Exploring ALDH2 Expression and Its Immune Infiltration in HNSC and Its Prognositic Correlation With Gender or Alcohol Intake

Senbang Yao
The Second Affiliated Hospital of Anhui Medical University
Xiangxiang Yin
The Second Affiliated Hospital of Anhui Medical University
Tingting Chen
The Second Affiliated Hospital of Anhui Medical University
Wenjun Chen
Anhui Medical University
He Zuo
The Second Affiliated Hospital of Anhui Medical University
Ziran Bi
The Second Affiliated Hospital of Anhui Medical University
Xiuqing Zhang
The Second Affiliated Hospital of Anhui Medical University
Yanyan Jing
The Second Affiliated Hospital of Anhui Medical University
Lulian Pang
The Second Affiliated Hospital of Anhui Medical University
Huaidong Cheng (chd1975ay@126.com)
The Second Affiliated Hospital of Anhui Medical University

Research Article

Keywords: Diagnosis, prognostic, ALDH2, head and neck squamous carcinoma (HNSC)

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-914685/v1

Exploring ALDH2 expression and its Immune infiltration in HNSC and its prognositic correlation with gender or alcohol intake

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] and Huaidong Cheng[1,2]

[1]Department of Oncology, The Second Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China;

[2]Department of Oncology, Anhui Medical University, Hefei, Anhui, China

[3]Department of Oncology, Anhui Chest Hospital, Hefei, Anhui, China

†These authors contributed equally to this work.

***Correspondence to:**

Huaidong Cheng, Department of Oncology, The Second Affiliated Hospital of Anhui Medical University, Anhui Medical University, 678 Furong Road, Hefei, Anhui 230601, China.

Email：chd1975ay@126.com

Abstract

Aldehyde dehydrogenase 2 (ALDH2) point mutation ALDH2*2 is the common frequent human gene variant, espically in East Asians. However, nothing is known about their expression and mechanism of action in HNSC.This research tried to explore the clinical significance and immune characteristics of ALDH2 in HNSC. The receiver operating characteristic (ROC) curve was analysed to assessment the diagnostic value of ALDH2 expression. ALDH2 expression in normal tissues and HNSC tissues were evaluated by IHC.Also we analyzed ALDH2 gene expression in 4 HNSC cell lines. ALDH2 expression was significantly reduced in HNSC tissues ($p< 0.05$). HNSC patients with highly expressed ALDH2 have a better prognosis ($p< 0.05$). Then, GSEA analysis result pointed that these gene sets were connected with signaling pathways including JAK-STAT signaling pathway. We unexpectedly found a significant prognostic effect of ALDH2 in alcohol consumption and male HNSC patients. The correlation between ALDH2 expression and immunoinhibitors showed a effect for ALDH2 in modifying tumor immunology in HNSC and there may be a possible mechanism by which ALDH2 regulates functions of T cell in HNSC. We developed a nomogram prognostic model for HNSC patients. Moreover, low ALDH2 expression was a poor prognostic factor in male alcoholics of HNSC.

Keywords

Diagnosis; prognostic; ALDH2; head and neck squamous carcinoma (HNSC)

Introduction

Head and neck squamous carcinoma (HNSC) is the eighth common malignancy according to the information reported in Global Cancer Statistics 2021 [1]. The most promising way to reducing mortality is early cancer diagnosis, as early detection is correlated with more favorable prognosis for approximately all kinds of cancer[2]. Currently, there is lack of availability of accurate biomarkers for the detection of HNSC. With the development of technology in high-throughput sequencing, many important diagnostic genes are constantly being discovered. However, key

biomarkers that can be used to enhance the prognosis of patients with head and neck squamous cell carcinoma remain to be confirmed.

Acetaldehyde dehydrogenase 2 (ALDH2) is a main enzyme for acetaldehyde metabolism during alcohol metabolism,ALDH2 is acknowledged 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 ALDH2 gene defect is correlated with an increased risk of hepatocellular carcinoma in patients with hepatitis B cirrhosis due to excessive alcohol consumption [3].The results of Jin S show that ALDH2 plays the key role of cancer suppressor by sustaining the stability of the liver genome, and the common human ALDH2 mutation may be an important risk factor for liver cancer[4]. A recent research found a significant positive dose response correlation between DFS and pre-treatment drinking history in HNSC patients with ALDH2 Glu/Glu[5]. However, the function of ALDH2 in HNSC remains unclear.

In this research, we explored ALDH2 expression in numerous neoplasms by using The Cancer Genome Atlas (TCGA) and its association with HNSC patient prognosis. Gene set enrichment analysis (GSEA) was applied to further evaluate the biological functions of the ALDH2 regulatory network correlated with HNSC pathogenesis. Since infiltration of immune cells is vital for the prognosis of HNSC patients, we also analysed the connection between expression of ALDH2 and the immune cells infiltration score. Our research provides new insights into the function of ALDH2 in head and neck squamous cell carcinoma.

Results

Patient characteristics

Gene sequence expression data of 502 HNSC cases were downloaded from the TCGA website(https://portal.gdc.cancer.gov/). Clinical characteristics data of 502 HNSC cases was shown in Table 1. The group of age≤60 accounted for the half proportion (50.00%), followed by the age > 60 group (50.00%). There were 33 cases (6.80%) in group T1, 144 cases (29.50%) in group T2, 131 cases (26.90%) in group T3 and 179 cases (36.8%) in group T4. N0 type cases were the majority (49.80%, N=239), N1 type 16.70% [80], N2 type 32.10% [154] and N3 type 1.40% [7]. Stage I HNSC cases accounted for 3.90% [19], Stage II for 19.50% [95], Stage III for 20.90% [102], and Stage IV for 55.80% [272]. Out of the 502 HNSC patinets, 5 (1.0%) cases had distant metastasis. 134 females (26.60%) and 368 (73.40%) males were composited for the composition of gender.

Table 1 TCGA-HNSC patient characteristics

Characteristic	levels	Low expression of ALDH2	High expression of ALDH2	*p*
n		251	251	
Gender, n (%)	Female	63 (12.5%)	71 (14.1%)	0.480
	Male	188 (37.5%)	180 (35.9%)	
Age, n (%)	<=60	124 (24.8%)	121 (24.2%)	0.893

	>60	127 (25.3%)	129 (25.7%)	
T stage, n (%)	T1	13 (2.7%)	20 (4.1%)	0.092
	T2	63 (12.9%)	81 (16.6%)	
	T3	75 (15.4%)	56 (11.5%)	
	T4	90 (18.5%)	89 (18.3%)	
N stage, n (%)	N0	106 (22.1%)	133 (27.7%)	0.042
	N1	38 (7.9%)	42 (8.8%)	
	N2	85 (17.7%)	69 (14.4%)	
	N3	6 (1.2%)	1 (0.2%)	
M stage, n (%)	M0	234 (49.1%)	238 (49.9%)	1.000
	M1	2 (0.4%)	3 (0.6%)	
Clinical stage, n (%)	Stage I	7 (1.4%)	12 (2.5%)	0.403
	Stage II	42 (8.6%)	53 (10.9%)	
	Stage III	54 (11.1%)	48 (9.8%)	
	Stage IV	138 (28.3%)	134 (27.5%)	
Alcohol history, n (%)	No	70 (14.3%)	88 (17.9%)	0.094
	Yes	176 (35.8%)	157 (32%)	
Age, meidan (IQR)		61 (52, 69)	61 (55, 69)	0.362

The expression of the ALDH2 in transcriptional level and its diagnostic function

ALDH2 expression from TCGA pan-cancer data were analyzed. As **Figure 1A** shows, compared with normal tissues, the expression level of ALDH2 in cancer tissues was significantly reduced. In order to verify the expression of ALDH2 in HNSC patients, we used a total of 3 related GEO data sets. And they generated sequencing results with similar ALDH2 expression difference in TCGA-HNSC (**Figure 1B-D**).ALDH2 expression differences between non-paired samples were statistically significant as shown in **Figure 1E**. Among 36 pairs of matched tissues, the expression of ALDH2 in tumor tissues and paraneoplastic tissues is also significantly different (**Figure 1F**). In order to evaluate the diagnostic efficacy of ALDH2, we performed ROC curve analysis on the expression data from tumor and normal tissues. The area under the ROC curve was 0.833 [95% confidence interval (CI): 0.793–0.872] (**Figure 1G**).

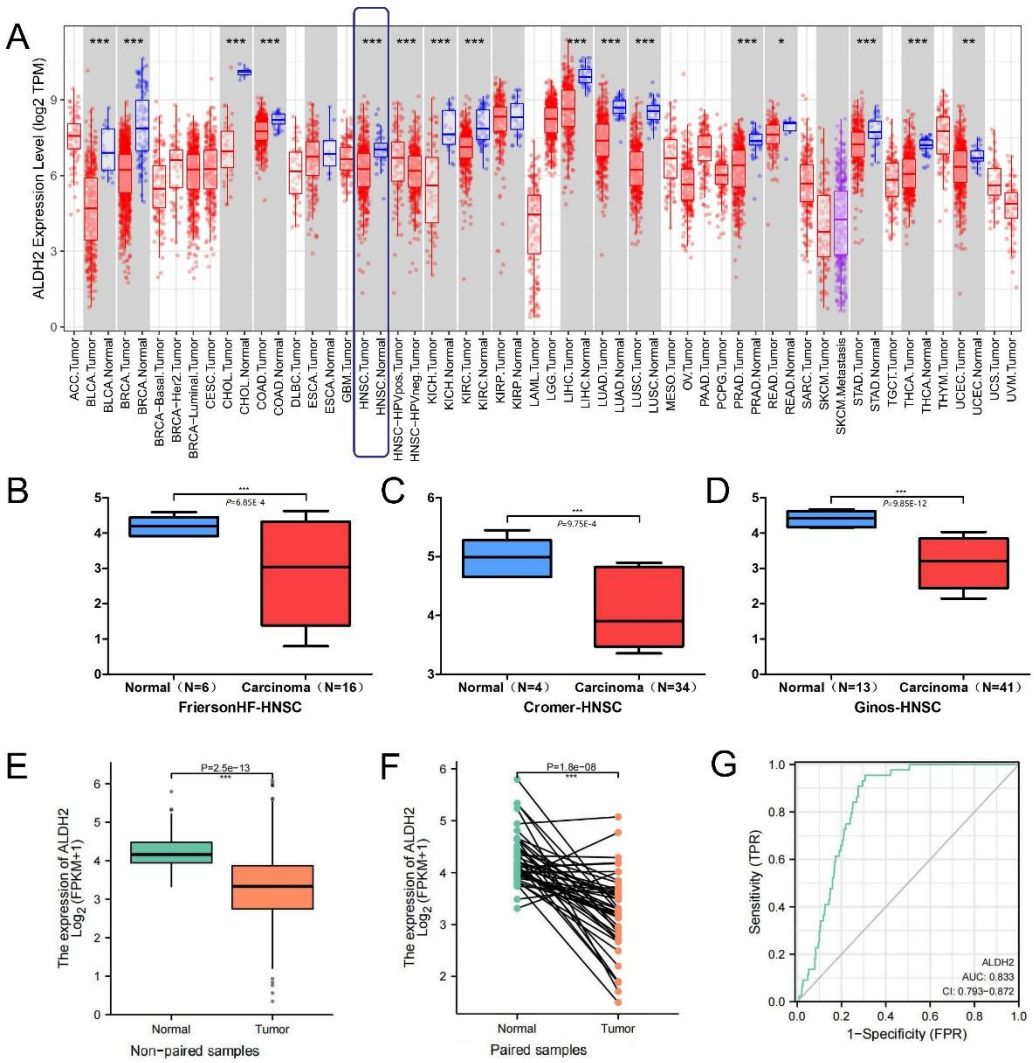


Figure 1 | The expression of the ALDH2 in transcriptional level and its diagnostic value. **(A)** TCGA pan-cancer data shows the expression of ALDH2 in tumors and normal tissues. **(B-D)** The expression of ALDH2 was signifcantly reduced in 91 cases of HNSC than 23 normal tissues in the three datasets (FriersonHF-HNSD, Cromer-HNSC, Ginos-HNSC). **(E)** ALDH2 expression in unpaired normal and tumor tissues of HNSC from TCGA-GTEx. **(F)** ALDH2 expression in paired normal and tumor tissues of HNSC from TCGA. **(G)** ROC curves of ALDH2 expression in normal and tumor tissues of HNSC.

