<0.01.

by this factor, which should to be investigated by further researches.

To explore possible mechanisms that regulate the abnormal expression of ALDH2, we explored the co-expression genes network of ALDH2. The results suggest that the functional elements of ALDH2 mostly include cell cycle, mRNA surveillance while it enhances the processes of metabolism, such as fatty acid, and cellular modified amino acid metabolic process. The molecular pathways of HCC carcinogenesis are closely related to these findings [19].

Our research also identified ALDH2-related miRNAs, these short chains of noncoding RNAs affect gene expression by participating in the transcription can promote the progress of tumor. miR-339, miR-206 can be used as prognostic and diagnostic markers of HCC [20-22].

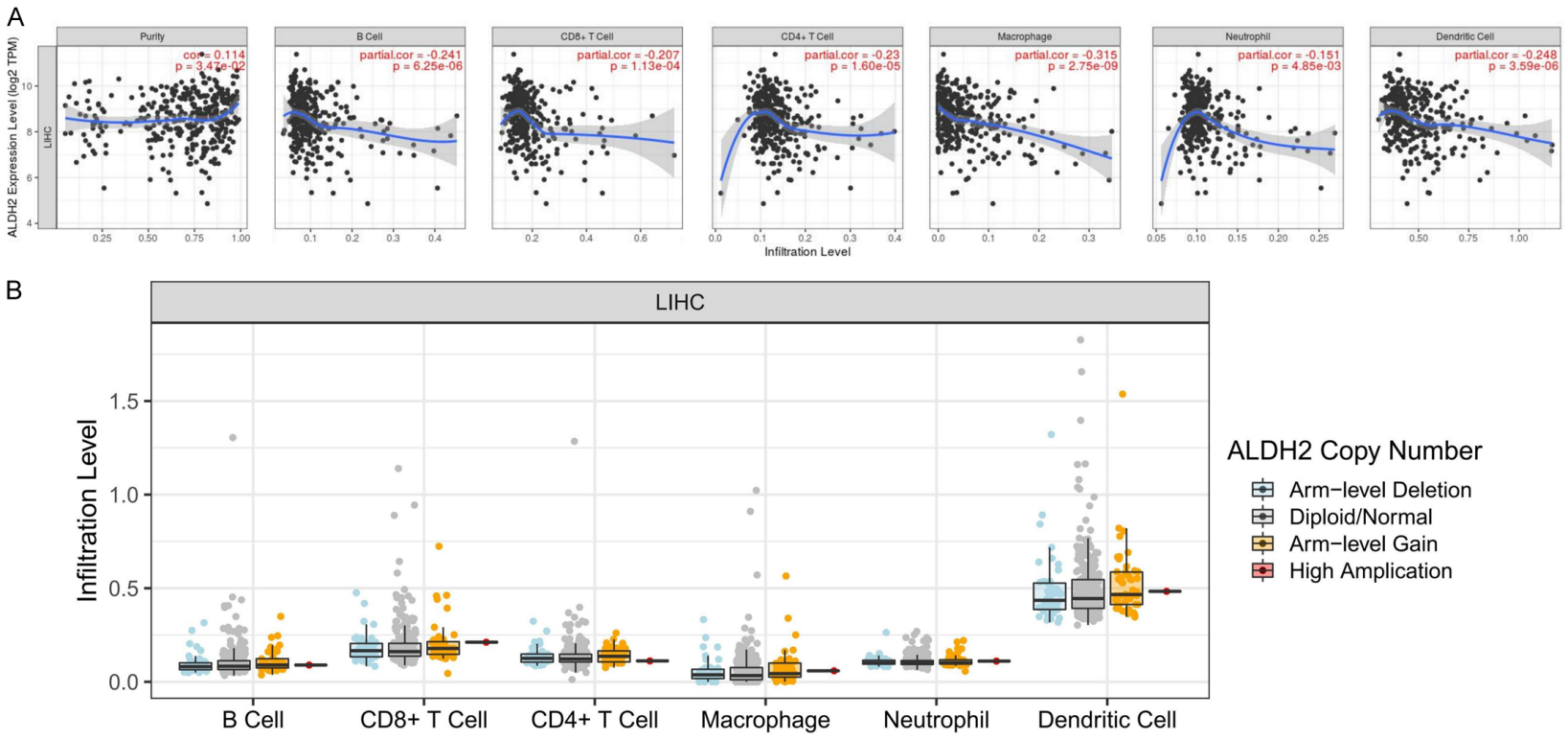
The results of our study provide a sufficient theoretical basis for the important role of ALDH2 in progression and occurrence of HCC. It also reveals its potential as an HCC tumor biomarker. Our research found that ALDH2 low expression in HCC has thoughtful effects on genomic stability, gene expression and a series process of cell cycle. ALDH2 is specially correlated to several cancer-correlated kinases (for instance ATM), MicroRNAs (for instance miRNA-339), and transcription factor (for instance E2F1). In this study, we used advanced cancer bioinformatics analysis tools to comprehensively analyze the target gene ALDH2 in cancer data from the public database. Compared with the previous analysis method, this study has a large sample size, simple operation and low cost. This is of great significance for the future HCC genomics research and future functional mechanism exploration.

