Singh *et al.*
Egyptian Journal of Medical Human Genetics (2023) 24:17
https://doi.org/10.1186/s43042-023-00401-5

Egyptian Journal of Medical Human Genetics

RESEARCH

Open Access

Understanding the role of adipokines and adipogenesis family in hepatocellular carcinoma

Prithvi Singh[1†], Rishabh Gurung[1†], Armiya Sultan[2] and Ravins Dohare[1*]

Abstract

Background Hepatocellular carcinoma (HCC) is the most common primary liver cancer. It has the sixth most incident cases with poor prognosis. Adipokines are known to have been linked with oncogenesis and progression of HCC.

Methods We extracted TCGA-HCC data and identified differentially expressed genes (DEGs) using R. Genes of adipokines and adipogenesis family were scrutinized from DEGs and expression of genes in normal versus tumor patients was studied. Prognostic and stage plot analyses were performed, and key genes were selected. Pathway and gene ontology (GO) enrichment analysis was conducted. Expression analysis based on nodal metastasis, tumor protein p53 (TP53) mutation and tumor grade, and mutation analysis was performed using UALCAN and cBioPortal. Tumor infiltration analysis was performed to study the correlation of gene expression with tumor-infiltrating immune cells.

Results We found four genes apelin (APLN), aldehyde dehydrogenase, mitochondrial (ALDH2), E2F transcription factor 1 (E2F1) and phosphoenolpyruvate carboxykinase, cytosolic (PCK1) highly associated with HCC. APLN and E2F1 were upregulated and ALDH2 and PCK1 were downregulated in HCC patients. High expression of APLN and E2F1 and low expression of ALDH2 and PCK1 resulted in poor prognosis of HCC patients. In expression analysis, ALDH2 showed significant change in all three categories. PCK1 showed highest mutation of out all 4 genes in HCC patients. T cell CD8+ is found to be positively correlated with APLN, ALDH2 and E2F1 and macrophages showed a positive correlation with APLN and E2F1.

Conclusions ALDH2 and PCK1 are great prognostic biomarkers and play a vital role in the development of HCC. Overexpression of ALDH2 and PCK1 can be a potential treatment strategy for HCC.

Keywords HCC, Adipokines, Adipogenesis, DEGs

Background

Liver cancer is the world's sixth most prevalent cancer and the third most significant cause of cancer-related death. According to the Global Cancer Statistics report of 2020 [23], there were 905,677 new cases of liver cancer and 830,180 deaths globally, underscoring its poor prognosis. An increase of 58.6% in incident cases of liver cancer is estimated by 2040. Liver hepatocellular carcinoma (HCC) is the most common cancer. HCC causes more than 80% of all primary liver cancer cases worldwide.

†Prithvi Singh and Rishabh Gurung contributed equally to this study as co-first authors

*Correspondence:
Ravins Dohare
ravinsdohare@gmail.com
[1] Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi 110025, India
[2] Department of Biotechnology, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi 110025, India

Primary risk factors include chronic hepatitis B virus (HBV) and hepatitis C virus (HCV) infection, excessive alcohol consumption, non-alcoholic fatty acid liver disease (NAFLD), obesity, diabetes mellitus and aflatoxins. HBV accounted for 219,000 incident cases of HCC and 192,000 deaths related to HCC globally in 2019 [27]. HCV about 17-fold increases the chance of acquiring HCC. After acquiring HCV, 80% of people develop chronic hepatitis, with 20% developing cirrhosis [2]. According to a meta-analysis of alcohol and liver cancer, those who consume three or more drinks per day have a 16% increased risk of liver cancer, while those who consume six or more drinks per day have a 22% increased risk [25]. NAFLD is prevalent in roughly 70% of diabetic individuals and 90% of obese people [9]. Diabetic people have a 2.5-fold increased risk of developing HCC [7]. According to a 2012 meta-analysis, aflatoxin B1 increased HCC risk by *six*fold, HBV raised HCC risk by 11-fold, and the two variables together elevated HCC risk by 54-fold [17].

Adipose tissue functions not only as a reservoir for excess energy but also as an endocrine organ. It secretes bioactive molecules termed adipokines. Adipokines are hormones that function as growth factors that regulate insulin resistance, affect fat and glucose metabolism, and play a role in pro- and anti-inflammatory responses. Evidence suggests that dysregulated production of adipokines plays a role in the development of obesity-related diseases [18, 19].

RNA sequencing (RNA-Seq) is enhancing transcriptome research. Researchers can use RNA-Seq to detect known and novel features in a single assay, allowing them to detect transcript isoforms, gene fusions, single nucleotide variants (SNVs), and other features without knowing anything beforehand. RNA-Seq has quickly established itself as the most effective method for high-throughput transcriptome profiling, thanks to the true discovery power of unbiased RNA detection. RNA-Seq gene expression analysis is critical for understanding cancer mechanisms and aiding genetic disease research. RNA-Seq is a powerful sequencing-based method for capturing a broad range of gene expression information. Compared to existing technologies like gene expression microarrays, RNA-Seq offers many significant advantages. RNA-Seq, unlike hybridization-based methods, is not limited to detecting transcripts that match existing genomic sequences. This makes RNA-Seq particularly appealing for non-model organisms with unknown genomic sequences. RNA-Seq can reveal the precise location of transcription boundaries to a single-base resolution. RNA-Seq has very little to no background noise relative to DNA microarrays. The dynamic range of expression levels over which transcripts can be detected is quite wide. On the other hand, DNA microarrays lack sensitivity for genes expressed at low or very high levels, resulting in a much smaller dynamic range. RNA-Seq has shown to be highly accurate for quantifying expression levels. RNA-Seq results also show high levels of reproducibility, both in terms of technical and biological replicates. Because there are no cloning steps in RNA-Seq, it requires fewer RNA samples [28].

In the current study, we extracted the TCGA-HCC dataset from the UCSC Xena browser and performed differential expression analysis using R to identify differentially expressed genes (DEGs). Genes of adipokines and adipogenesis family were filtered out of it, and survival analysis and stage analyses were conducted, giving us key genes associated with HCC. Pathway and gene ontology (GO) term enrichment analysis was performed on these key genes to identify the pathway and processes they are involved in. Expression analysis based on nodal metastasis, TP53 mutation and tumor grade, and mutation analysis of key genes were also carried out. Finally, tumor infiltration analysis was carried out for our key genes.

Methods

HCC mRNA data extraction and differential expressions analysis (DEA)

The UCSC Xena browser (https://xenabrowser.net/) was utilized to retrieve the mRNA HTSeq raw count data of HCC patients. We included only those patient samples which fulfilled the following criteria: (1) mRNA-Seq data of only primary solid tumor and solid tissue normal samples should be available; (2) all the selected samples must have survival information available; (3) all the selected samples must be present within the TCGA-GDC data portal; (4) information on race and gender of the selected patients must be available. Original raw read counts were backlog-transformed to obtain raw integer counts. We utilized the DESeq2 R package to obtain $\log_2$ transformed and normalized expression values via variance stabilizing transformation (VST). The ARSyNseq function within the NOISeq R package was utilized for batch correcting (with unknown batch setting) normalized expression values. Next, the biomaRt R package was utilized to map the Ensembl IDs to their corresponding HUGO gene nomenclature committee (HGNC) symbol(s) where only protein-coding genes were retained for further analysis. Expression values were taken as an average for those genes mapping to multiple Ensembl IDs in order to avoid redundancy. DEGs were recognized utilizing the limma R package corresponding to a cut-off: $\left|\log_2(\text{fold change})\right| > 2$ and a Benjamini–Hochberg (BH)-adjusted P value < 0.05.

Detection of HCC-specific adipokines and adipogenesis

Gene set of the adipokines family was retrieved from [20] and [26], while "Hallmark adipogenesis" and "Nakamura adipogenesis early up" gene sets corresponding to the adipogenesis family were retrieved from Molecular Signature Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/). Only overlapping genes among the two adipogenesis gene sets were finalized. The HCC-specific adipokines and adipogenesis were finalized by checking overlap of HCC DEGs and adipokines/adipogenesis gene sets.