The expression of the ALDH2 in tissues and cell lines

The immunohistochemistry (IHC) staining results showed the different expression of ALDH2 in the HNSC tissues. The staining intensity and quantity were significantly down-regulated in HNSC. (**Figure 2A**). We also examined ALDH2 alteration frequency and result shown at **Figure 2B**. It should be noted that HNSC cell lines differed in ALDH2 expression. ALDH2 showed the highest expression in the CAL33 HNSC cell. YD15 had the lowest expression values of all cell lines(**Figure 2C**).

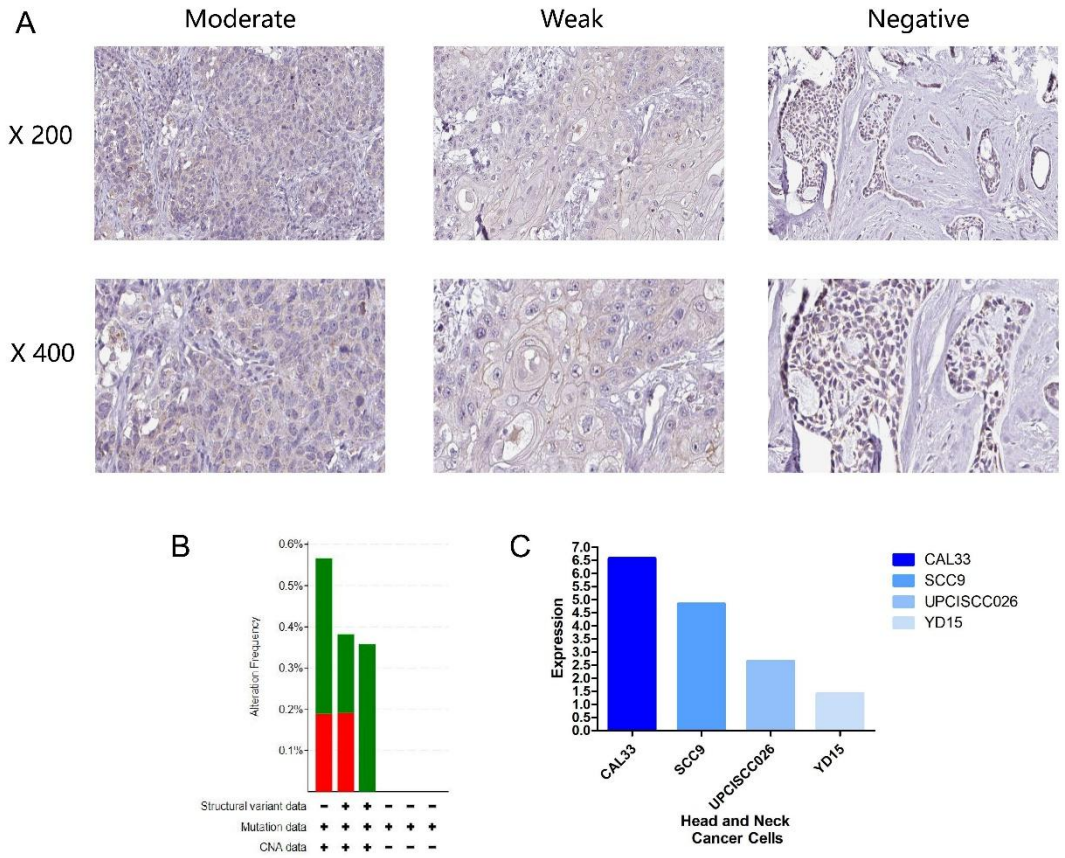


Figure 2 | ALDH2 expression in tissues and cell lines. **(A)** ALDH2 protein expression staining results: upper, ×200; lower, ×400， suggested that ALDH2 protein expression reduced in tumor tissues. **(B)** ALDH2 mutation frequencies across HNSC as obtained from cBioPortal. **(C)** Relative expression of ALDH2 in different HNSC cell lines.

ALDH2 expression and clinicopathological parameters

Then, the clinicopathological data of 502 HNSC patients from TCGA were analyzed, including clinical stage, age at diagnosis, alcohol history, race, residual tumor, as well as primary tumor, node, and metastasis classification(TNM). As shown in **Figure 3A-H**, decrease of ALDH2 expression was significantly correlated with staging and TNM staging.

The univariate logistic regression confirmed that expression of ALDH2 was characterized as a classified dependent variable with better prognosis of clinicopathology(**Table 2**). Decreasing HNSC ALDH2 expression was correlated with T stage (T2 vs. T1 OR =0.836, T3 vs. T1 OR =0.485, T4 vs. T1 OR =0.643).

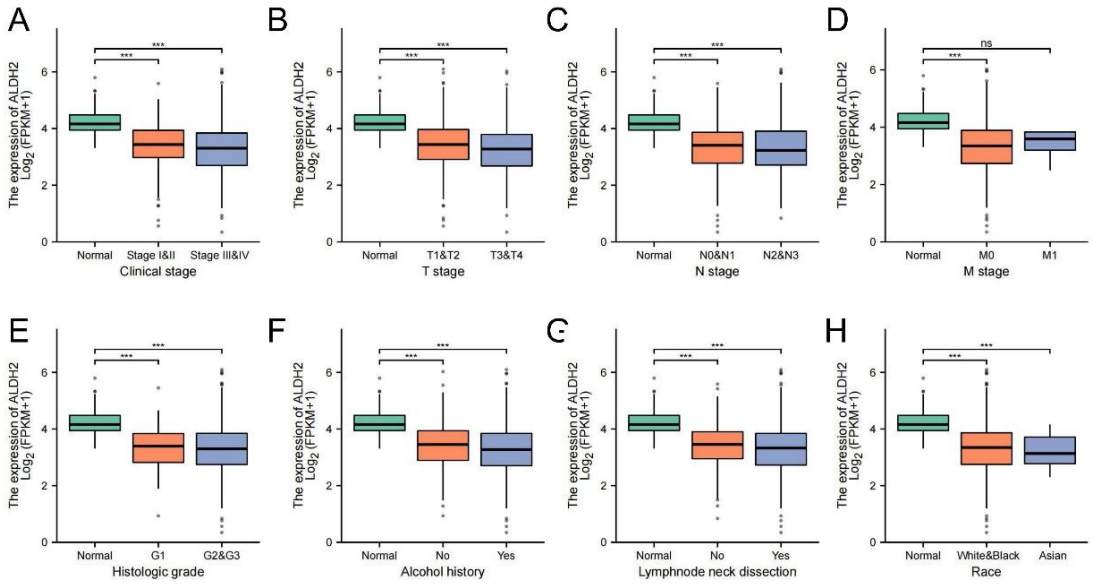


Figure 3 | Association between clinicopathological characteristics and ALDH2 expression. **(A)** Clinical stage. **(B)** T stage. **(C)** N stage. **(D)** M stage. **(E)** Histological grade. **(E)** Alcohol history. **(E)** Lymphnode dissection. **(E)** Race.

Table 2 Logistic regression of clinical pathological characteristics and ALDH2 expression

Clinical Characteristics	Total(N)	OR	95% CI	*P* value
Clinical stage				
Stage II vs. Stage I	114	0.736	0.254-1.996	0.555
Stage III vs. Stage I	121	0.519	0.180-1.396	0.202
Stage IV vs. Stage I	291	0.566	0.205-1.452	0.247
T stage				
T2 vs. T1	177	0.836	0.379-1.793	0.649
T3 vs. T1	164	0.485	0.218-1.048	0.069
T4 vs. T1	212	0.643	0.295-1.358	0.253
N stage				
N1 vs. N0	319	0.881	0.530-1.467	0.624
N2 vs. N0	393	0.647	0.430-0.971	0.036
N3 vs. N0	246	0.133	0.007-0.794	0.064
M stage				

	M1 vs. M0	477	1.475	0.242-11.273	0.672
Gender					
	Male vs. Female	502	0.850	0.571-1.262	0.420
Age					
	>65 vs <=65	501	1.041	0.733-1.478)	0.822

Network analysis of the differentially expressed genes correlated with ALDH2 in HNSC

The mRNA sequencing data of 502 HNSC patients from TCGA was analyzed by function module of LinkedOmics. As shown in volcano plot (**Figure 4A**), red dots indicate a significant positive correlation with ALDH2 gene, whereas green dots represent genes showed a significant negative correlation (false discovery rate [FDR] < 0.01). Interacting genes network was predicted by using GENEMANIA（**Figure 4B**）. The heat map shows 50 gene sets that are significantly positively or negatively correlated with ALDH2 (**Figure 4C**).

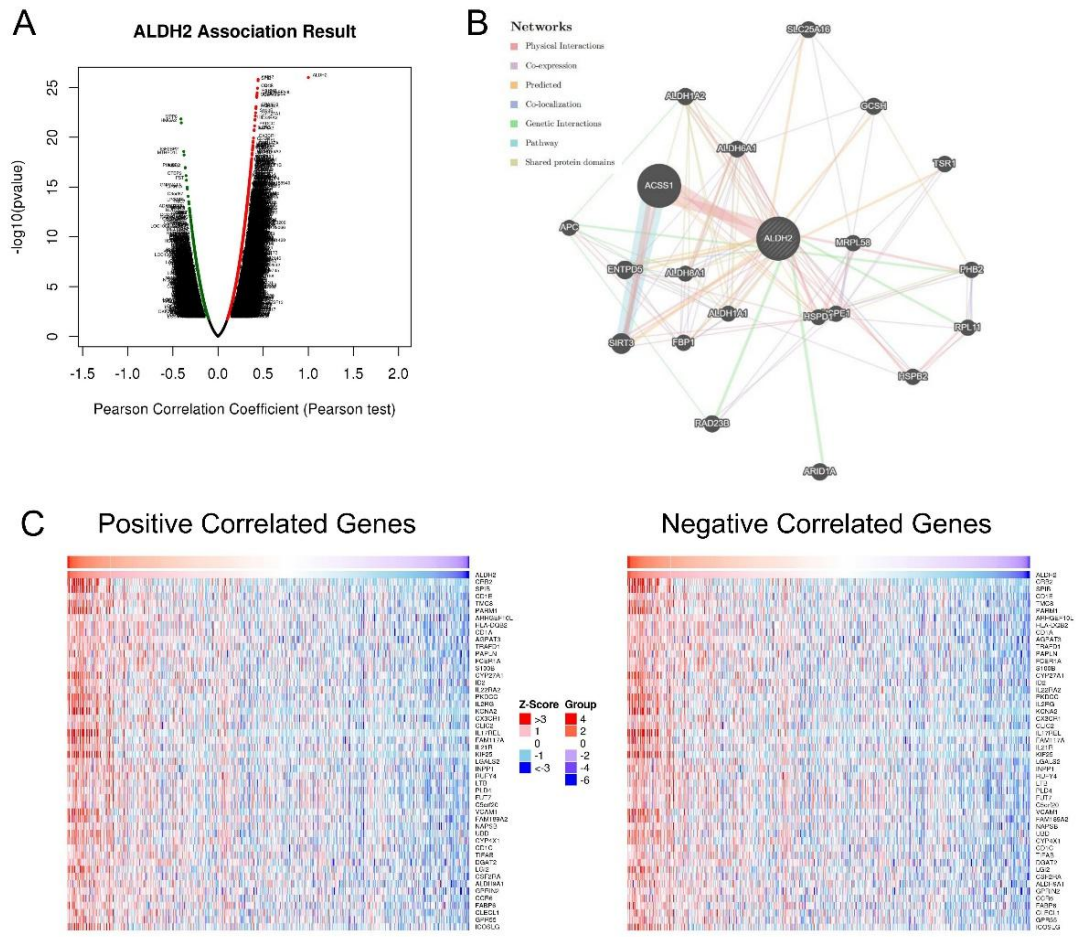


Figure 4 | Network analysis of the differentially expressed genes correlated with ALDH2 in HNSC. **(A)** The volcano plot was used to display genes that are highly correlated with ALDH2 in HNSC. Green and red dots represent negatively and positively significantly correlated genes with

ALDH2, respectively. **(B)** The main ALDH2 gene cluster in GeneMania analysis results. **(C)** Heatmaps showing the top 50 genes in HNSC that are positively or negatively related to ALDH2.