Through the analysis of more than 1000 clinical samples, we found that ALDH2 mRNA expression level in hepatocellular carcinoma was significantly down-regulated compared with that of corresponding non-cancer liver tissue. Low expression of ALDH2 is often associated with higher tumor grade and worse survival and prognosis. We believe that ALDH2 could play an important role in inhibiting the progress of malignant tumors in HCC, and its specific biological mechanism remains to be further explored.

Although we have preliminarily clarified the biological functions of ALDH2 in immune infiltration and immunotherapy on the basis of previous studies, the current research also had some limitations. Further in vivo model studies are needed to confirm the effect of ALDH2 on immune infiltration and its correlation with PD-1/PD-L1 inhibitor efficacy.

Conclusion

Due to the prevalence of ALDH2 mutations in populations, ALDH2 may be the most likelihood prognostic marker and immunotherapeutic effector for HCC in clinical application. On the whole, our study provided a comprehensive signal for the role of ALDH2 in the progression of HCC and its prospective as a biological prognostic predictor and therapeutic


A
Purity
cor = 0.114
p = 3.47e-02
B Cell
partial.cor = -0.241
p = 6.25e-06
CD8+ T Cell
partial.cor = -0.207
p = 1.13e-04
CD4+ T Cell
partial.cor = -0.23
p = 1.60e-05
Macrophage
partial.cor = -0.315
p = 2.75e-09
Neutrophil
partial.cor = -0.151
p = 4.85e-03
Dendritic Cell
partial.cor = -0.248
p = 3.59e-06
ALDH2 Expression Level (log2 TPM)
LIHC
Infiltration Level
B
LIHC
Infiltration Level
B Cell
CD8+ T Cell
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
ALDH2 Copy Number
Arm−level Deletion
Diploid/Normal
Arm−level Gain
High Amplication