Prognostic and stage plot analyses of HCC-specific adipokines and adipogenesis

Kaplan–Meier (KM) plotter (https://kmplot.com/analysis/) was accessed for prognostic analysis of HCC-specific adipokines and adipogenesis. We generated overall survival (OS) and relapse-free survival (RFS) KM plots of HCC-specific adipokines and adipogenesis across TCGA-HCC cohort. All the patients were bifurcated into higher and lower expression groups based on 'auto select best cutoff' option. Logrank P value < 0.05 between the two expression groups were considered as a statistically significant threshold for assessing the prognosis of HCC-specific adipokines and adipogenesis. Next, we queried the Gene Expression Profiling Interactive Analysis (GEPIA2) (http://gepia2.cancer-pku.cn/#index) for stage plot analysis of HCC-specific adipokines and adipogenesis. It is a web server that uses a common processing pipeline to analyze the RNA sequencing expression data of 9736 cancers and 8587 normal samples from the TCGA and GTEx projects. Tumor/normal DEA, profiling according to cancer kinds or pathological stages, patient survival analysis, similar gene recognition, correlation analysis, and dimensionality reduction analysis are some of the functions that can be customized [24]. The genes with logrank P value < 0.05 for both OS and RFS and Pr $(> F)$ value < 0.05 in stage plots were considered as statistically significant.

Pathway and GO term enrichment analyses of key adipokines and adipogenesis

For pathway and GO term enrichment analyses, Enrichr (https://maayanlab.cloud/Enrichr/) was accessed. It is a web-based enrichment analysis tool that includes over 30 gene-set libraries, an alternative approach to rank enriched terms, and various interactive visualization approaches to display enrichment results [5, 10, 29]. BioPlanet 2019, GO-Biological Process (BP), GO-Molecular Function (MF), GO-Cellular Compartment (CC) libraries within Enrichr were utilized for selecting top 10 significantly (P value < 0.05) enriched pathway and GO terms.

Validation of key adipokines and adipogenesis using UALCAN and cBioPortal

Using UALCAN (http://ualcan.path.uab.edu/), expression analysis of key adipokines and adipogenesis was performed based on tumor grade, TP53 mutation and nodal metastasis status. UALCAN is a web-based resource for analyzing cancer OMICS data. Researchers can use UALCAN to obtain Level 3 RNA-Seq data from The Cancer Genome Atlas (TCGA) and perform gene expression and survival analyses on over 20,500 protein-coding genes in 33 tumor types [4]. Using cBioPortal (https://www.cbioportal.org/) OncoPrint, cancer type summary and mutual exclusivity of key adipokines and adipogenesis were studied. The cBioPortal is a free, interactive website for exploring multidimensional cancer genomics data sets. It enables access to molecular profiles and clinical features from large-scale cancer genomics research in a fast, intuitive, and high-quality manner [3, 8]. The HCC (TCGA, Firehose legacy) dataset was chosen and 363 tumor samples present in our HCC mRNA dataset (extracted from UCSC Xena) were selected for mutation analysis.

Tumor infiltration analysis

We investigated the correlation between mRNA expression levels of key adipokines and adipogenesis with tumor-infiltrating immune cells such as macrophages and neutrophils across TCGA-HCC patients using TIMER 2.0 (http://timer.cistrome.org/). Spearman correlation was utilized to evaluate the statistical significance. TIMER2.0 is a web server that analyzes and visualizes tumor-infiltrating immune cells in real time. It includes four modules for looking into immune infiltrates and genetic or clinical features, as well as four modules for looking into cancer-related associations in the TCGA cohorts [11–13].

Results

HCC mRNA data extraction and DEA

HCC-specific mRNA cohort comprised 413 patient samples (i.e., 363 tumor and 50 normal samples). Additional file 1: Table S1 summarizes the clinical information associated with TCGA-HCC cohort. Post pre-processing (i.e., normalization, $\log_2$ transformation and batch correction), we obtained 19575 unique protein-encoding genes and their respective expression values across the samples. We identified 562 DEGs in accordance with the aforementioned cutoff, i.e., $\left|\log_2(\text{fold change})\right| > 2$ and a BH-P value < 0.05 utilizing limma. A total of 125 and 437 DEGs were filtered as up and downregulated, respectively.

Detection of HCC-specific adipokines and adipogenesis

Nineteen members within the adipokines gene set and 321 within the adipogenesis gene set were finalized. However, only five (i.e., APLN, ITLN1, SERPINE1, LCN2, SFRP5) and fourteen (i.e., ACADL, ALDH2, CNTFR, COL15A1, CYP26A1, E2F1, EGR2, GADD45B, LIFR, ORM1, PCK1, PTGIS, SERPINE1, SOCS3) genes of adipokines and adipogenesis gene sets were present in HCC-specific DEGs list, respectively. Box-and-whisker plots showing relative expression levels of HCC-specific adipokines and adipogenesis are shown in Fig. 1. We observed higher expression levels of APLN, LCN2, COL15A1 and E2F1 in tumor patients as compared to normal.

