


ORIGINAL RESEARCH

In silico analysis of the association between long non-coding RNA family with sequence similarity 99 member A (*FAM99A*) and hepatic cancer

Manyi Sun¹  | Shuhua Lv² | Jin Zhong³

¹Department of Gastroenterology, Tianjin Union Medical Center, Tianjin, China

²Department of Pathology, Tianjin Union Medical Center, Tianjin, China

³Department of Radiology, Tianjin Union Medical Center, Tianjin, China

Correspondence

Manyi Sun, Department of Gastroenterology, Tianjin Union Medical Center, No. 190, Jieyuan Road, Hongqiao District, Tianjin 300121, China. Email: sunmanyi@umc.net.cn

Abstract

The link between family with sequence similarity 99 member A (*FAM99A*), a type of long non-coding RNA, and tumourigenesis remains ambiguous. Therefore, in this study, the authors conducted an expression profile analysis of *FAM99A* based on 33 types of cancer within The Cancer Genome Atlas project. The expression profile data revealed low expression levels of *FAM99A* in hepatocellular carcinoma, cholangiocarcinoma, and testicular germ cell tumour tissues than in the normal control tissues. Survival analysis indicated a correlation between low *FAM99A* expression and worse survival outcome in patients with hepatic cancer. Further investigation revealed the possible implication of DNA methylation, but not copy number variation, in *FAM99A*-associated hepatic tumourigenesis. The authors also identified a set of differentially expressed genes between patients with hepatic cancer and negative controls, which were found to be related to biochemical metabolism or the cell cycle. Additionally, *FAM99A* expression may be associated with the infiltration status of several immune cells, such as dendritic cells, natural killer cells, and neutrophils. Overall, *FAM99A* may function as a prognostic marker that is potentially associated with DNA methylation, immune cell infiltration, and biochemical metabolism in hepatic cancer.

KEYWORDS

cancer, genetics, genomics, tumours

1 | INTRODUCTION

Growing evidence suggests a functional link between long non-coding RNAs (lncRNAs) and tumourigenesis [1–5]. Information on various characteristics, such as expression, methylation, copy number variation (CNV), and clinical traits, are available on The Cancer Genome Atlas (TCGA) database [6, 7]. Based on the available datasets on liver hepatocellular carcinoma (LIHC) cohort, we screened various lncRNAs related to the pathological stages of hepatic cancer using bioinformatic methods, such as random forest (data not shown). The family with sequence similarity 99 member A (*FAM99A*) gene was identified. Very recently, Mo et al. reported the suppressor and prognostic effects of *FAM99A* on hepatocellular carcinoma (HCC) [8]. However, the deeper molecular

mechanism remains unclear. It is worth further exploring the role of *FAM99A* in the oncogenesis of hepatic cancer from more analytical perspectives. There was still no in silico analysis of *FAM99A* in a total of 33 types of cancers as well.

FAM99A is an lncRNA that functions as a foetal imprinted gene [9]. Three studies have investigated the association between *FAM99A* and pregnancy-related issues [9–11]. Two studies have reported the role of *FAM99A* in hypoxia-induced HCC metastasis [12] and glucose transporter 1-mediated glycolysis in HCC cells [13]. In this study, we aimed to analyse the expression profile of *FAM99A* in different cancer types. We investigated the functional links between lncRNA *FAM99A* and hepatic cancer in terms of gene expression differences, CNVs, DNA methylation level, and infiltration status of immune cells using datasets from TCGA, Gene

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *IET Systems Biology* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

Expression Omnibus (GEO), and Genotype-Tissue Expression (GTEx) databases.

2 | MATERIALS AND METHODS

2.1 | Expression feature

We downloaded the expression dataset of *FAM99A* [ENSG00000205866] gene and relative clinical traits from TCGA, GSE76427, and GSE64041 projects using R scripts, and determined the expression differences between the tissues of 33 types of cancers and negative controls using the Wilcoxon test. Bodymap with the indicated median expression values in LIHC, cholangiocarcinoma (CHOL), and testicular germ cell tumours (TGCT) tissues and normal controls was obtained through the Gene Expression Profiling Interactive Analysis 2 (GEPIA2) tool (<http://gepia2.cancer-pku.cn/#analysis>). GEPIA2 can be used to determine the gene expression differences between the tumour tissues and normal controls based on TCGA datasets [14, 15]. Since normal controls are missing for some types of cancer in TCGA, we obtained negative controls from the GTEx project. RNA-sequencing datasets of TCGA/GTEx with batch correction were obtained from University of California Santa Cruz (UCSC) Xena browser (<https://xenabrowser.net/datapages/>) through the Toil process [16]. After log₂ transformation of the data, we conducted an expression comparison analysis.