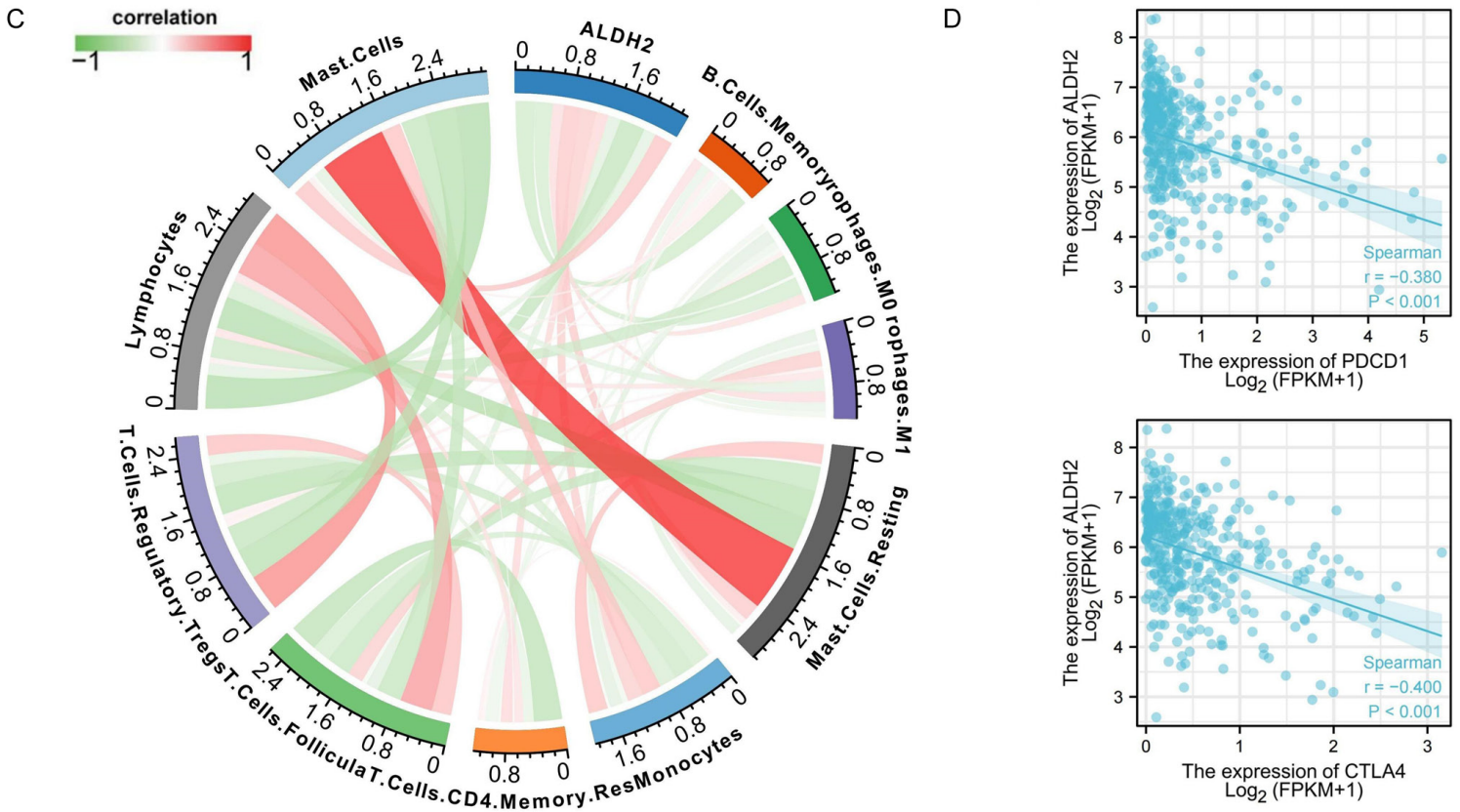


Figure 8. The correlation analysis between ALDH2 and immune cells infiltration in HCC. A. The correlation between ALDH2 and immunocytes infiltration. B. Comparison of infiltration levels among HCC with different ALDH2 copy number alterations. C. The correlation between ALDH2 and levels of immune cells infiltration, green represents negative correlation, red represents positive correlation, and the deeper the color, the stronger the correlation. D. The correlation between ALDH2 and PD1, CTLA4 in HCC.

target. Our researches suggest that the down-regulation of ALDH2 in HCC is associated with poor prognostic survival, which could be due to complicated steps that weaken the stability of the genome. Also, we found that there was a substantial correlation between ALDH2 and most immune characteristics, this will be an important research direction from this time forth.

Acknowledgements

We acknowledge the National Natural Science Foundation of China (No. 81872504) for its financial support. We are deeply grateful for the cooperation of cancer patients and their families.

Disclosure of conflict of interest

None.

Address correspondence to: Huaidong Cheng, Department of Oncology, The Second Affiliated Hospital of Anhui Medical University, 678 Furong Road, Hefei 230601, Anhui, China. Tel: +86-13955112735; E-mail: chd1975ay@126.com

References

[1] Siegel RL, Miller KD, Fuchs HE and Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021; 71: 7-33.

[2] Zhu CP, Wang AQ, Zhang HH, Wan XS, Yang XB, Chen SG and Zhao HT. Research progress and prospects of markers for liver cancer stem cells. World J Gastroenterol 2015; 21: 12190-12196.

[3] Sun N, Lee YT, Zhang RY, Kao R, Teng PC, Yang Y, Yang P, Wang JJ, Smalley M, Chen PJ, Kim M, Chou SJ, Bao L, Wang J, Zhang X, Qi D, Palomique J, Nissen N, Han SB, Sadeghi S, Finn RS, Saab S, Busuttil RW, Markovic D, Elashoff D, Yu HH, Li H, Heaney AP, Posadas E, You S, Yang JD, Pei R, Agopian VG, Tseng HR and Zhu Y. Purification of HCC-specific extracellular vesicles on nanosubstrates for early HCC detection by digital scoring. Nat Commun 2020; 11: 4489.

[4] Seo W, Gao Y, He Y, Sun J, Xu H, Feng D, Park SH, Cho YE, Guillot A, Ren T, Wu R, Wang J, Kim SJ, Hwang S, Liangpunsakul S, Yang Y, Niu J and Gao B. ALDH2 deficiency promotes alcohol-associated liver cancer by activating oncogenic pathways via oxidized DNA-enriched extracellular vesicles. J Hepatol 2019; 71: 1000-1011.