Correlation and Enrichment Analyses

The R software clusterProfiler software package is used to analyze highly correlated gene groups to explore possible functional pathways. GO functional enrichment analysis pointed that ALDH2 was correlated primarily with immune cells proliferation-related pathways including T cells proliferation and regulation of leukocyte proliferation (**Figure 5A-C**).GSEA was used to search the Reactome pathway and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The KEGG results indicated that JAK-STAT signaling pathway, Transcriptional misregulation in cancer, and MicroRNAs in cancer were enriched significantly (**Figure 5D**). The Reactome pathways analysis discovered significant enrichment in VEGFA-VEGFR2 Pathway, Death Receptor Signalling, and Adaptive Immune System pathways (**Figure 5E**). Such results indicate that ALDH2 expression is connected with complicated oncogenic pathways hyperactivation in head and neck squamous carcinoma, especially signalings which correlated with cell proliferation and immune system.

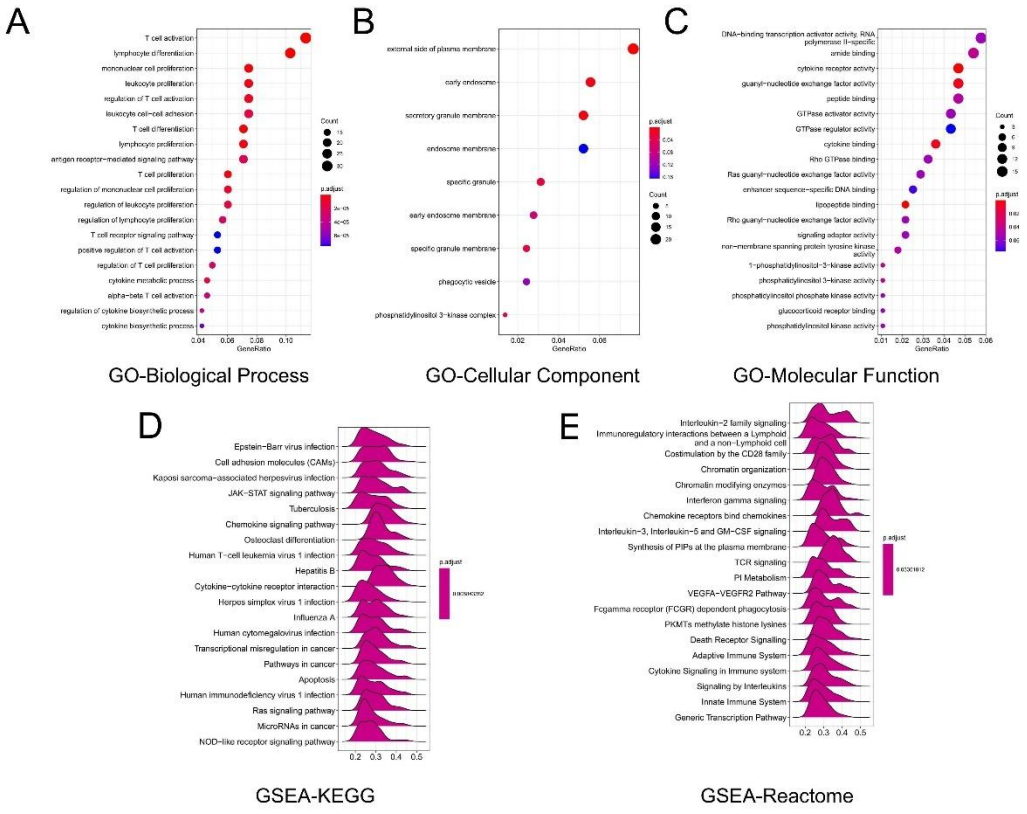


Figure 5 | Enrichment analysis of ALDH2 in HNSC. **(A)** Significant Gene Ontology iterms of 300 top genes most related to ALDH2 in biological processes. **(B)** Significant Gene Ontology iterms of 300 top genes most related to ALDH2 in cell component. **(C)** Significant Gene Ontology iterms of 300 top genes most related to ALDH2 in molecular function. **(D)** Significant gene set enrichment analysis (GSEA) analysis of ALDH2 in KEGG pathways. **(E)** Significant gene set enrichment analysis (GSEA) analysis of ALDH2 in Reactome pathways.

ALDH2 Expression is Correlated With Immune Infiltration Level in HNSC

Tumor-inundated lymphocyte is independent predictor of the survival and state of cancer outpost lymph nodes[6]. Then, we explored whether ALDH2 was correlated with immune infiltration level in head and neck squamous carcinoma. The expression of ALDH2 has significant positive correlations with infiltrating level of CD8+ T cell (r = 0.207, P = 5.27e-06), CD4+ T cell (r = 0.302, P = 1.37e-11), macrophages (r =0.18, P = 7.26e-05), neutrophils (r = 0.238, P = 1.35e-07) and DCs (r =0.304, P = 8.50e-12) in HNSC **(Figure 6A-B)**. In addition to this, we investigated immunoinhibitors that might be regulated by ALDH2**(Figure 6C)**. The expression of ALDH2 also has significant positive correlation with immunoinhibitors of CD244 (r = 0.331, P = 1.14e-14), CD96 (r = 0.282, P = 6.15e-11) and TIGIT (r =0.292, P = 1.21e-11) in HNSC**(Figure 6D-F)**.

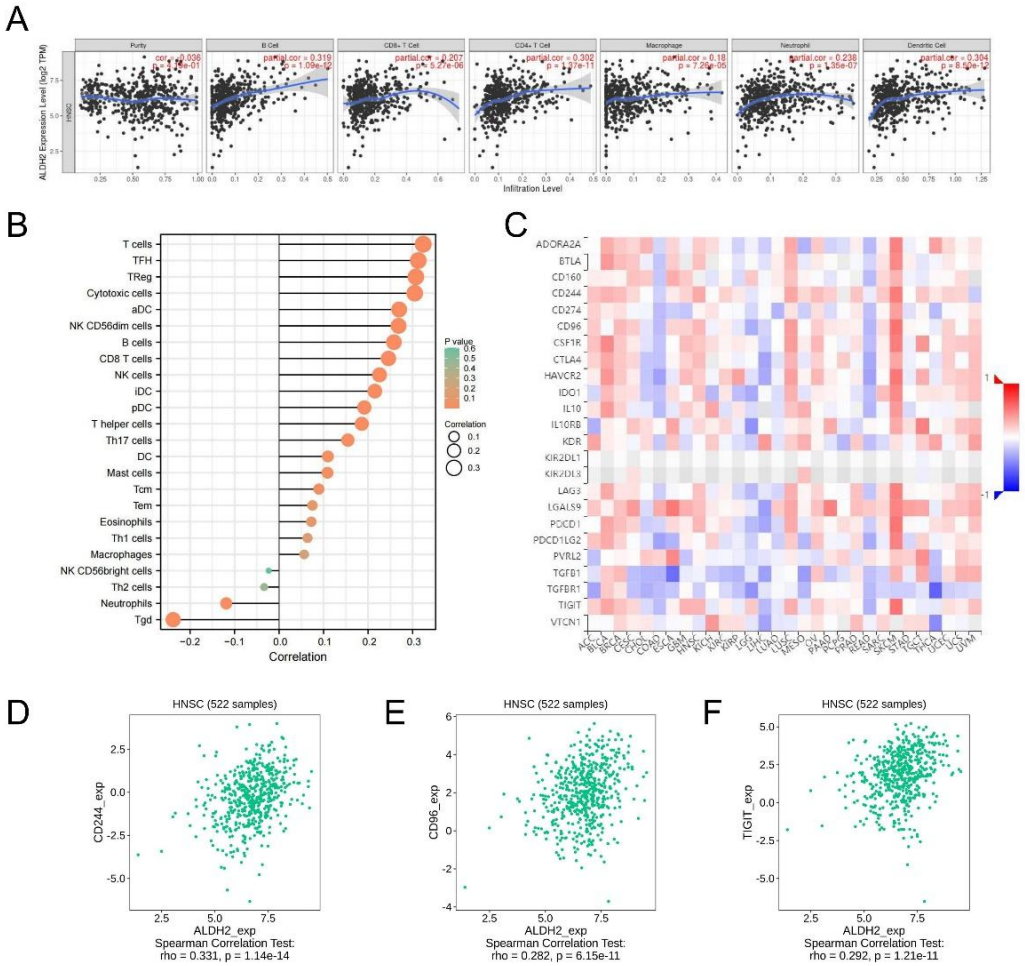


Figure 6 | ALDH2 Expression is Correlated With Level of Immune Infiltration in HNSC. **(A)** Correlation of immune infiltration cell with ALDH2 expression gained from TIMER analysed by purity-corrected Spearman test. **(B)** Immune infiltration levels of ALDH2 shown as lollipop diagram. **(C)** Immunoinhibitors that might be regulated by ALDH2. **(D-F)** The expression level of ALDH2 has significant positive correlations with immunoinhibitors(CD244, CD96 and TIGIT).

Survivability analysis suggested prognostic significance of ALDH2

Disease specific survival, Overall survival, and progression free survival were analyzed using Kaplan-Meier method. Interestingly, high ALDH2 expression correlates with well prognosis of HNSC patients（**Figure 7A-C**）. Also a multivariable cox proportional hazard model was shown in

Figure 7D. The results also showed age（$P < 0.01$）and stage（$P < 0.05$） as important covariates in predicting the survival.

Surprisingly, our subgroup analysis of overall survival showed significant difference in overall survival between the alcohol history and gender （**Figure 8A-D**）. It underlines the importance given to ALDH2 in predicting outcomes of male patient who is regular drinker. Finally, we established an nomogram combining prognostic information from clinicopathologic data and ALDH2 expression to predict the prognosis of gastric cancer patients（**Figure 8E**）. It can be seen in the figure that the level of ALDH2 expression has important implications for survival prediction of gastric cancer patients.