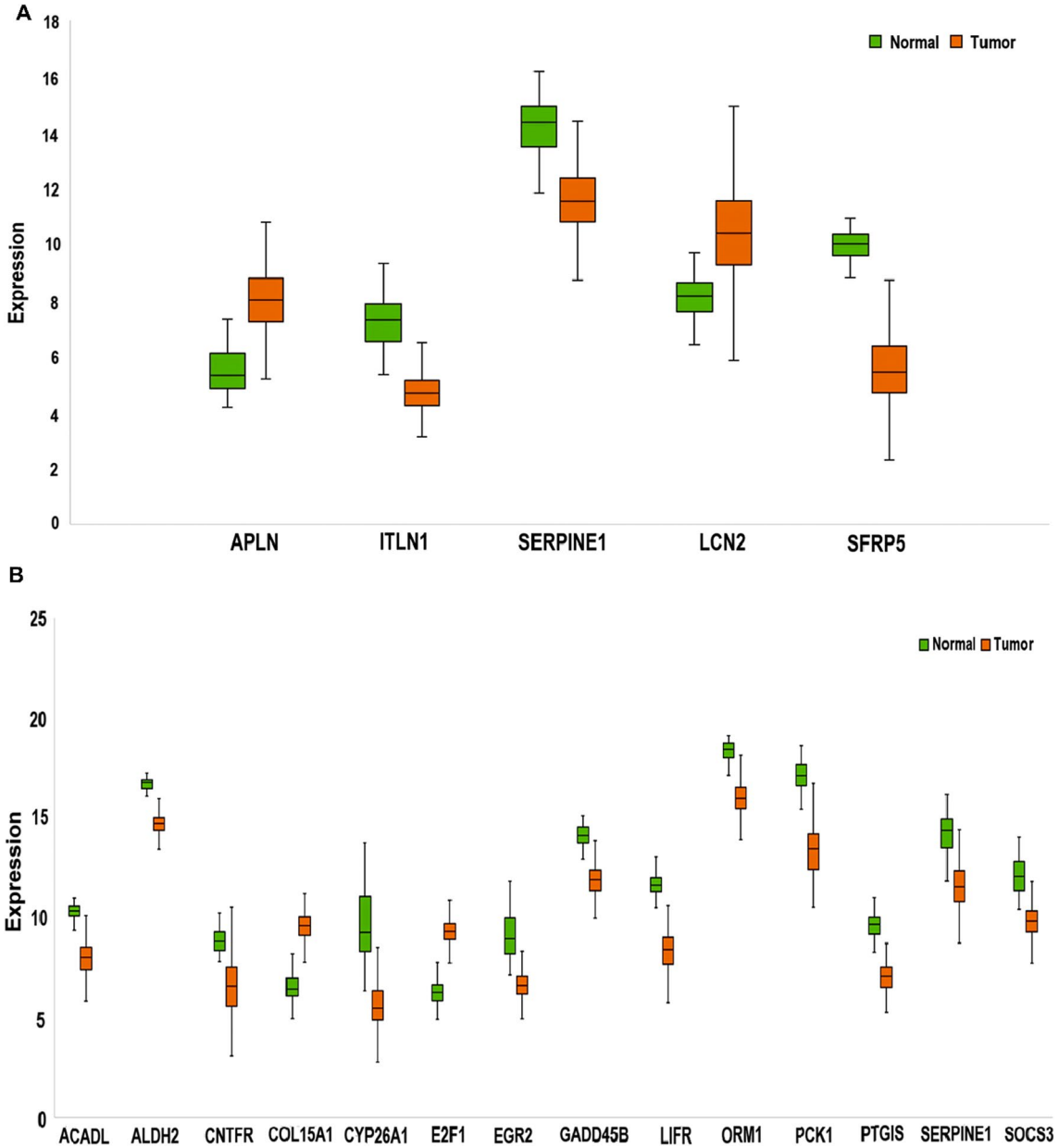

Fig. 1 Box-and-whisker plots displaying the expression distribution of **A** HCC-specific adipokines and **B** HCC-specific adipogenesis across TCGA-HCC cohort samples. Green- and orange-colored areas signify normal and tumor patient samples. The top and bottom of the boxes signify 75th and 25th percentile of distribution. Horizontal lines within the boxes represent the median values, while the axes end points are labeled by minimum and maximum values

Prognostic and stage plot analyses of HCC-specific adipokines and adipogenesis

Using KM plotter, prognostic analysis was performed on our 19 genes to determine the correlation between their expression levels and risk of HCC. We reported only those genes whose significant survival trend matched with their expression status. Higher mRNA expression levels of APLN and E2F1 worsened the OS of HCC patients. Also, lower mRNA expression levels of ACADL, ALDH2, GADD45B, and PCK1 worsened the OS of HCC patients (Fig. 2). We observed that higher mRNA expression levels of APLN and E2F1 worsened the RFS of HCC patients. Also, lower mRNA expression levels of SFRP5, ACADL, ALDH2, CNTFR, CYP26A1, EGR2, GADD45B, LIFR, PCK1, and PTGIS worsened the RFS of HCC patients (Fig. 3). ALDH2 reported highest hazard ratio (HR = 2.38) in OS category, while, in the RFS category, E2F1 reported highest hazard ratio (HR = 1.97). The median survival times w.r.t OS and RFS for prognostically significant HCC-specific adipokines and adipogenesis is reported in Additional file 1: Tables S2 and S3, respectively. Stage plot analysis (Fig. 4) revealed high expression of APLN across stages II and III of HCC as compared to stages I and IV. SERPINE1 expression incremented gradually with cancer stage advancement from stages I to IV, whereas ALDH2 expression decremented gradually from stages I to IV. ORM1 showed almost stable expression trend across all four stages. PCK1 expression decremented from stages I to III but elevated in

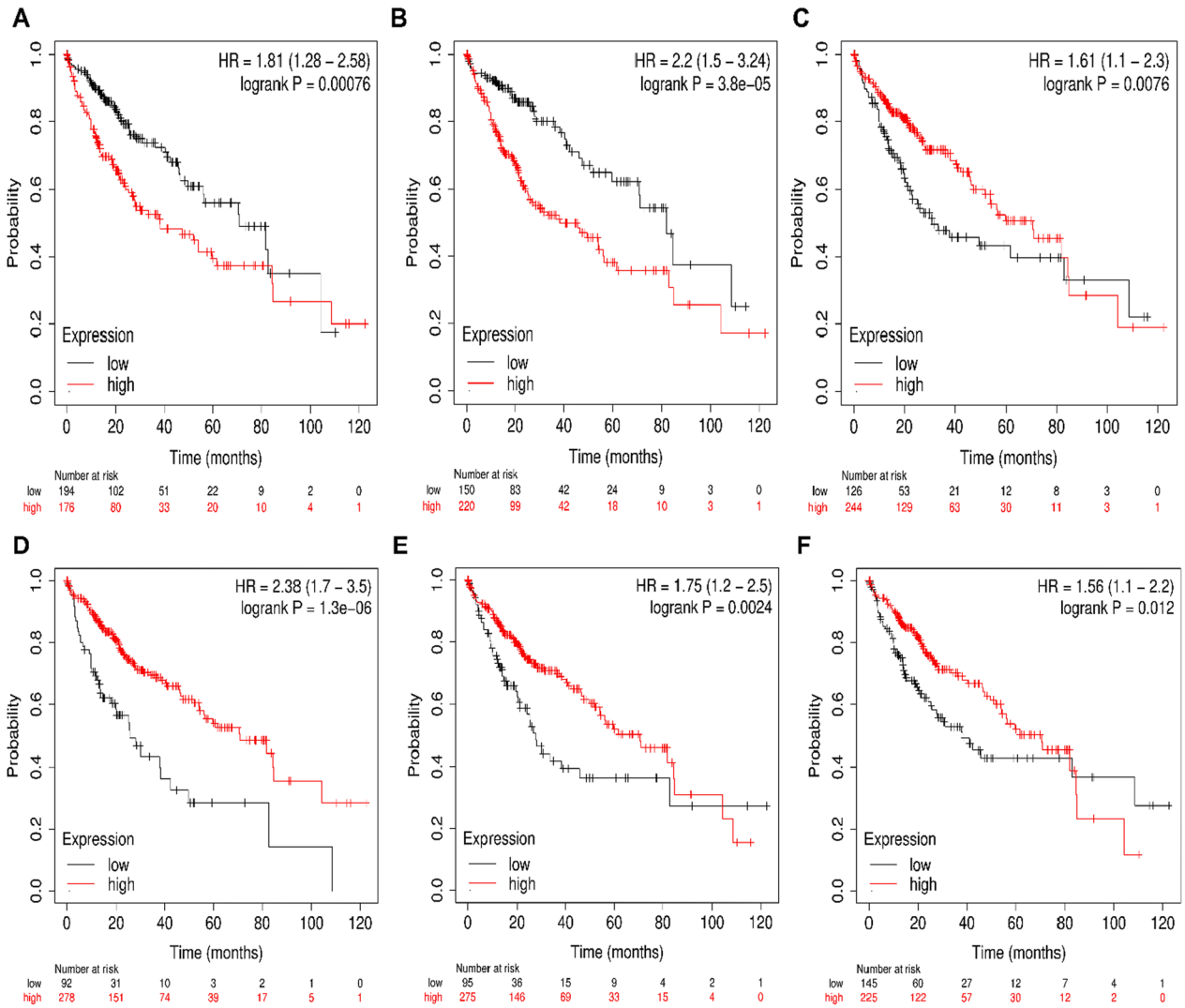

Fig. 2 KM plots showing the OS of **A** APLN, **B** E2F1, **C** ACADL, **D** ALDH2, **E** GADD45B, and **F** PCK1 across TCGA-HCC cohort. Higher and lower expression groups are signified by red and black colors