[5] Jin S, Chen J, Chen L, Histen G, Lin Z, Gross S, Hixon J, Chen Y, Kung C, Chen Y, Fu Y, Lu Y, Lin H, Cai X, Yang H, Cairns RA, Dorsch M, Su SM, Biller S, Mak TW and Cang Y. ALDH2(E487K) mutation increases protein turnover and promotes murine hepatocarcinogenesis. Proc Natl Acad Sci U S A 2015; 112: 9088-9093.

[6] Hou G, Chen L, Liu G, Li L, Yang Y, Yan HX, Zhang HL, Tang J, Yang YC, Lin X, Chen X, Luo GJ, Zhu Y, Tang S, Zhang J, Liu H, Gu Q, Zhao LH, Li Y, Liu L, Zhou W and Wang H. Aldehyde dehydrogenase-2 (ALDH2) opposes hepatocellular carcinoma progression by regulating AMP-activated protein kinase signaling in mice. Hepatology 2017; 65: 1628-1644.

[7] Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B and Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 2017; 77: e108-e110.

[8] Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW and Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. Bioinformatics 2019; 35: 4200-4202.

[9] Lian Q, Wang S, Zhang G, Wang D, Luo G, Tang J, Chen L and Gu J. HCCDB: a database of hepatocellular carcinoma expression atlas. Genomics Proteomics Bioinformatics 2018; 16: 269-275.

[10] Vasaikar SV, Straub P, Wang J and Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res 2018; 46: D956-D963.

[11] Nagy Á, Munkácsy G and Győrffy B. Pancancer survival analysis of cancer hallmark genes. Sci Rep 2021; 11: 6047.

[12] Lu HP, Du XF, Li JD, Huang SN, He RQ, Wu HY, Li MF, Wu WZ, Chen JT, Mo WJ and Chen G. Expression of cell division cycle protein 45 in tissue microarrays and the CDC45 gene by bioinformatics analysis in human hepatocellular carcinoma and patient outcomes. Med Sci Monit 2021; 27: e928800.

[13] Paschalis A, Sheehan B, Riisnaes R, Rodrigues DN, Gurel B, Bertan C, Ferreira A, Lambros MBK, Seed G, Yuan W, Dolling D, Welti JC, Neeb A, Sumanasuriya S, Rescigno P, Bianchini D, Tunariu N, Carreira S, Sharp A, Oyen W and de Bono JS. Prostate-specific membrane antigen heterogeneity and DNA repair defects in prostate cancer. Eur Urol 2019; 76: 469-478.

[14] Guo R, Berry LD, Aisner DL, Sheren J, Boyle T, Bunn PA Jr, Johnson BE, Kwiatkowski DJ, Drilon A, Sholl LM and Kris MG. MET IHC is a poor screen for MET amplification or MET exon 14 mutations in lung adenocarcinomas: data from a tri-institutional cohort of the lung cancer mutation consortium. J Thorac Oncol 2019; 14: 1666-1671.

[15] Chen CH, Ferreira JC, Gross ER and Mochly-Rosen D. Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. Physiol Rev 2014; 94: 1-34.
[16] Guo R, Xu X, Babcock SA, Zhang Y and Ren J. Aldehyde dedydrogenase-2 plays a beneficial role in ameliorating chronic alcohol-induced hepatic steatosis and inflammation through regulation of autophagy. J Hepatol 2015; 62: 647-656.
[17] Zahid KR, Yao S, Khan ARR, Raza U and Gou D. mTOR/HDAC1 crosstalk mediated suppression of ADH1A and ALDH2 links alcohol metabolism to hepatocellular carcinoma onset and progression in silico. Front Oncol 2019; 9: 1000.
[18] Johnson N, Li YC, Walton ZE, Cheng KA, Li D, Rodig SJ, Moreau LA, Unitt C, Bronson RT, Thomas HD, Newell DR, D'Andrea AD, Curtin NJ, Wong KK and Shapiro GI. Compromised CDK1 activity sensitizes BRCA-proficient cancers to PARP inhibition. Nat Med 2011; 17: 875-882.
[19] Couri T and Pillai A. Goals and targets for personalized therapy for HCC. Hepatol Int 2019; 13: 125-137.
[20] Zeng H, Zheng J, Wen S, Luo J, Shao G and Zhang Y. MicroRNA-339 inhibits human hepatocellular carcinoma proliferation and invasion via targeting ZNF689. Drug Des Devel Ther 2019; 13: 435-445.
[21] Hu W, Feng H, Xu X, Huang X, Huang X, Chen W, Hao L and Xia W. Long noncoding RNA FOXD2-AS1 aggravates hepatocellular carcinoma tumorigenesis by regulating the miR-206/MAP3K1 axis. Cancer Med 2020; 9: 5620-5631.
[22] Lu H, Hu J, Li J, Lu W, Deng X and Wang X. miR-328-3p overexpression attenuates the malignant proliferation and invasion of liver cancer via targeting endoplasmic reticulum metallo protease 1 to inhibit AKT phosphorylation. Ann Transl Med 2020; 8: 754.