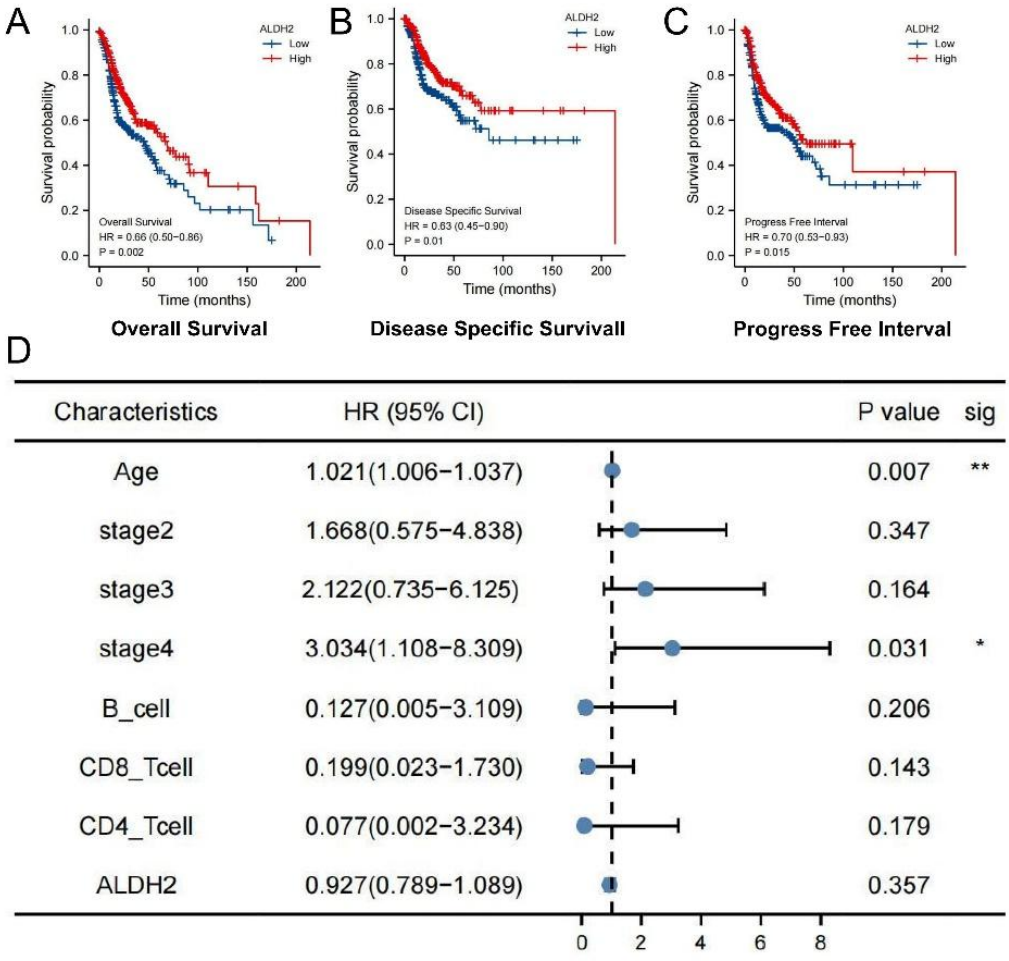


Figure 7 | Survivability analysis suggested prognostic significance of ALDH2. **(A)** The correlation between ALDH2 expression and the overall survival of HNSC. **(B)** The correlation between ALDH2 expression and the disease specific survival of HNSC. **(C)** The correlation between ALDH2 expression and the progress free interval of HNSC. **(D)** Multivariable cox proportional hazard analysis of ALDH2 expression in HNSC patients.

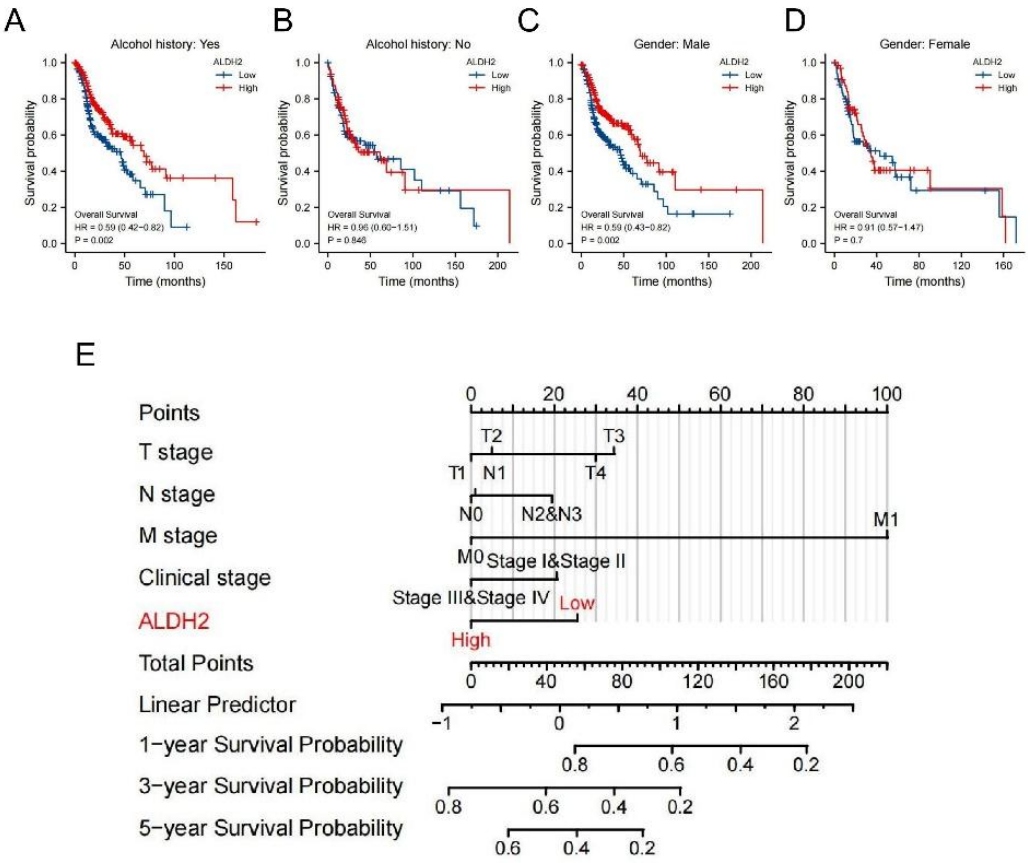


Figure 8 | ALDH2 has significant implications for HNSC patients. **(A-D)** ALDH2 expression had different effects of survilibity on drinking behaviors and gender. **(E)** Nomograms Predicting Survival in HNSC Patients.

Discussion

ALDH2, located on chromosome 12q24.12, belongs to the aldehyde dehydrogenase family of proteins[7].Although previous studies have shown the significant differences in ALDH2 genotype lead to different prognosis of cancer patients[8,9], its biological roles and prognostic value in HNSC has rarely been characterized. And to our knowledge, it is the first study to assess the influence of alcohol intake combined with ALDH2 expression on clinical survival in HNSC patients. We hope that our findings will contribute to existing knowledge, strengthen treatment design, and improve the correctness of the prognosis of HNSC patients.

With this research, we found that ALDH2 was down-regulated in HNSC cancer tissues and predicts a well prognosis significantly; also, high expression is associated with lower tumor stage. The result of univariate and multivariate cox analyses indicate that ALDH2 might be a prospective independent biomarker for prognosis of HNSC. And then we explored the regulators networks and co-expression genes of ALDH2. At last, a correlation analysis between immune signatures or immune infiltration and ALDH2 was conducted, result shows that ALDH2 was correlated to most immune marker genes, and its T cell infiltration may be one important factor for its ability of prognostic. The study we have done is intended to guide research of HNSC in the future.

In order to explore the mechanism of ALDH2 preventing the progression and occurrence of

HNSC, DEGs were screened out through correlation analysis, and functional annotation and pathways examination were performed. Then the genes network was established, the main gene clusters were discovered, and the functional annotation and pathways analysis of these genes were explored. Finally, we found that ALDH2 mainly affects the occurrence and progression of HNSC through the JAK-STAT signaling pathway (**Figure 5**).

JAK-STAT pathway was one of the most significant pathways in HNSC. In head and neck carcinoma, STATs can be initiated through a variety of signal transduction pathways, including epidermal growth factor receptor (EGFR), erythropoietin receptors, and interleukin (IL) receptor pathways[10]. Since ALDH2 was found to be related to the JAK-STAT pathway through functional enrichment analysis, we assumed that the possible mechanism of the differential expression and prognostic value of ALDH2 in HNSC is the crossover with the JAK-STAT pathway. Further experiments were needed to confirm the conclusions of this research. Furthermore, we also observed that the expression of ALDH2 is related to the Death Receptor Signalling pathway of GSEA. Previous studies have observed that death receptor signaling is related to antigen-independent drug resistance in leukemia by inducing CAR-T cell dysfunction[11]. Thus whether ALDH2 modulates HNSC progression via death receptor pathway remains further exploration.

Immune cells in the tumor microenvironment are important substances that regulate the progression behaviors of tumor cells [12-15]. Another significant feature of this research is the correlation between ALDH2 expression and diverse levels of immune infiltration in HNSC. Our results show that ALDH2 expression is related to the infiltration level of macrophage, neutrophil and CD8+ T cells, and is significantly related to the infiltration level of DCs, CD4+ T cells and B cells in HNSC (**Figure 6A-B**). In addition, the correlation between immunoinhibitors and ALDH2 expression suggests that ALDH2 regulates tumor immunity in HNSC. These correlations suggest the potential mechanism of ALDH2 regulating T cell function in HNSC. Therefore, ALDH2 plays a key role in the recruitment and regulation of HNSC immune infltrating cells.

Levels of alcohol drinking within a person are closely related to various types of cancer[16]. And ALDH2 disfunction will initiate numerous diseases, such as cardiovascular diseases and cancer[17]. Previous studies did not highlight the independent prognostic ability of ALDH2, which was confirmed in our work. We unexpectedly found a significant prognostic effect of ALDH2 in alcohol consumption and male HNSC patients. Cox analysis suggests that ALDH2 may be a potential independent biomarker for the prognosis of HNSC.

In summary, the medical and biological research of ALDH2 has received more and more attention[7]. Additionally, Alda-1, which can restoring activity of the ALDH2*2 enzyme, has the potential to reverse some of the ALDH2*2 population. Therefore, it may be used to improve the normal function of ALDH2, and improve prognosis of HNSC patients[18,19].This piolot study has some limits and this research is a retrospective study. The role and prognosis of the ALDH2 gene in HNSC require a prospective and functional study to provide more accurate information.

Conclusion

In conclusion, this is a first research describing the correlation between the prognosis of alcohol drinker and ALDH2 gene expression level. This research provided all-round evidence for the function of ALDH2 in the progression of HNSC and its potential as a prognostic predictor and biotherapy target. This finding can be used to forecast the survival prognosis of patients,

especially those alcoholics with HNSC.

Methods

TCGA RNA-sequencing data

A total of 502 HNSC cases and 44 normal samples are included in the gene sequence expression data. The detailed information of clinicopathological data was downloaded from TCGA data portal (https://www.cancer.gov/). The RNA-Seq gene expression data and clinicopathological data of 502 patients were processed and further analyzed (Table 1). This research was conducted under the provisions of the Declaration of Helsinki (revised in 2013).

Differently Expressed Genes (DEGs) and Gene Functional Analysis

we performed pearson correlation analysis to explore the genes significantly correlated to the expression of ALDH2, and the differently expressed genes (DEGs) was retrieved by using heatmap package of R language. Clinical and mutation information for HNSC patients were regained from cbioPortal[20]. Gene Interactings were predicted by using GENEMANIA[21]. LinkedOmics includes multiple sets of data of all 32 cancer types from TCGA, which is an open portal site[22]. By using the"LinkFinder" function of LinkedOmics, we performed pearson test statistical analysis on ALDH2 co-expression, and the result was displayed in volcano map. The functional annotations of DEGs were performed by KEGG and Gene Ontology (GO). Gene signaling pathways were analyzed by using clusterProfiler and pathview packages of R language.

Immunohistochemistry and evaluation of immunostaining intensity

ALDH2 expression was analyzed by Human Protein Atals[23]. Immunostaining was performed using the Anti-ALDH2 antibody produced in rabbit HPA051065. The IHC staining was graded as high; medium; low; not detected. The IHC intensity was graded as strong; moderate; weak; negative. The IHC quantity intensity was graded as > 75%; 75%-25%; < 25%; none.