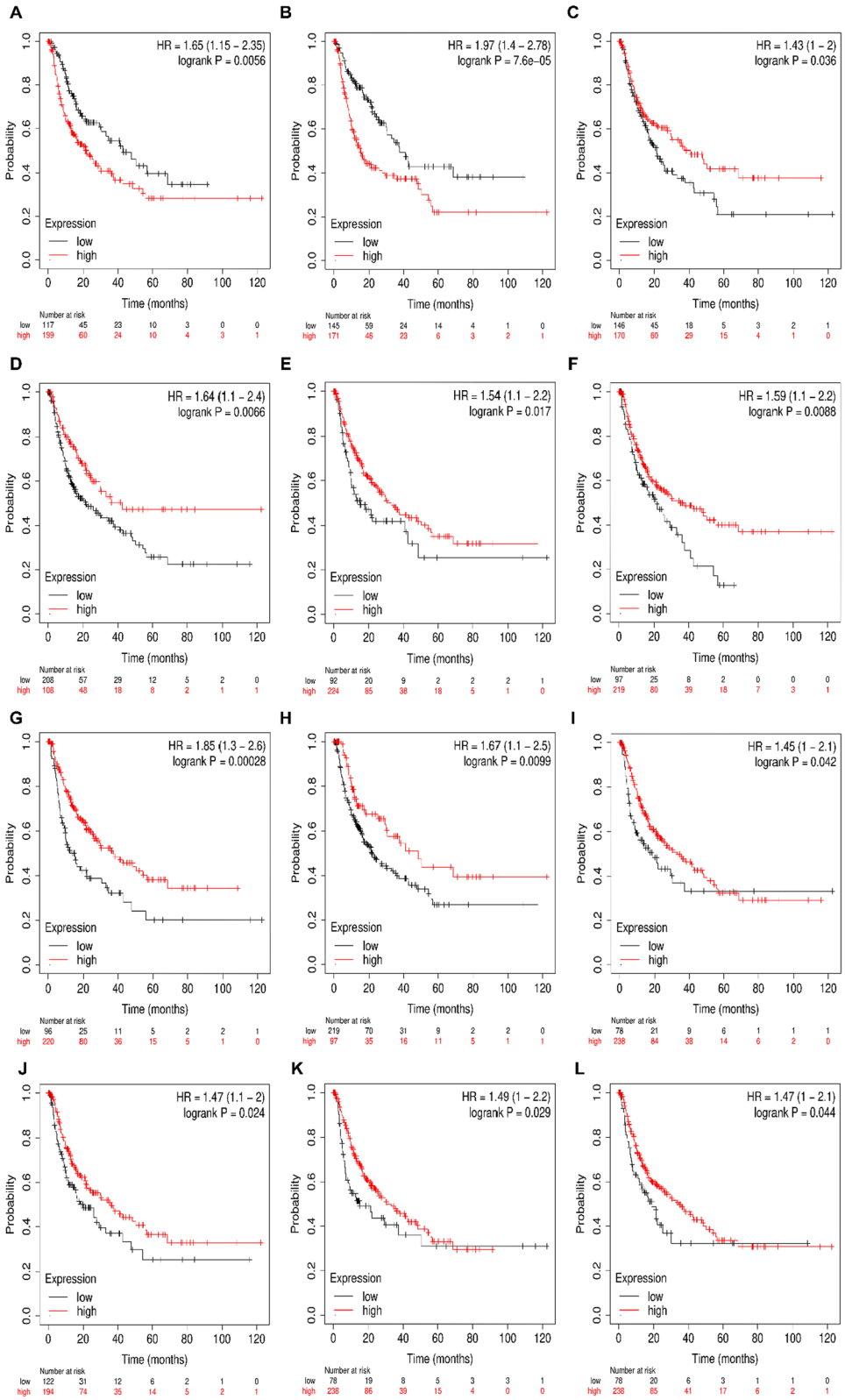

Fig. 3 KM plots showing the RFS of **A** APLN, **B** E2F1, **C** SFRP5, **D** ACADL, **E** ALDH2, **F** CNTFR, **G** CYP26A1, **H** EGR2, **I** GADD45B, **J** LIFR, **K** PCK1, **L** PTGIS across TCGA-HCC cohort. Higher and lower expression groups are signified by red and black colors

stage IV. We finalized APLN, ALDH2, E2F1 and PCK1 as our key HCC-specific adipokines and adipogenesis because they only reported significant results in both survival and stage plot analyses.

Pathway and GO term enrichment analyses of key adipokines and adipogenesis

Utilizing abovementioned Enrichr libraries, we performed pathway and GO term enrichment analyses for our 4 key genes (i.e., APLN, ALDH2, E2F1 and PCK1) and top 10 entries with P value < 0.05 (Additional file 1: Fig. S1 and S2). Most significant pathway, GO-BP, and GO-MF terms were pyruvate metabolism (P value $= 3.96 \times 10^{-5}$), oxaloacetate metabolic process (P value $= 9.99 \times 10^{-4}$), and aldehyde dehydrogenase (NAD+) activity (P value $= 2.59 \times 10^{-3}$). In GO-CC, all entries were statistically insignificant, i.e., (P value > 0.05).

Validation of key adipokines and adipogenesis using UALCAN and cBioPortal

Expression analysis based on nodal metastasis of our 4 genes revealed a constant significant decrease in ALDH2 expression as the cancer metastasized in 1 to 3 axillary lymph nodes. No significant change in expression of APLN, E2F1 and PCK1 was observed as metastasis progressed from N0 to N1. Expression analysis based on TP53 mutation of our 4 genes revealed a significant difference in the expression of ALDH2, E2F1 and PCK1 in HCC patients with mutant and non-mutant TP53. ALDH2 and PCK1 showed high expression in non-mutant TP53 compared to mutant TP53, while E2F1 showed low expression in non-mutant TP53 compared to mutant TP53 in HCC patients. Expression analysis based on tumor grade of our 4 key genes revealed a constant decrease in the expression of ALDH2 and PCK1 as the tumor grade increased. E2F1 showed a gradual increase in expression, while APLN expression showed no pattern with tumor grade (Fig. 5). cBioPortal was

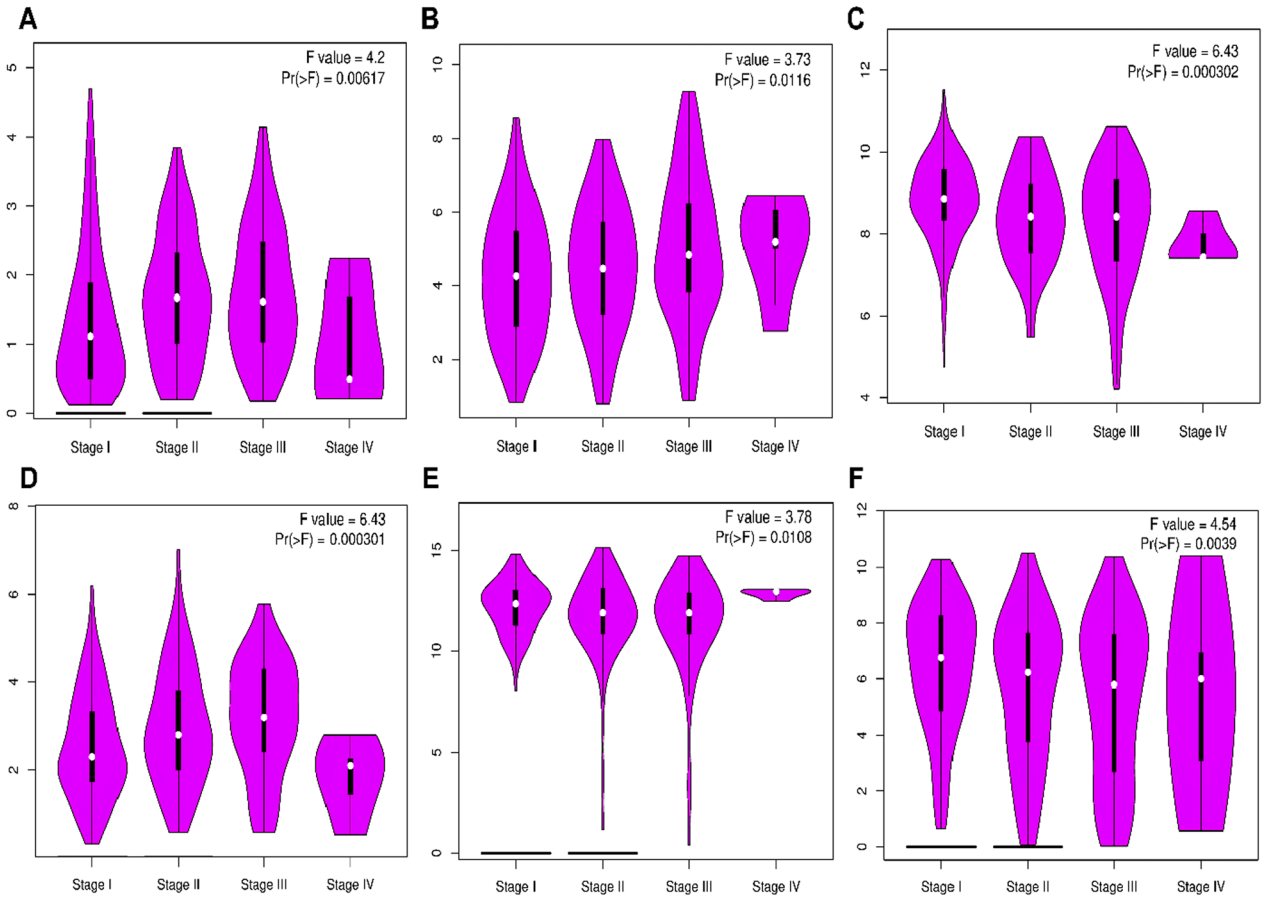

Fig. 4 Violin plots displaying association between significant TNM stages and **A** APLN, **B** SERPINE1, **C** ALDH2, **D** E2F1, **E** ORM1, **F** PCK1 across TCGA-HCC cohort. The black-colored bars and white-colored dots signify interquartile ranges and median, respectively. The ordinate and abscissa depict expression levels of these key genes and various stages