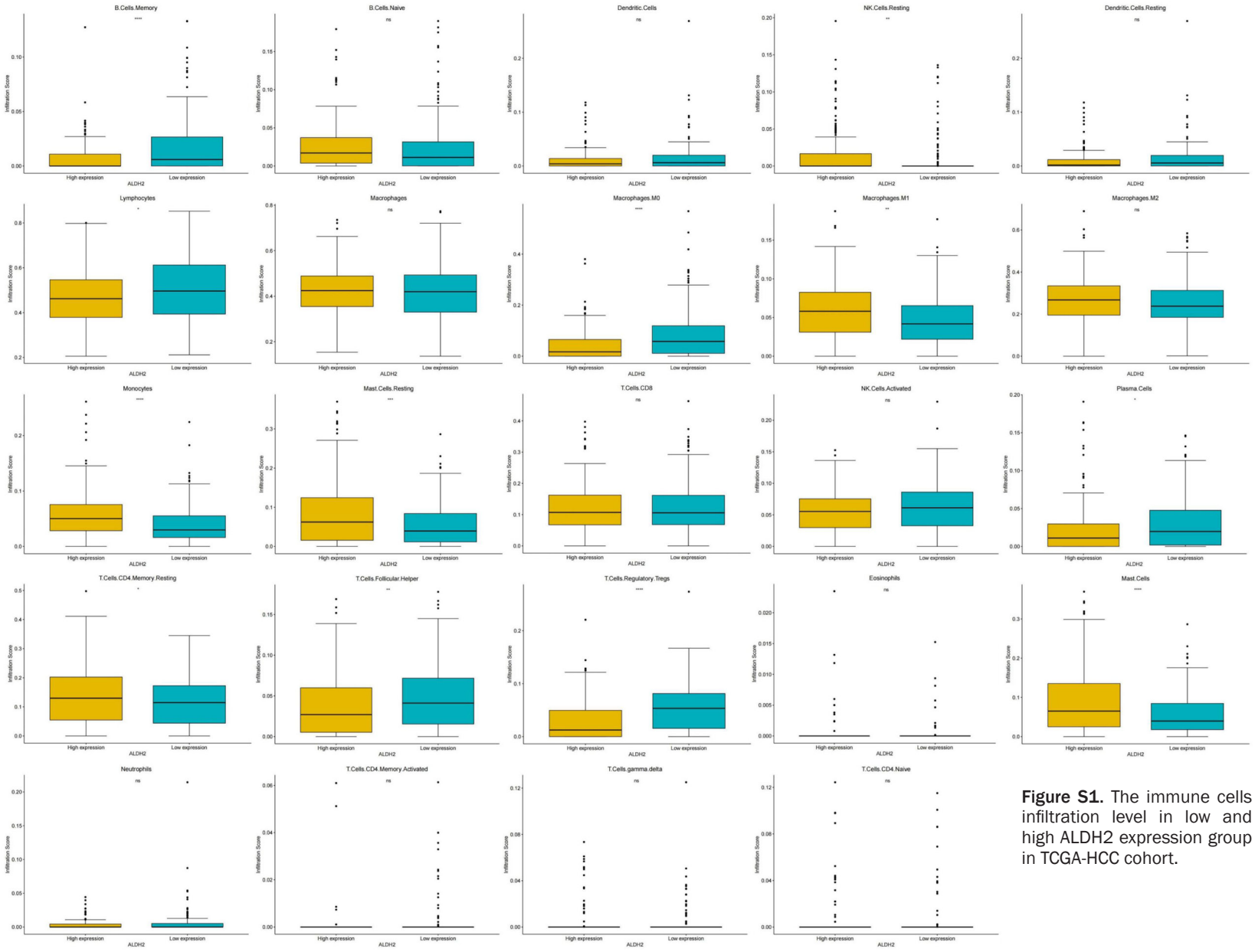


Figure S1. The immune cells infiltration level in low and high ALDH2 expression group in TCGA-HCC cohort.