Statistical analysis

R (v.3.6.3) was used for all statistical analysis. The relationship between clinicopathological characteristics was assessed by logistic regression and wilcoxon signed rank test.
Clinicopathological characteristics associated with overall survival (OS) in HNSC patients from TCGA were analyzed by Kaplan-Meier methods and cox regression. The receiver operating characteristic (ROC) curve was analysed using Wilson's method. Univariate logistic regression was used to analyze the correlation between clinicopathological characteristics and ALDH2 expression. Univariate and multifactorial cox analysis was used to assess the effect of ALDH2 expression on survival with other clinicopathological characteristics (e.g., grading, staging, lymph node status, distant metastatic status, age) . ALDH2 expression cut-off value was set as the median, and HNSC patients were divided into two groups.

Author Contributions

The manuscript was organized and wrote by Senbang Yao and Huaidong Cheng. Senbang Yao produced and designed the figures. He Zuo contributed to the documentation of the manuscript. Ziran Bi and Xiuqing Zhang grammar edited the manuscript. The manuscript was revised by Lulian Pang, Yanyan Jing, and Xiangxiang Yin. The manuscript was reviewed by all authors and approved for publication.

Ethical Approval

All aspects of the work are the responsibility of the author to ensure that issues related to the

accuracy or completeness of any part of the work are properly investigated and resolved. This research was conducted under the provisions of the Declaration of Helsinki (revised in 2013). All data used in this research were publicly available and approval from the local Ethics Committee was not required.

Data availability

The Research Data belongs to TCGA and is available in https://www.cancer.gov/tcga

Acknowledgments

We would like to express our most sincere thanks to TCGA for the data support provided.

Funding

1 Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA Cancer J Clin* **71**, 7-33, doi:10.3322/caac.21654 (2021).

2 Gregory, M. *et al.* Targeted single molecule mutation detection with massively parallel sequencing. *Nucleic acids research* **44**, e22, doi:10.1093/nar/gkv915 (2016).

3 Seo, W. *et al.* ALDH2 deficiency promotes alcohol-associated liver cancer by activating oncogenic pathways via oxidized DNA-enriched extracellular vesicles. *Journal of hepatology* **71**, 1000-1011, doi:10.1016/j.jhep.2019.06.018 (2019).

4 Jin, S. *et al.* ALDH2(E487K) mutation increases protein turnover and promotes murine hepatocarcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 9088-9093, doi:10.1073/pnas.1510757112 (2015).

5 Joo Kang, S., Shin, C., Sung, J. & Kim, N. Association Between ALDH2 Polymorphism and Gastric Cancer Risk in Terms of Alcohol Consumption: A Meta-Analysis. *Alcoholism, clinical experimental research* **45**, 6-14, doi:10.1111/acer.14508 (2021).

6 Okazaki, Y. *et al.* Prognostic Significance of the Immunological Indices in Patients Who Underwent Complete Resection of Pulmonary Metastases of Colorectal Cancer. *In vivo* **35**, 1091-1100, doi:10.21873/invivo.12354 (2021).

7 Wang, L. S. & Wu, Z. X. ALDH2 and Cancer Therapy. *Adv Exp Med Biol* **1193**, 221-228, doi:10.1007/978-981-13-6260-6_13 (2019).

8 Kawakita, D. *et al.* Prognostic Value of Drinking Status and Aldehyde Dehydrogenase 2 Polymorphism in Patients With Head and Neck Squamous Cell Carcinoma. *J Epidemiol* **26**, 292-299, doi:10.2188/jea.JE20140240 (2016).

9 Ji, Y. *et al.* ADH1B and ALDH2 polymorphisms and their associations with increased risk of squamous cell carcinoma of the head and neck in the Korean population. *Oral oncology* **47**, 583-587, doi:10.1016/j.oraloncology.2011.04.007 (2011).

10 Lai, S. & Johnson, F. Defining the role of the JAK-STAT pathway in head and neck and thoracic malignancies: implications for future therapeutic approaches. *Drug resistance updates : reviews commentaries in antimicrobial anticancer chemotherapy* **13**, 67-78, doi:10.1016/j.drup.2010.04.001 (2010).

11 Singh, N. *et al.* Impaired Death Receptor Signaling in Leukemia Causes Antigen-Independent Resistance by Inducing CAR T-cell Dysfunction. *Cancer discovery* **10**, 552-567, doi:10.1158/2159-8290.Cd-19-0813 (2020).

12 Xie, T. *et al.* PAFAH1B3 Expression Is Correlated With Gastric Cancer Cell Proliferation and Immune Infiltration. *Frontiers in oncology* **11**, 591545, doi:10.3389/fonc.2021.591545 (2021).

13 Vilariño, N., Bruna, J., Bosch-Barrera, J., Valiente, M. & Nadal, E. Immunotherapy in NSCLC patients with brain metastases. Understanding brain tumor microenvironment and dissecting outcomes from immune checkpoint blockade in the clinic. *Cancer treatment reviews* **89**, 102067, doi:10.1016/j.ctrv.2020.102067 (2020).

14 Sahin Ozkan, H., Ugurlu, M., Yumuk, P. & Kaya, H. Prognostic Role of Immune Markers in Triple Negative Breast Carcinoma. *Pathology oncology research : POR* **26**, 2733-2745, doi:10.1007/s12253-020-00874-4 (2020).

15 Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature medicine* **24**, 541-550, doi:10.1038/s41591-018-0014-x (2018).

16 Clarke, T. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Molecular psychiatry* **22**, 1376-1384, doi:10.1038/mp.2017.153 (2017).

17 Hu, Y. *et al.* A phenomics-based approach for the detection and interpretation of shared genetic influences on 29 biochemical indices in southern Chinese men. *BMC genomics* **20**, 983, doi:10.1186/s12864-019-6363-0 (2019).

18 Chen, C. *et al.* Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. *Science* **321**, 1493-1495, doi:10.1126/science.1158554 (2008).

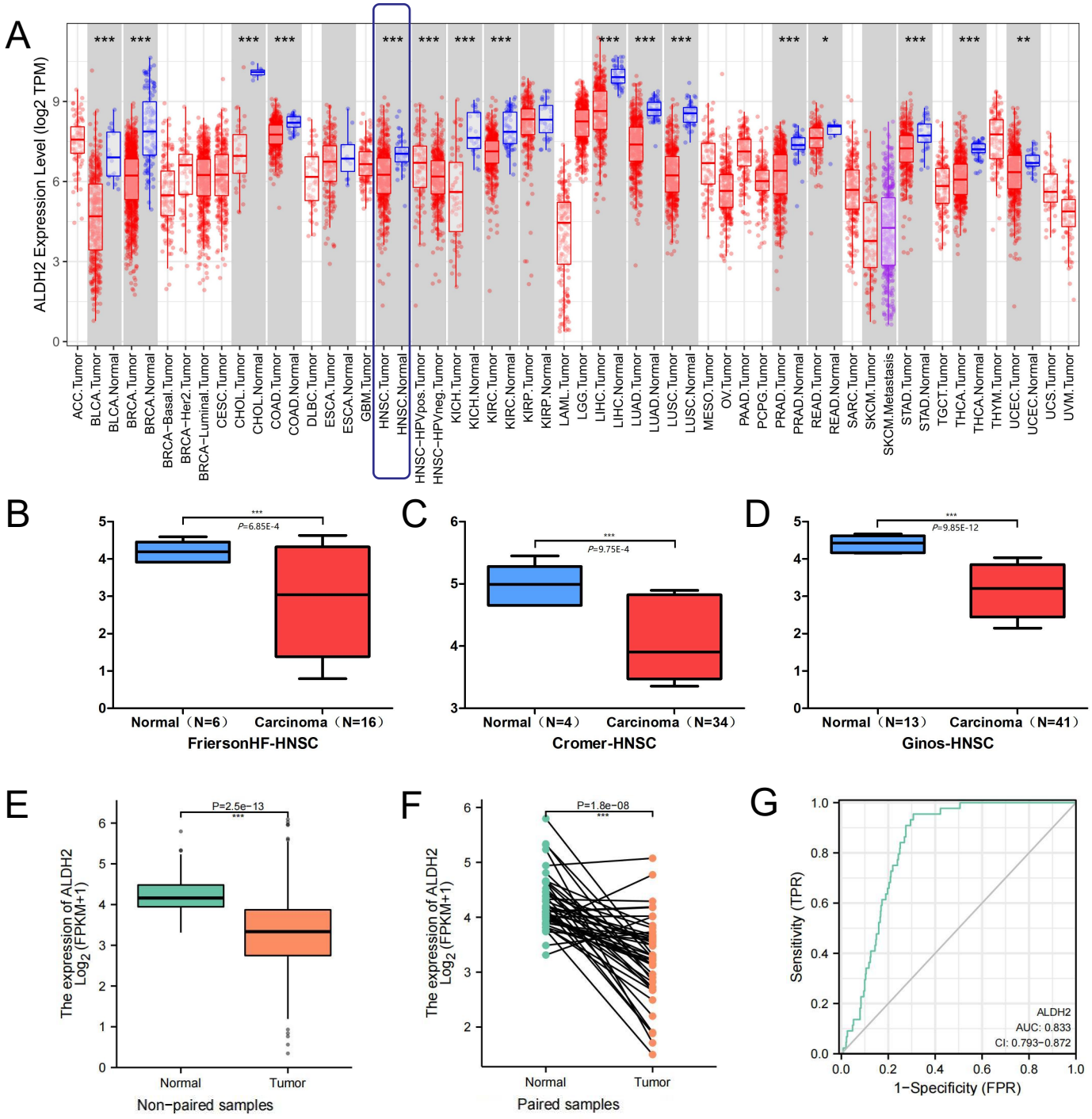
19 Li, R., Zhao, Z., Sun, M., Luo, J. & Xiao, Y. ALDH2 gene polymorphism in different types of cancers and its clinical significance. *Life sciences* **147**, 59-66, doi:10.1016/j.lfs.2016.01.028 (2016).

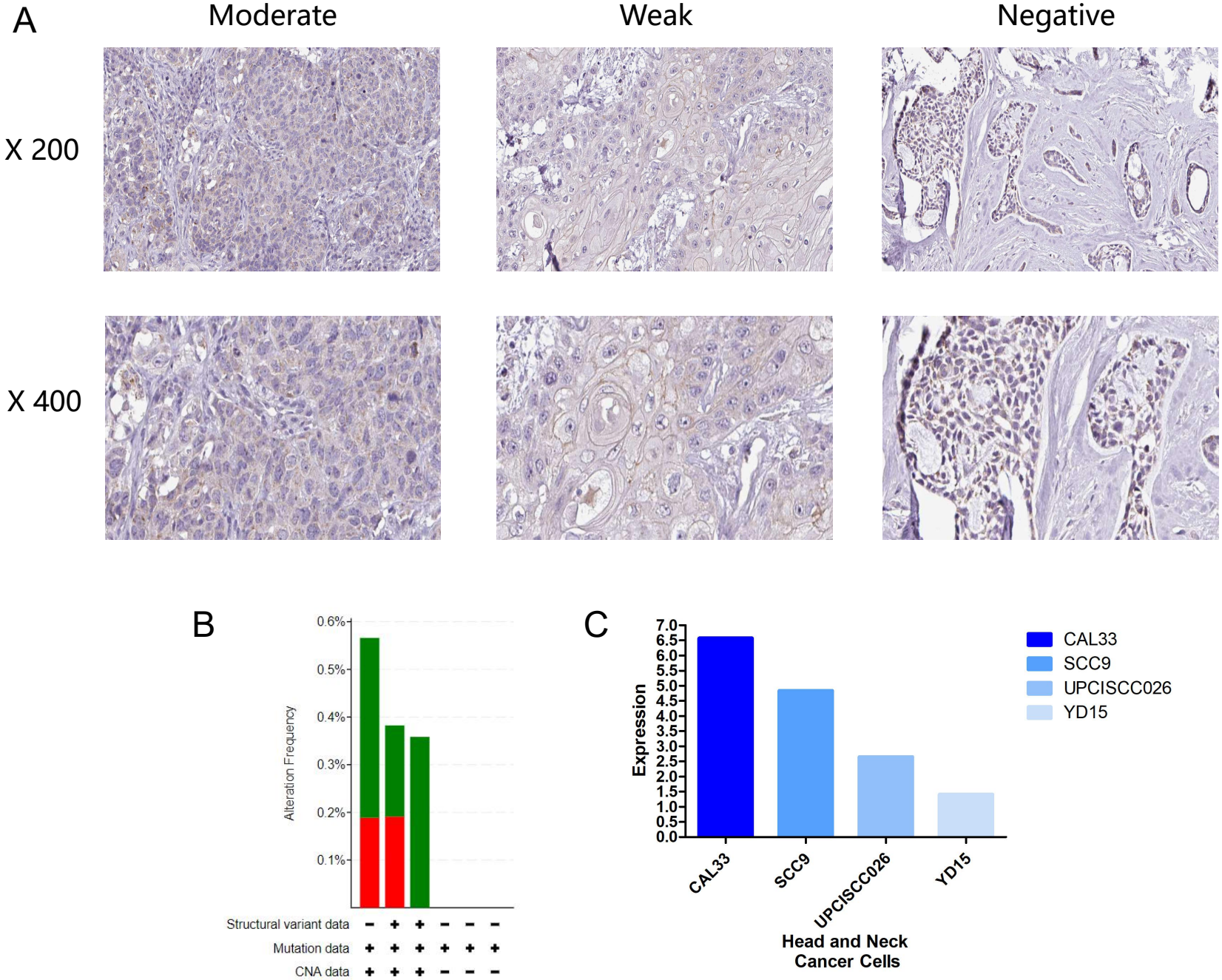
20 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404, doi:10.1158/2159-8290.Cd-12-0095 (2012).