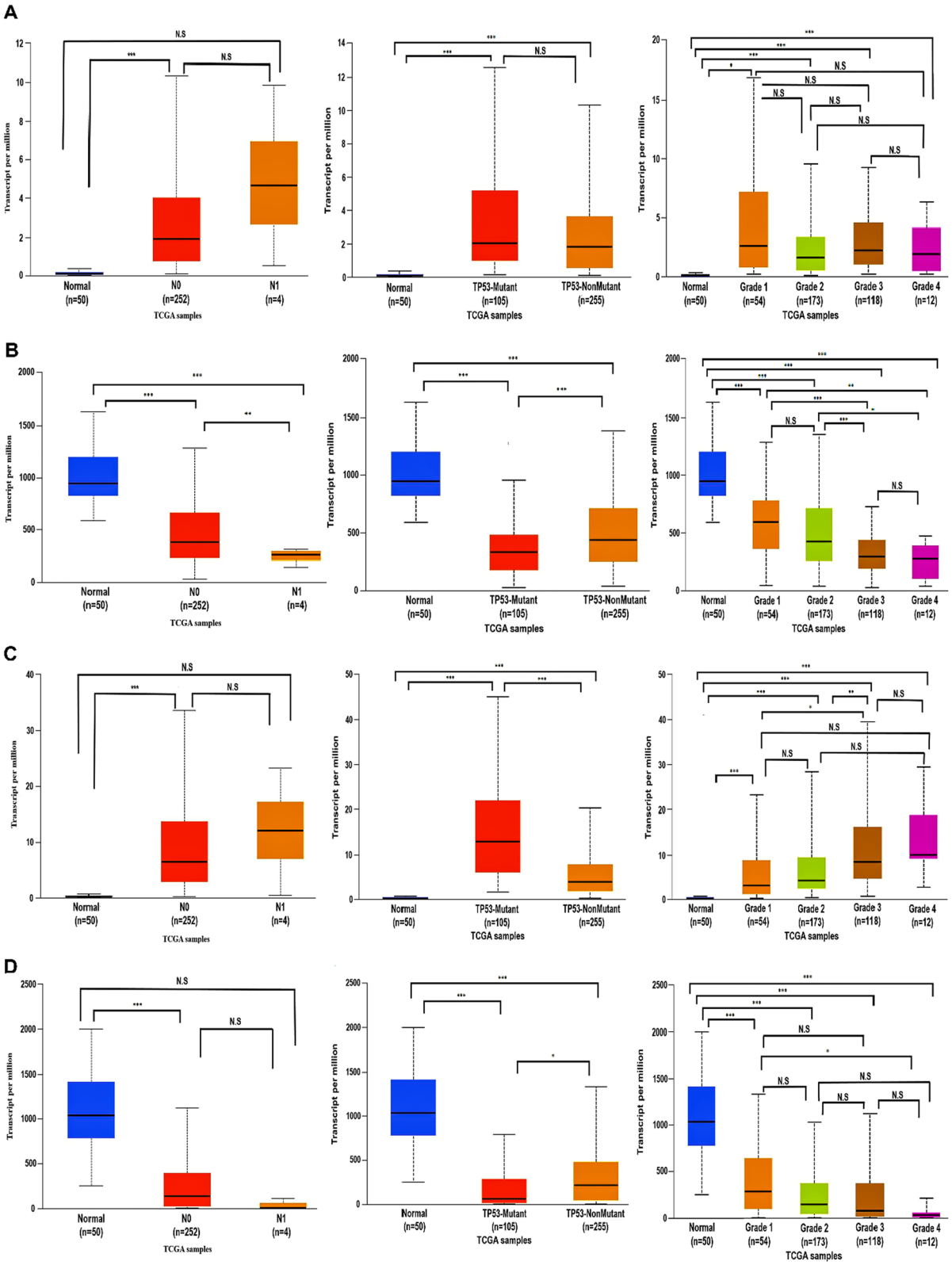

Fig. 5 Box-and-whisker plots displaying expression distribution of **A** APLN, **B** ALDH2, **C** E2F1, and **D** PCK1 based on nodal metastasis status (N0 means no regional lymph node metastasis; N1 means metastases in 1 to 3 axillary lymph nodes), TP53 mutation status, and tumor grade across TCGA-HCC cohort. *P value < 0.05, **P value < 0.01, ***P value < 0.001

used to validate the specific genetic modifications associated with key adipokines and adipogenesis (i.e., APLN, ALDH2, E2F1, PCK1) across the HCC dataset (TCGA, Firehose legacy) comprising 363 primary tumor samples. OncoPrint results for these queried genes as shown in Fig. 6 revealed genetic alterations in 7% (24/363) cases. As observed, PCK1 showed maximum mutation frequency (3%). The cancer type summary analysis revealed the overall alteration frequency of these genes as shown in Fig. 7. APLN, ALDH2, E2F1, and PCK1 were altered in 1.65%, 0.55%, 1.93%, and 3.03% of 363 total HCC patient samples. We observed 1.1% (4/363 cases) amplifications, 0.28%(1/363 case) deep deletion, and 0.28%(1/363 case) missense mutation in case of APLN. In case of ALDH2, we observed 0.28%(1/363 case) amplification and 0.28%(1/363 case) missense mutation. In case of E2F1, we observed 1.1%(4/363 cases) amplification and 0.83%(3/363 cases) missense mutation. In case of

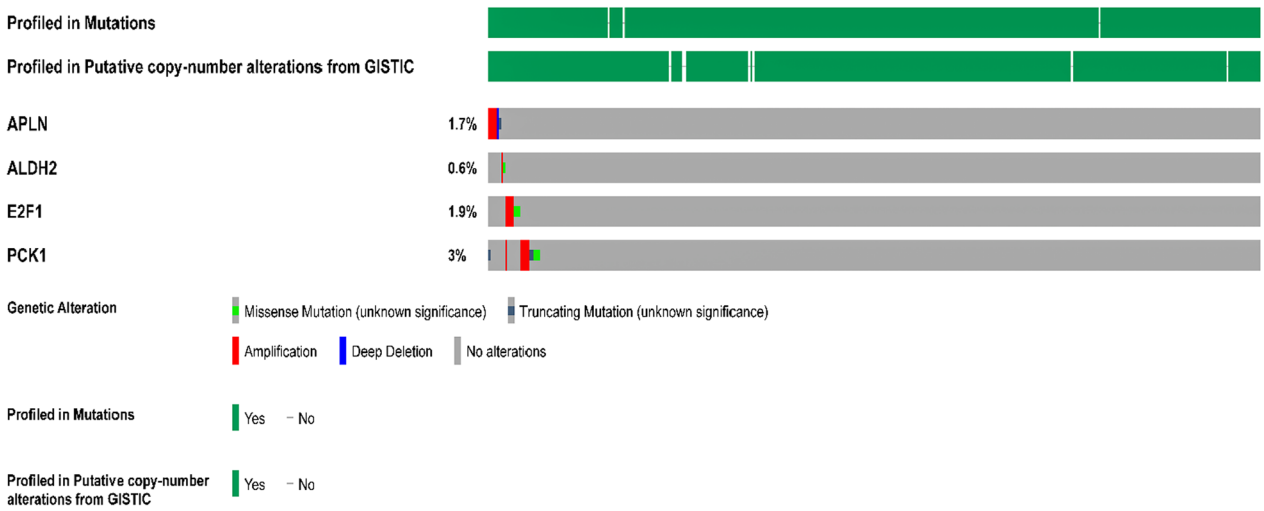

Fig. 6 OncoPrint summarizes genomic alterations in key adipokines and adipogenesis across TCGA-HCC cohort comprising 363 samples. The bottom row represents frequency of genomic alterations in APLN, ALDH2, E2F1, PCK1 with red, blue, green, and gray bars signifying amplifications, deep deletions, missense and truncating mutations, respectively