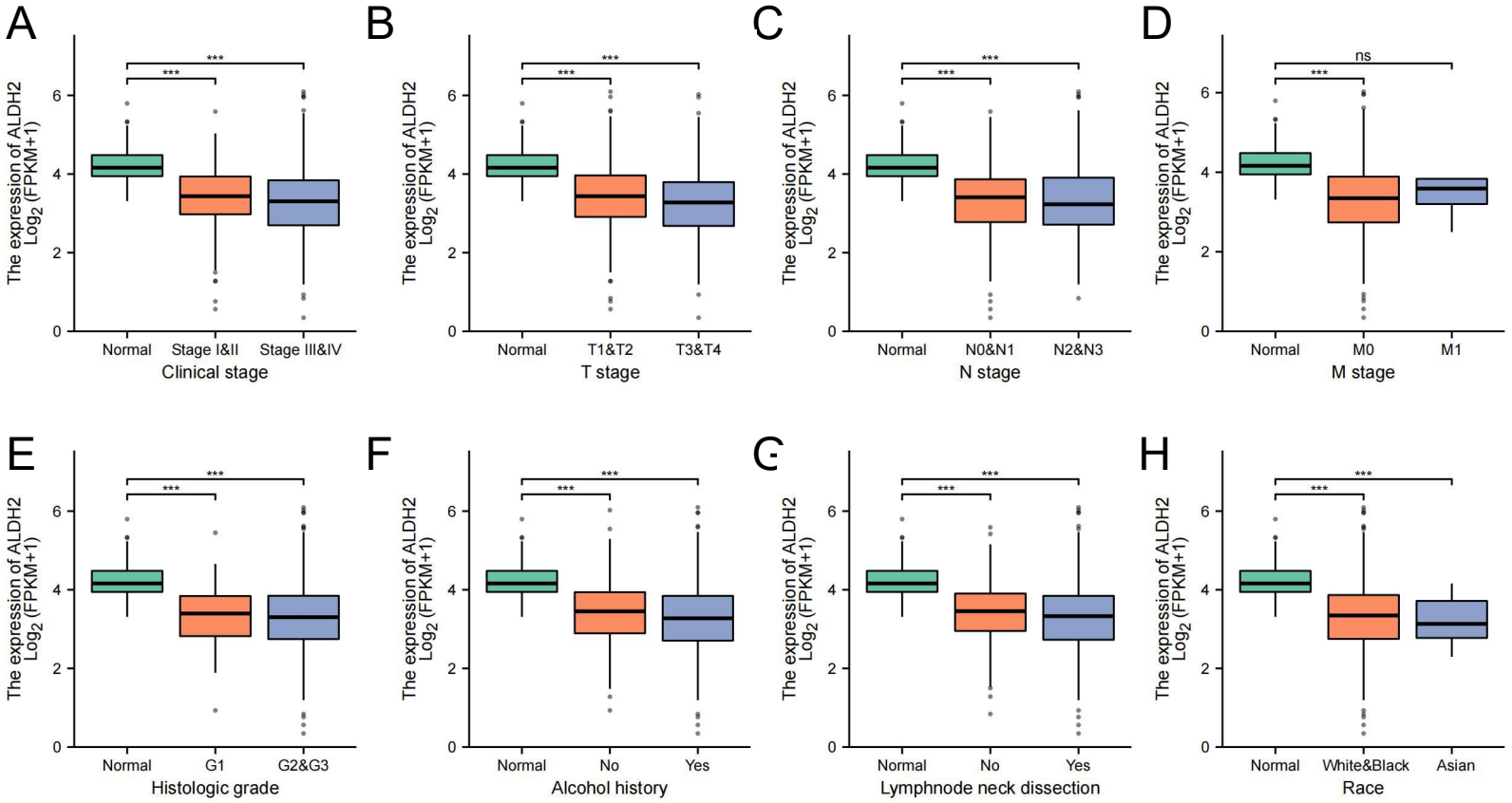
21 Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research* **38**, W214-220, doi:10.1093/nar/gkq537 (2010).

22 Vasaikar, S., Straub, P., Wang, J. & Zhang, B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic acids research* **46**, D956-D963, doi:10.1093/nar/gkx1090 (2018).

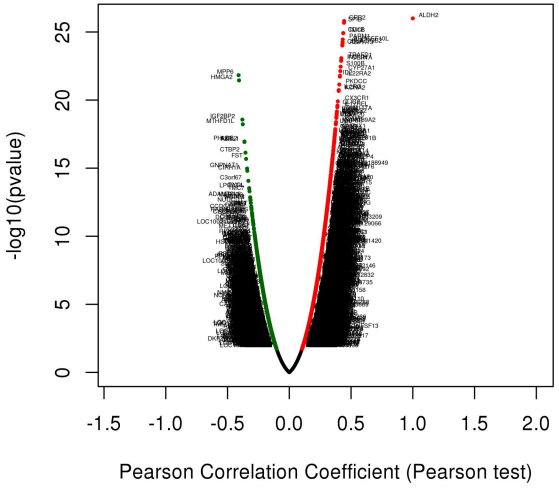
23 Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, doi:10.1126/science.aan2507 (2017).


A
ALDH2 Expression Level (log2 TPM)
ACC.Tumor
BLCA.Tumor
BLCA.Normal
BRCA.Tumor
BRCA.Normal
BRCA−Basal.Tumor
BRCA−Her2.Tumor
BRCA−Luminal.Tumor
CESC.Tumor
CHOL.Tumor
CHOL.Normal
COAD.Tumor
COAD.Normal
DLBC.Tumor
ESCA.Tumor
ESCA.Normal
GBM.Tumor
HNSC.Tumor
HNSC.Normal
HNSC−HPVpos.Tumor
HNSC−HPVneg.Tumor
KICH.Tumor
KICH.Normal
KIRC.Tumor
KIRC.Normal
KIRP.Tumor
KIRP.Normal
LAML.Tumor
LGG.Tumor
LIHC.Tumor
LIHC.Normal
LUAD.Tumor
LUAD.Normal
LUSC.Tumor
LUSC.Normal
MESO.Tumor
OV.Tumor
PAAD.Tumor
PCPG.Tumor
PRAD.Tumor
PRAD.Normal
READ.Tumor
READ.Normal
SARC.Tumor
SKCM.Tumor
SKCM.Metastasis
STAD.Tumor
STAD.Normal
TGCT.Tumor
THCA.Tumor
THCA.Normal
THYM.Tumor
UCEC.Tumor
UCEC.Normal
UCS.Tumor
UVM.Tumor
B
P=6.85E-4
Normal（N=6）
Carcinoma（N=16）
FriersonHF-HNSC
C
P=9.75E-4
Normal（N=4）
Carcinoma（N=34）
Cromer-HNSC
D
P=9.85E-12
Normal（N=13）
Carcinoma（N=41）
Ginos-HNSC
E
P=2.5e−13
The expression of ALDH2 Log2 (FPKM+1)
Normal
Tumor
Non−paired samples
F
P=1.8e−08
The expression of ALDH2 Log2 (FPKM+1)
Normal
Tumor
Paired samples
G
Sensitivity (TPR)
1−Specificity (FPR)
ALDH2
AUC: 0.833
CI: 0.793−0.872

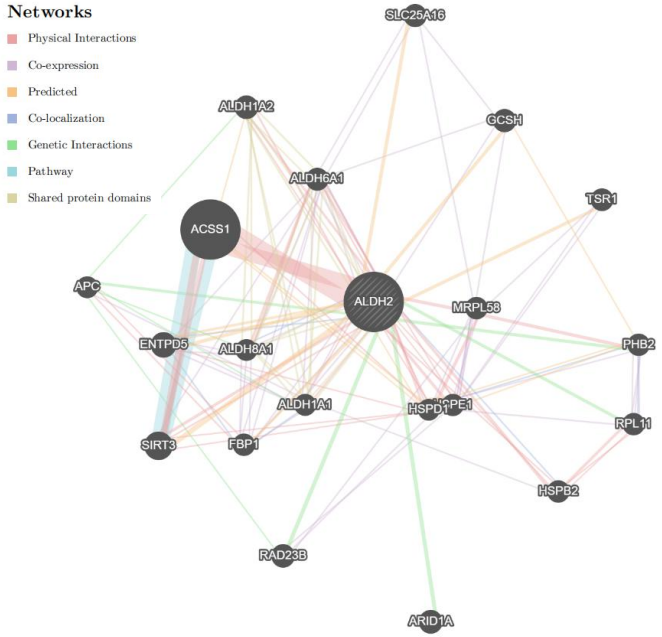

A
Moderate
Weak
Negative
X 200
X 400
B
Alteration Frequency
0.6%
0.5%
0.4%
0.3%
0.2%
0.1%
Structural variant data
Mutation data
CNA data
C
Expression
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
CAL33
SCC9
UPCISCC026
YD15
Head and Neck Cancer Cells


A
B
C
D
E
F
G
H
The expression of ALDH2
Log2 (FPKM+1)
***
ns
Normal
Stage I&II
Stage III&IV
Clinical stage
T1&T2
T3&T4
T stage
N0&N1
N2&N3
N stage
M0
M1
M stage
G1
G2&G3
Histologic grade
No
Yes
Alcohol history
Lymphnode neck dissection
White&Black
Asian
Race

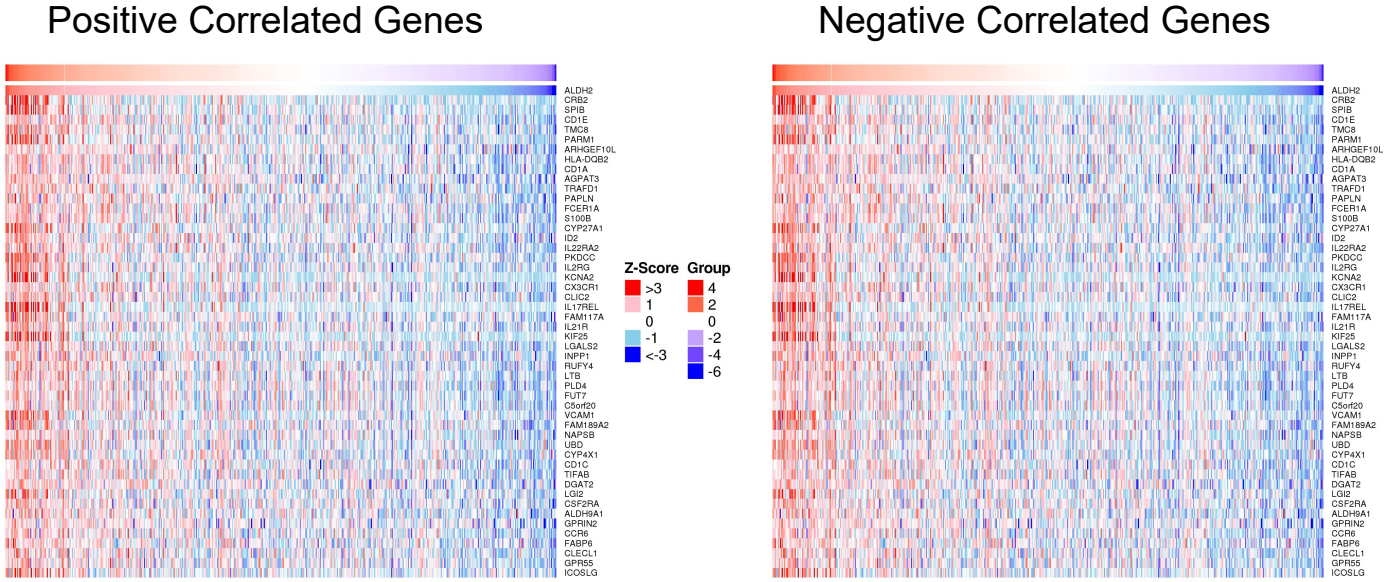
A

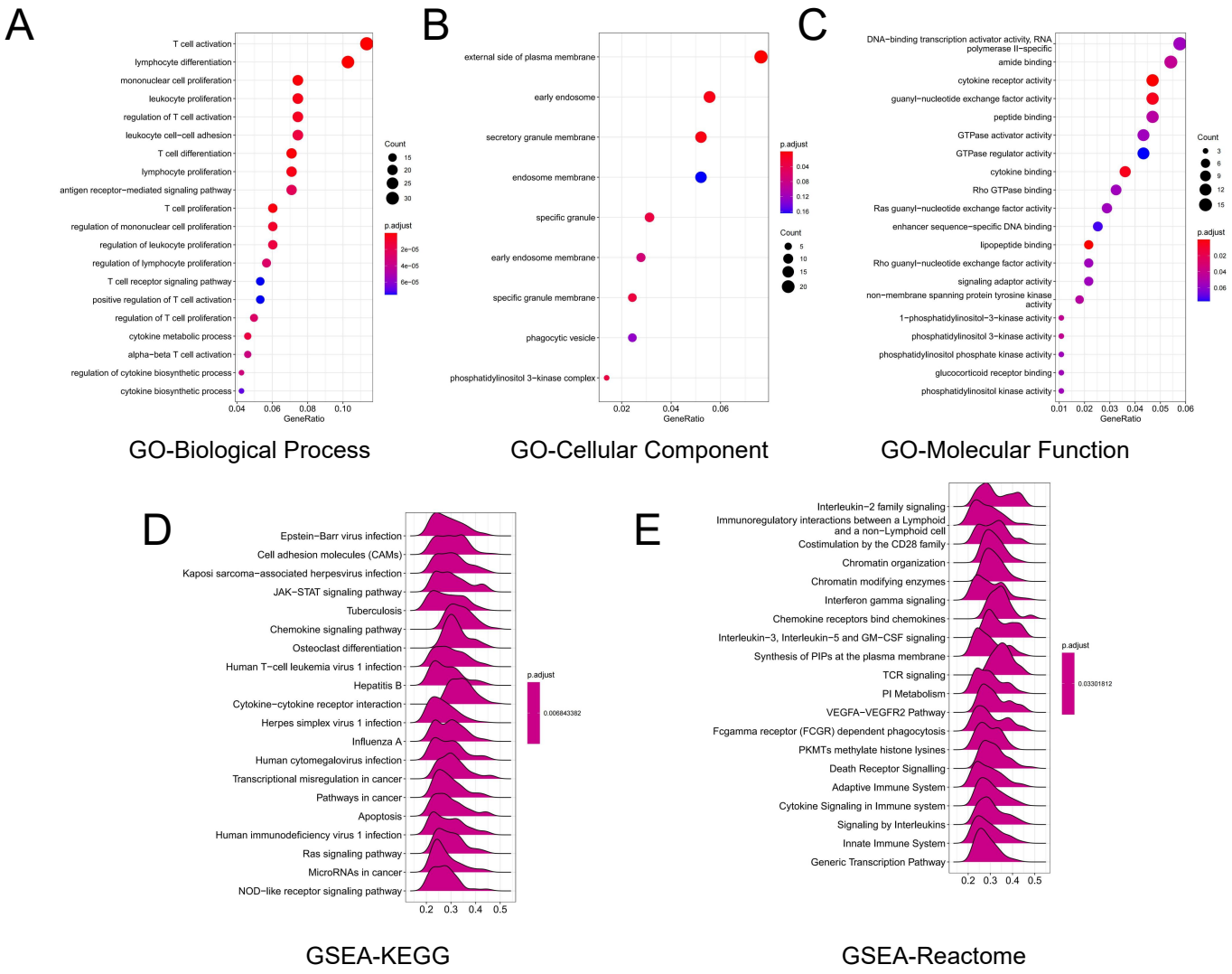

ALDH2 Association Result
-log10(pvalue)
0
5
10
15
20
25
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Pearson Correlation Coefficient (Pearson test)

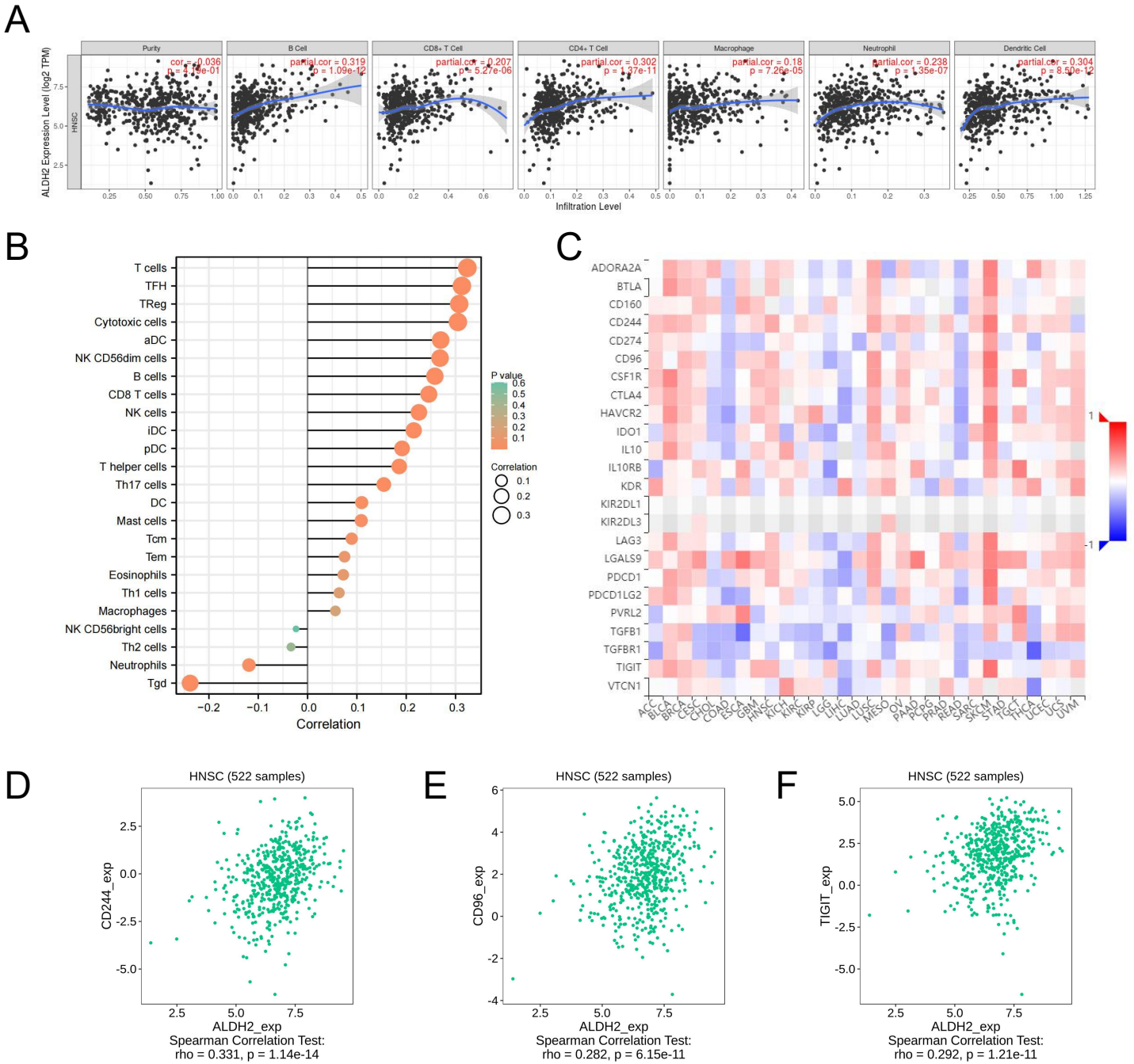
B


Networks
Physical Interactions
Co-expression
Predicted
Co-localization
Genetic Interactions
Pathway
Shared protein domains
SLC25A16
ALDH1A2
GCSH
ALDH6A1
TSR1
ACSS1
APC
ALDH2
MRPL58
ENTPD5
ALDH8A1
PHB2
ALDH1A1
HSPD1
PPE1
RPL11
SIRT3
FBP1
HSPB2
RAD23B
ARID1A

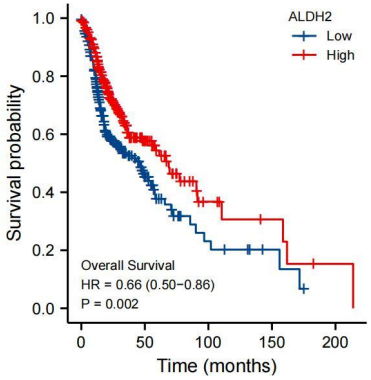
C


Positive Correlated Genes
Negative Correlated Genes
Z-Score
Group
>3
1
0
-1
<-3
4
2
0
-2
-4
-6
ALDH2
CRB2
SPIB
CD1E
TMC8
PARM1
ARHGEF10L
HLA-DQB2
CD1A
AGPAT3
TRAFD1
PAPLN
FCER1A
S100B
CYP27A1
ID2
IL22RA2
PKDCC
IL2RG
KCNA2
CX3CR1
CLIC2
IL17REL
FAM117A
IL21R
KIF25
LGALS2
INPP1
RUFY4
LTB
PLD4
FUT7
C5orf20
VCAM1
FAM189A2
NAPSB
USD
CYP4X1
CD1C
TIFAB
DGAT2
LGI2
CSF2RA
ALDH9A1
GPRIN2
CCR6
FABP6
CLECL1
GPR55
ICOSLG