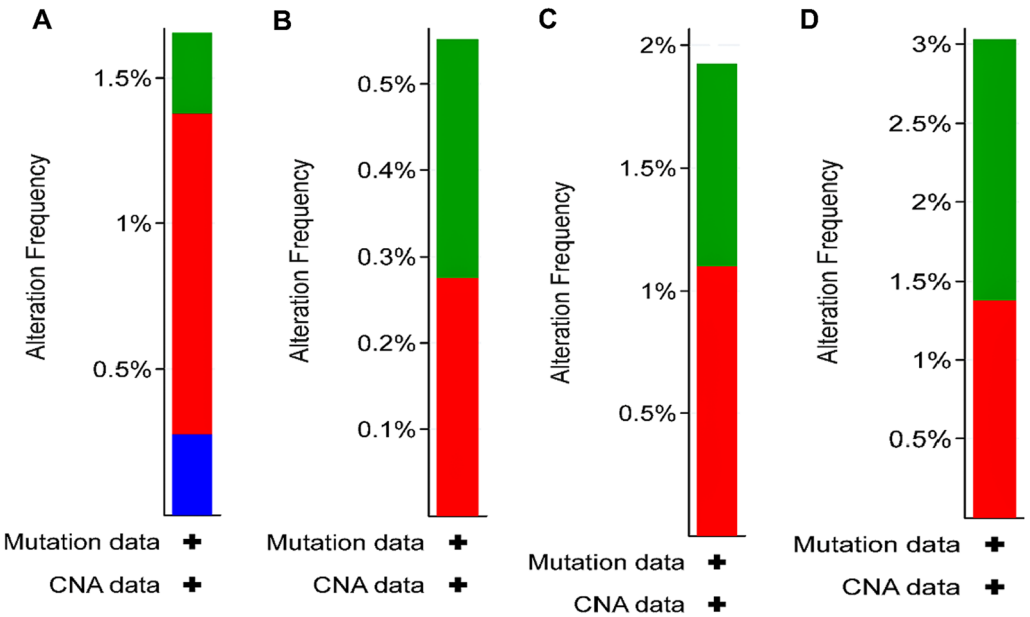

Fig. 7 Barplots showing alteration frequencies of **A** APLN, **B** ALDH2, **C** E2F1, and **D** PCK1 across TCGA-HCC cohort. Red, blue, and green colored bars signify amplifications, deep deletions, and missense mutations

PCK1, we observed 1.38%(5/363 cases) amplification and 1.65%(6/363 case) missense mutation.

Tumor infiltration analysis

Correlation of APLN, ALDH2, E2F1, and PCK1 mRNA expression levels with tumor purity and infiltrating levels of neutrophils, macrophages, T cell CD4+, and T cell CD8+ across TCGA-HCC cohort are shown by scatterplots in Fig. 8A–D. APLN displayed significant positive correlations with infiltrating levels of neutrophils ($r = 0.228$, $p = 1.83 \times 10^{-5}$) and macrophages ($r = 0.12$, $p = 2.59 \times 10^{-2}$). ALDH2 displayed significant negative correlations with infiltrating levels of neutrophils ($r = -0.197$, $p = 2.28 \times 10^{-4}$) and macrophages ($r = -0.152$, $p = 4.68 \times 10^{-3}$). E2F1 displayed significant positive correlations with infiltrating levels of neutrophils ($r = 0.104$, $p = 0.05$) and macrophages ($r = 0.272$, $p = 2.91 \times 10^{-7}$). PCK1 displayed significant negative correlations with infiltrating levels of neutrophils ($r = -0.22$, $p = 3.72 \times 10^{-5}$) and macrophages ($r = -0.215$, $p = 5.73 \times 10^{-5}$). APLN ($r = 0.224$, $p = 2.70 \times 10^{-5}$), ALDH2 ($r = 0.114$, $p = 3.47 \times 10^{-2}$), E2F1 ($r = 0.241$, $p = 5.8 \times 10^{-6}$) displayed significant positive correlations with tumor purity across HCC. PCK1 ($r = 0.002$, $p = 9.71 \times 10^{-1}$) displayed nonsignificant positive correlations with tumor purity across HCC.

Discussion

Analysis of gene expression data can help to identify genes that are differentially expressed in disease states, which can suggest potential drug targets. By identifying genes that are differentially expressed in disease states, researchers can gain insights into the underlying biological processes involved in the disease and identify potential drug targets [1]. In a disease state, abnormal differential gene expression can play a key role in the development and progression of the disease. For example, the over-expression or under-expression of certain genes may contribute to the development of cancer by promoting cell proliferation, evasion of programmed cell death, or the ability to invade other tissues. Moreover, the differential expression levels of certain genes may be used to diagnose cancer or to predict the likelihood of a cancer recurrence.

Understanding the patterns of differential gene expression in HCC can provide insights into the underlying molecular mechanisms of the disease and may inform the development of new diagnostic and therapeutic approaches. In the present, we found over-expression of APLN, LCN2, COL15A1 and E2F1 and under-expression of ITLN1, SERPINE1, SFRP5, ACADL, ALDH2, CNTFR, CYP26A1, EGR2, GADD45B, LIFR, ORM1, PCK1, PTGIS and SOCS3 in tumor patients as compared to normal patients.

Prognostic analysis of these nineteen genes showed worsened OS and RFS for patients exhibiting higher expression of APLN, E2F1 and lower expression of ALDH2 and PCK1. These four genes were considered as key genes. Among these four key genes, highest hazard ratio of 2.2 reflected that high expression of E2F1 in HCC patients is highly lethal.

Pathway enrichment analysis of these four key genes showed significant involvement of PCK1 and ALDH2 in pyruvate metabolism and glucose metabolism (glycolysis and gluconeogenesis) from which we concluded that HCC is significantly related with metabolic pathways. GO-BP enrichment analysis showed significant involvement of PCK1 in oxaloacetate metabolism which is a component of the TCA cycle and gluconeogenesis. GO-MF analysis revealed that ALDH2 codes for aldehyde dehydrogenase (NAD+) which plays role in alcohol metabolism. Even in the presence of oxygen, tumor cells conduct aerobic glycolysis known as the Warburg effect, which is essential for fulfilling the metabolic needs of rapid cancer cell growth and multiplication [14]. The liver is the primary site of gluconeogenesis, with the conversion of oxaloacetate (OAA) to phosphoenolpyruvate (PEP) by phosphoenolpyruvate carboxykinase serving as the initial rate-limiting step. PCK1 restricts glycolysis and promotes gluconeogenesis in HCC cell lines, reducing cancer cell survival, inducing apoptosis, and inhibiting migration. [16]. ALDH2 has been linked to malignancies caused by alcohol consumption. ALDH2 is a crucial enzyme in the detoxification of the ethanol metabolite acetaldehyde. ALDH2 deficiency results in the accumulation of acetaldehyde, which promotes liver inflammation, HBV infection persistence, T cell inactivation, and HCC. ALDH2 under-expression induces acetaldehyde build-up in cells, which enhances the activation of the AMP-activated protein kinase (AMPK) pathway in HCC, which is one of the mechanisms of ALDH2-modulated HCC progression. [30]. Furthermore, a recent study found that ALDH2 deficiency promotes HCC growth by transferring oxidized mitochondrial DNA-rich extracellular vesicles from weaker hepatocytes to HCC cells. These extracellular vesicles, in conjunction with acetaldehyde, activate multiple oncogenic pathways (including C-Jun N-terminal kinase, signal transducer and activator of transcription 3, BCL-2, and transcriptional co-activator with PDZ-binding motif) and promote HCC carcinogenesis after chronic exposure to alcohol [22].

Further, expression analysis based on nodal metastasis, TP53 mutation and tumor grade of these four key genes revealed significant changes in ALDH2 expression in all three categories. This reflects that ALDH2 is differentially

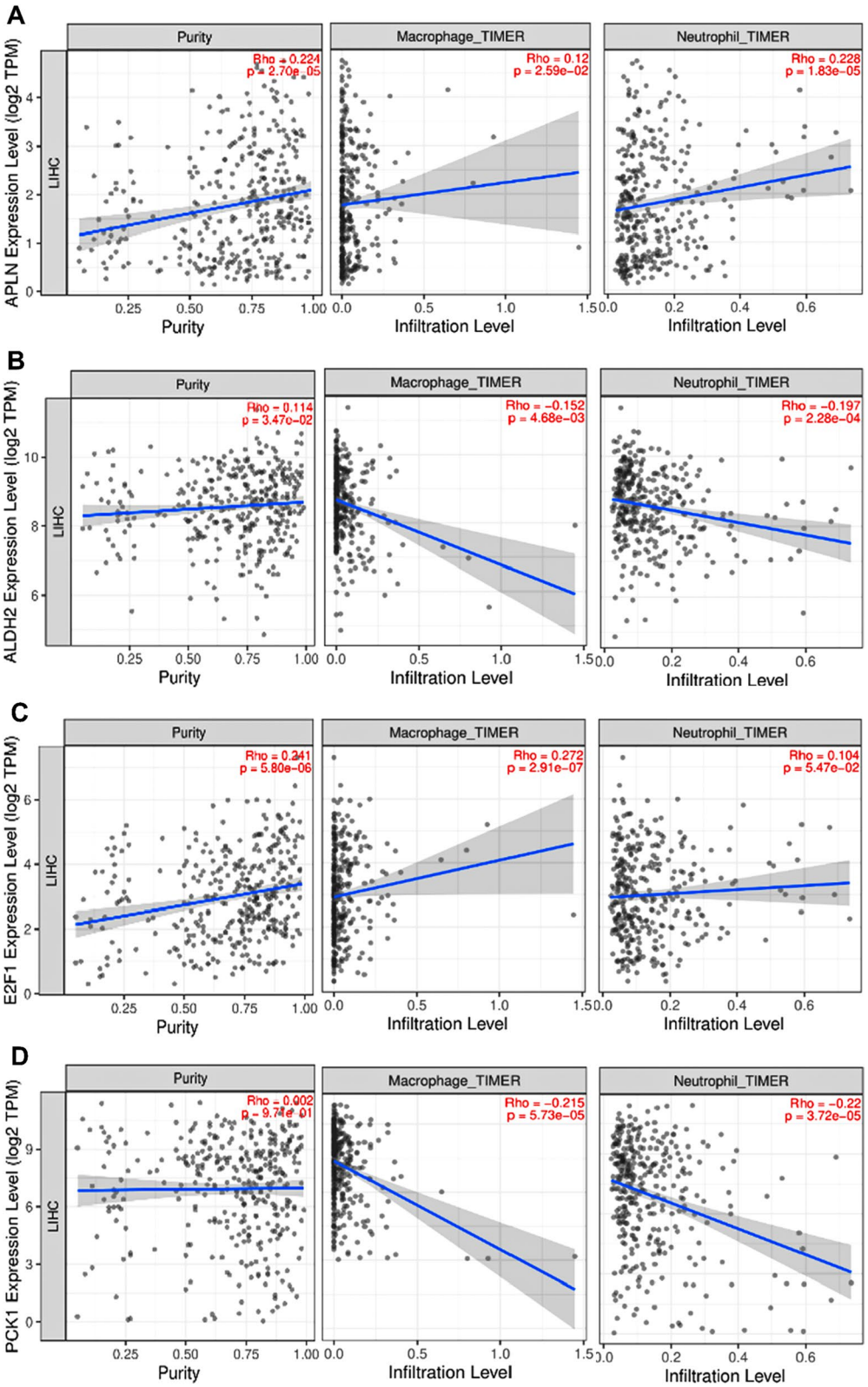

Fig. 8 Scatterplots showing significant correlations of **A** APLN, **B** ALDH2, **C** E2F1, and **D** PCK1 with infiltrating levels of neutrophils and macrophages across TCGA-HCC cohort. Spearman's correlation value and estimated statistical significance were shown as the legends for each scatter plot

regulated in cancer samples with different characteristics and may be involved in the development or progression of the disease. OncoPrint analysis is often used to identify patterns of genetic changes in cancer samples and to investigate how these changes might be related to the development or progression of the disease [8]. It can also be used to compare the genetic changes that occur in different types of cancer or to identify potential therapeutic targets for the treatment of cancer.

OncoPrint analysis of these four genes revealed that PCK1 has the highest frequency of genetic alteration (3%) as compared to the other three genes in HCC patients. Also, amplification was found to be the most prominent type of genetic alteration in all four genes. No mutual exclusivity was found between our set of four genes.

Macrophage showed positive correlation with APLN and E2F1 while negative correlation with ALDH2 and PCK1. Majority of the studies have shown that macrophage mainly M2 macrophages inhibits antitumor immunity and aids in HCC progression. [6]. Recent evidence suggested that M2 macrophages promote pathogenic angiogenesis, tumor cell invasion and migration, epithelial-mesenchymal transition (EMT), and cancer stem cell-like characteristics, all of which contribute to the advancement of liver cancer [31]. Studies have shown increased macrophages in HCC tumor tissue as compared to healthy liver [21]. Neutrophils are abundant in the microenvironment of liver cancer and exhibit different functional characteristics after being converted into tumor-associated neutrophils (TANs). In the tumor microenvironment, TANs promote HCC tumor formation, growth, and metastasis by promoting hepatoma cell proliferation, migration, invasion, colony formation, and the negative regulation of antitumor immunity. TANs can also release neutrophil extracellular traps (NETs) to promote the progression of liver cancer, induce tumor-related thrombosis, aggravate the body's hypercoagulable state, and increase the risk of tumor-related complications, such as organ failure. However, the antitumor effect of neutrophils in the tumor microenvironment should not be overlooked. Neutrophils induced by tumor necrosis factor alpha (TNF-α) have the capacity to slow tumor growth and metastasis through HGF/MET (MET proto-oncogene, receptor tyrosine kinase)-dependent nitric oxide release, whereas TANs can directly kill tumor cells by releasing ROS, stimulating the T cell response, assisting in antigen presentation, inhibiting early tumor formation, and inhibiting the formation of metastatic foci. [15].

By understanding the patterns of differential gene expression in HCC, this study showed a strong association of four key genes namely APLN, ALDH2, E2F1, and PCK1 with HCC. These four genes can be validated and developed as prognostic biomarkers. Overexpression or under-expression of these genes may be a targeted for potential treatment plan for HCC.

Conclusion

In conclusion, our studies showed that APLN, ALDH2, E2F1 and PCK1 were significantly associated with HCC. APLN and E2F1 showed high expression in HCC patients which was further associated with poor prognosis. ALDH2 and PCK1 showed low expression in HCC patients which was also related with poor prognosis. High expression of E2F1 showed to be highly lethal in HCC patients. APLN and PCK1 expression showed constant decrease as the tumor progressed. We also found that under-expression of PCK1 inhibits gluconeogenesis which promotes HCC. ADLH2 expression was found positively correlated with metastasis in 1 to 3 axillary lymph nodes. We found high mutation of PCK1 in HCC. Macrophages were negatively correlated with ALDH2 and PCK1. This study revealed significant association of four genes named APLN, ALDH2, E2F1, and PCK1 with HCC of which ALDH2 and PCK1 are great prognostic biomarkers for HCC and play a pivotal role in development of cancer. Overexpression of PCK1 and ALDH2 can be considered and targeted for potential treatment strategy in HCC.