A
T cell activation
lymphocyte differentiation
mononuclear cell proliferation
leukocyte proliferation
regulation of T cell activation
leukocyte cell−cell adhesion
T cell differentiation
lymphocyte proliferation
antigen receptor−mediated signaling pathway
T cell proliferation
regulation of mononuclear cell proliferation
regulation of leukocyte proliferation
regulation of lymphocyte proliferation
T cell receptor signaling pathway
positive regulation of T cell activation
regulation of T cell proliferation
cytokine metabolic process
alpha−beta T cell activation
regulation of cytokine biosynthetic process
cytokine biosynthetic process
0.04 0.06 0.08 0.10
GeneRatio
Count
15 20 25 30
p.adjust
2e−05 4e−05 6e−05
GO-Biological Process
B
external side of plasma membrane
early endosome
secretory granule membrane
endosome membrane
specific granule
early endosome membrane
specific granule membrane
phagocytic vesicle
phosphatidylinositol 3−kinase complex
0.02 0.04 0.06
GeneRatio
p.adjust
0.04 0.08 0.12 0.16
Count
5 10 15 20
GO-Cellular Component
C
DNA−binding transcription activator activity, RNA polymerase II−specific
amide binding
cytokine receptor activity
guanyl−nucleotide exchange factor activity
peptide binding
GTPase activator activity
GTPase regulator activity
cytokine binding
Rho GTPase binding
Ras guanyl−nucleotide exchange factor activity
enhancer sequence−specific DNA binding
lipopeptide binding
Rho guanyl−nucleotide exchange factor activity
signaling adaptor activity
non−membrane spanning protein tyrosine kinase activity
1−phosphatidylinositol−3−kinase activity
phosphatidylinositol 3−kinase activity
phosphatidylinositol phosphate kinase activity
glucocorticoid receptor binding
phosphatidylinositol kinase activity
0.01 0.02 0.03 0.04 0.05 0.06
GeneRatio
Count
3 6 9 12 15
p.adjust
0.02 0.04 0.06
GO-Molecular Function
D
Epstein−Barr virus infection
Cell adhesion molecules (CAMs)
Kaposi sarcoma−associated herpesvirus infection
JAK−STAT signaling pathway
Tuberculosis
Chemokine signaling pathway
Osteoclast differentiation
Human T−cell leukemia virus 1 infection
Hepatitis B
Cytokine−cytokine receptor interaction
Herpes simplex virus 1 infection
Influenza A
Human cytomegalovirus infection
Transcriptional misregulation in cancer
Pathways in cancer
Apoptosis
Human immunodeficiency virus 1 infection
Ras signaling pathway
MicroRNAs in cancer
NOD−like receptor signaling pathway
0.2 0.3 0.4 0.5
p.adjust
0.006843382
GSEA-KEGG
E
Interleukin−2 family signaling
Immunoregulatory interactions between a Lymphoid and a non−Lymphoid cell
Costimulation by the CD28 family
Chromatin organization
Chromatin modifying enzymes
Interferon gamma signaling
Chemokine receptors bind chemokines
Interleukin−3, Interleukin−5 and GM−CSF signaling
Synthesis of PIPs at the plasma membrane
TCR signaling
PI Metabolism
VEGFA−VEGFR2 Pathway
Fcgamma receptor (FCGR) dependent phagocytosis
PKMTs methylate histone lysines
Death Receptor Signalling
Adaptive Immune System
Cytokine Signaling in Immune system
Signaling by Interleukins
Innate Immune System
Generic Transcription Pathway
0.2 0.3 0.4 0.5
p.adjust
0.03301812
GSEA-Reactome

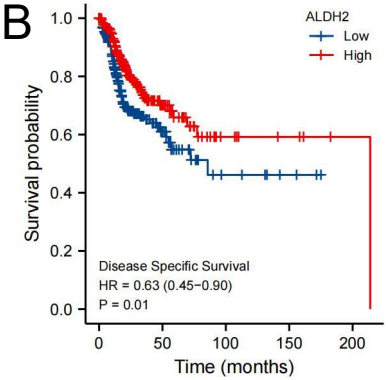

A
Purity
B Cell
CD8+ T Cell
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
cor = -0.036
p = 4.7e-01
partial.cor = 0.319
p = 1.09e-12
partial.cor = 0.207
p = 5.27e-06
partial.cor = 0.302
p = 1.27e-11
partial.cor = 0.18
p = 7.26e-05
partial.cor = 0.238
p = 1.35e-07
partial.cor = 0.304
p = 8.5e-12
HNSC
ALDH2 Expression Level (log2 TPM)
Infiltration Level
B
T cells
TFH
TReg
Cytotoxic cells
aDC
NK CD56dim cells
B cells
CD8 T cells
NK cells
iDC
pDC
T helper cells
Th17 cells
DC
Mast cells
Tcm
Tem
Eosinophils
Th1 cells
Macrophages
NK CD56bright cells
Th2 cells
Neutrophils
Tgd
P value
Correlation
Correlation
C
ADORA2A
BTLA
CD160
CD244
CD274
CD96
CSF1R
CTLA4
HAVCR2
IDO1
IL10
IL10RB
KDR
KIR2DL1
KIR2DL3
LAG3
LGALS9
PDCD1
PDCD1LG2
PVRL2
TGFB1
TGFBR1
TIGIT
VTCN1
D
HNSC (522 samples)
CD244_exp
ALDH2_exp
Spearman Correlation Test:
rho = 0.331, p = 1.14e-14
E
HNSC (522 samples)
CD96_exp
ALDH2_exp
Spearman Correlation Test:
rho = 0.282, p = 6.15e-11
F
HNSC (522 samples)
TIGIT_exp
ALDH2_exp
Spearman Correlation Test:
rho = 0.292, p = 1.21e-11

A



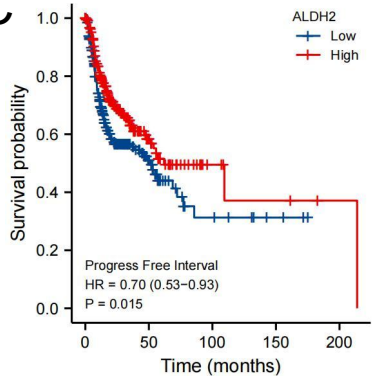
Overall Survival

B



Disease Specific Survivall

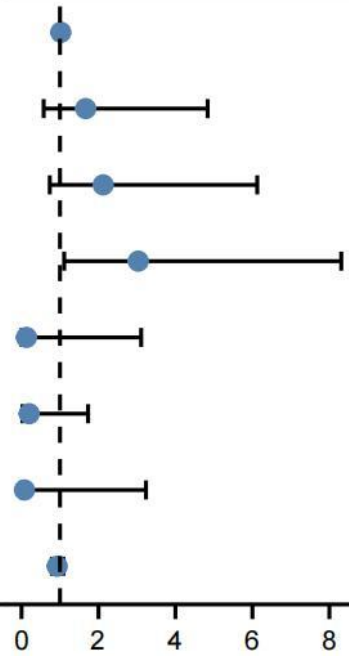
C

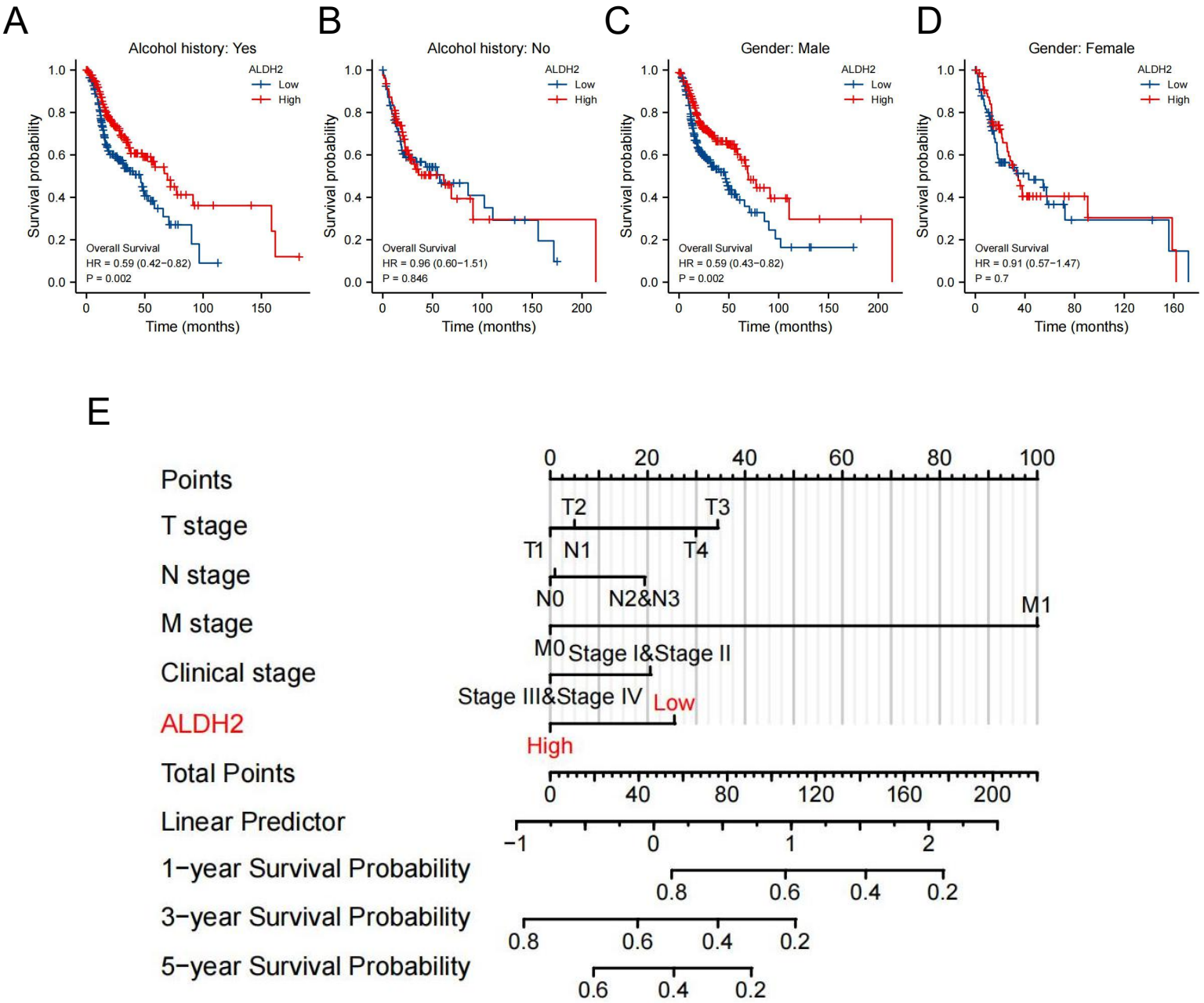


Progress Free Interval

D

Characteristics	HR (95% CI)	P value	sig
Age	1.021(1.006−1.037)	0.007	**
stage2	1.668(0.575−4.838)	0.347	
stage3	2.122(0.735−6.125)	0.164	
stage4	3.034(1.108−8.309)	0.031	*
B_cell	0.127(0.005−3.109)	0.206	
CD8_Tcell	0.199(0.023−1.730)	0.143	
CD4_Tcell	0.077(0.002−3.234)	0.179	
ALDH2	0.927(0.789−1.089)	0.357	




A
Alcohol history: Yes
ALDH2
Low
High
Survival probability
Overall Survival
HR = 0.59 (0.42−0.82)
P = 0.002
Time (months)
B
Alcohol history: No
Overall Survival
HR = 0.96 (0.60−1.51)
P = 0.846
C
Gender: Male
Overall Survival
HR = 0.59 (0.43−0.82)
P = 0.002
D
Gender: Female
Overall Survival
HR = 0.91 (0.57−1.47)
P = 0.7
E
Points
T stage
N stage
M stage
Clinical stage
ALDH2
Total Points
Linear Predictor
1−year Survival Probability
3−year Survival Probability
5−year Survival Probability
T1
T2
T3
T4
N0
N1
N2&N3
M0
M1
Stage I&Stage II
Stage III&Stage IV
Low
High