Abbreviations

HCC	Hepatocellular carcinoma
DEGs	Differentially expressed genes
GO	Gene ontology
TCGA	The Cancer Genome Atlas
HBV	Hepatitis B virus
HCV	Hepatitis C virus
NAFLD	Non-alcoholic fatty acid liver disease
VST	Variance stabilizing transformation
BH	Benjamini–Hochberg
KM	Kaplan–Meier
RFS	Relapse-free survival
OS	Overall survival
HR	Hazard ratio
BP	Biological process
MF	Molecular function
CC	Cellular component
OAA	Oxaloacetate
AMPK	AMP-activated protein kinase
PEP	Phosphoenolpyruvate
E2F1	E2F transcription factor 1
APLN	Apelin

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43042-023-00401-5.

Additional file 1. Clinical patients' information from TCGA-HCC cohort (Table S1); Pathway and GO term enrichment analysis (Fig S1 and S2); OS and RFS median time in different expression cohorts with respect to each gene (Table S2 and S3).

Acknowledgements
The authors would like to thank Jamia Millia Islamia for providing infrastructure, journal access, and Internet facilities. Prithvi Singh would like to thank the Indian Council of Medical Research (ICMR) for awarding him Senior Research Fellowship (Grant Number: BMI/11(89)/2020).

Author contributions
PS contributed to conceptualization, methodology, software, formal analysis, data curation, writing—original draft, writing—review and editing, visualization. RG contributed to methodology, software, formal analysis, data curation, writing—original draft, writing—review and editing, visualization. AS contributed to writing—review and editing. RD contributed to writing—review and editing, supervision, project administration. All authors read and approved the final manuscript.

Funding
This research work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials
The raw HTSeq count data of TCGA-HCC cohort used in our study were downloaded from UCSC Xena browser (https://xenabrowser.net/datapages/?dataset=TCGA-LIHC.htseq_counts.tsv&host=https%3A%2F%2Fgdc.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443, accessed on January 10, 2022).

Declarations

Ethics approval and consent to participate
Not applicable.

Human and animal ethics
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that there are no conflict of interests.

Received: 23 August 2022 Accepted: 16 February 2023
Published online: 22 February 2023

References

1. Alam A, Abubaker Bagabir H, Sultan A et al (2022) An integrative network approach to identify common genes for the therapeutics in tuberculosis and its overlapping non-communicable diseases. Front Pharmacol 12:4082
2. Asham EH, Kaseb A, Ghobrial RM (2013) Management of hepatocellular carcinoma. Surg Clin N Am 93:1423–1450. https://doi.org/10.1016/j.suc.2013.08.008
3. Cerami E, Gao J, Dogrusoz U et al (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2:401–404. https://doi.org/10.1158/2159-8290.CD-12-0095
4. Chandrashekar DS, Bashel B, Balasubramanya SAH et al (2017) UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 19:649–658. https://doi.org/10.1016/j.neo.2017.05.002
5. Chen EY, Tan CM, Kou Y et al (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinform 14:128. https://doi.org/10.1186/1471-2105-14-128
6. Dou L, Shi X, He X, Gao Y (2020) Macrophage phenotype and function in liver disorder. Front Immunol 10:3112
7. El-Serag HB (2012) Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology 142:1264-1273.e1. https://doi.org/10.1053/j.gastro.2011.12.061
8. Gao J, Aksoy BA, Dogrusoz U et al (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6:pl1. https://doi.org/10.1126/scisignal.2004088
9. Gomes MA, Priolli DG, Tralhão JG, Botelho MF (2013) Hepatocellular carcinoma: epidemiology, biology, diagnosis, and therapies. Rev Assoc Med Bras (1992) 59:514–524. https://doi.org/10.1016/j.ramb.2013.03.005
10. Kuleshov MV, Jones MR, Rouillard AD et al (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44:W90-97. https://doi.org/10.1093/nar/gkw377
11. Li B, Severson E, Pignon J-C et al (2016) Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol 17:174. https://doi.org/10.1186/s13059-016-1028-7
12. Li T, Fan J, Wang B et al (2017) TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 77:e108–e110. https://doi.org/10.1158/0008-5472.CAN-17-0307
13. Li T, Fu J, Zeng Z et al (2020) TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res 48:W509–W514. https://doi.org/10.1093/nar/gkaa407
14. Liberti MV, Locasale JW (2016) The Warburg effect: How does it benefit cancer cells? Trends Biochem Sci 41:211–218. https://doi.org/10.1016/j.tibs.2015.12.001
15. Liu K, Wang F-S, Xu R (2021) Neutrophils in liver diseases: pathogenesis and therapeutic targets. Cell Mol Immunol 18:38–44. https://doi.org/10.1038/s41423-020-00560-0
16. Liu M-X, Jin L, Sun S-J et al (2018) Metabolic reprogramming by PCK1 promotes TCA cataplerosis, oxidative stress and apoptosis in liver cancer cells and suppresses hepatocellular carcinoma. Oncogene 37:1637–1653. https://doi.org/10.1038/s41388-017-0070-6
17. Liu Y, Chang C-CH, Marsh GM, Wu F (2012) Population attributable risk of aflatoxin-related liver cancer: systematic review and meta-analysis. Eur J Cancer 48:2125–2136. https://doi.org/10.1016/j.ejca.2012.02.009
18. Pereira S, Alvarez-Leite J (2014) Adipokines: biological functions and metabolically healthy obese profile. JRLCR. https://doi.org/10.2147/JRLCR.S36060
19. Rajesh Y, Sarkar D (2021) Association of adipose tissue and adipokines with development of obesity-induced liver cancer. IJMS 22:2163. https://doi.org/10.3390/ijms22042163
20. Raucci R, Rusolo F, Sharma A et al (2013) Functional and structural features of adipokine family. Cytokine 61:1–14. https://doi.org/10.1016/j.cyto.2012.08.036
21. Rohr-Udilova N, Klinglmüller F, Schulte-Hermann R et al (2018) Deviations of the immune cell landscape between healthy liver and hepatocellular carcinoma. Sci Rep 8:6220. https://doi.org/10.1038/s41598-018-24437-5
22. Seo W, Gao Y, He Y et al (2019) ALDH2 deficiency promotes alcohol-associated liver cancer by activating oncogenic pathways via oxidized DNA-enriched extracellular vesicles. J Hepatol 71:1000–1011. https://doi.org/10.1016/j.jhep.2019.06.018
23. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/caac.21660
24. Tang Z, Kang B, Li C et al (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res 47:W556–W560. https://doi.org/10.1093/nar/gkz430
25. Turati F, Galeone C, Rota M et al (2014) Alcohol and liver cancer: a systematic review and meta-analysis of prospective studies. Ann Oncol 25:1526–1535. https://doi.org/10.1093/annonc/mdu020
26. Vachher M, Arora K, Burman A, Kumar B (2020) NAMPT, GRN, and SERPINE1 signature as predictor of disease progression and survival in gliomas. J Cell Biochem 121:3010–3023. https://doi.org/10.1002/jcb.29560
27. Vos T, Lim SS, Abbafati C et al (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 396:1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9
28. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

29. Xie Z, Bailey A, Kuleshov MV et al (2021) Gene set knowledge discovery with Enrichr. Curr Protoc 1:e90. https://doi.org/10.1002/cpz1.90
30. Zhang H, Fu L (2021) The role of ALDH2 in tumorigenesis and tumor progression: targeting ALDH2 as a potential cancer treatment. Acta Pharm Sin B 11:1400–1411. https://doi.org/10.1016/j.apsb.2021.02.008
31. Zheng X, Jin W, Wang S, Ding H (2021) Progression on the roles and mechanisms of tumor-infiltrating T lymphocytes in patients with hepatocellular carcinoma. Front Immunol 12:729705

Publisher's Note