We extracted the paired expression dataset of carcinoma and adjacent non-carcinoma tissues to perform the Wilcoxon signed-rank test. The results were visualised using the ggplot2 R package. Receiver operating characteristic (ROC) analyses of 'Normal versus Tumour' tissues for *FAM99A* expression levels in LIHC, CHOL, and TGCT were conducted using a pROC R package. ROC with area under the curve (AUC) was plotted using the ggplot2 R package. The expression differences of *FAM99A* were assessed in different groups of clinical traits for LIHC or CHOL, such as T/M/N pathologic stage, histologic grade, vascular invasion, residual tumour, age, gender, and race of the patient. Furthermore, we performed the Kruskal–Wallis and Dunn's tests to determine the statistical differences among three or more than three groups, and the Wilcoxon test to compare two groups.

2.2 | Survival status

Kaplan–Meier plotter (http://kmplot.com/analysis/index.php?p=service&cancer=liver_rnaseq) was used to analyse the overall survival (OS), relapse-free survival (RFS), progression-free survival (PFS), and disease-specific survival (DSS) in hepatic cancer cases within TCGA [17]. We also examined some clinical traits in the survival analysis. We analysed the OS, PFS, and DSS in CHOL cases within TCGA-CHOL cohort using survival and Survminer R packages. Furthermore, we conducted univariate/multivariate Cox survival analyses of *FAM99A* expression and OS in hepatic cancer cases of

GSE76247 TCGA-LIHC cohort using the survival R package. *p* values and hazard ratios (HRs) with 95% confidence intervals were calculated.

2.3 | Copy number variation

CNV analysis of lncRNA *FAM99A* was conducted for hepatic cancer cases within TCGA-LIHC projects using GSCALite (<http://bioinfo.life.hust.edu.cn/web/GSCALite/>) [18]. When conducting correlation analysis of continuous variables in the whole study, we performed the Shapiro–Wilk normality test to determine whether the variables were normally distributed. Pearson correlation analysis was conducted if the variables showed normal distribution, otherwise, the Spearman correlation analysis was conducted. The data of the rank variables were analysed using the Spearman correlation test. We conducted Spearman's correlation analysis between CNVs and expression levels of *FAM99A*.

2.4 | DNA methylation

We used MEXPRESS to analyse the DNA methylation status of *FAM99A* [19, 20]. Pearson correlation analysis was used to determine the statistical correlation between DNA methylation and expression level of *FAM99A*. Multiple hypothesis testing was used to calculate the correlation coefficients and Benjamini–Hochberg (BH)-adjusted *p* values for different methylation probes, such as cg24218935, cg01745044, cg04353359, cg04938738, and cg25356611.

2.5 | *FAM99A*-correlated gene enrichment

Hepatic cancer cases in TCGA-LIHC project were divided into high and low expression groups based on the median value of *FAM99A* gene expression. We used the DESeq2 R package [21] for lncRNA *FAM99A*-correlated differential genes. We used the following parameters: $|\log_2 \text{fold change}| > 2$ and $p \text{ adj} < 0.01$. Volcano plots were generated using the ggplot2 R package. Next, enrichment analyses of the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were conducted. Based on the hallmark gene sets, gene set enrichment analysis (GSEA) with the normalised enrichment score was conducted and *p*. adjust and false discovery rate were calculated using clusterProfiler and ggplot2 R packages [22]. Using the Search Tool for the Retrieval of Interacting Genes/Proteins database (<http://string-db.org>) [23], we analysed *FAM99A*-associated protein–protein interactions (PPIs) using igraph and ggraph R packages.

Spearman's correlation analysis of *FAM99A* was conducted using the Stat R package. We plotted a heat map targeting *FAM99A*-correlated genes using the ggplot2 R package. Then, a Venn diagram between the differential genes (Diff) and Spearman correlated genes (Cor) of *FAM99A* was plotted using the ggplot2 R package. We used the Shapiro–Wilk

normality test to determine the correlation between *FAM99A* and some target genes, such as *FAM99B*, *AP006285.1*, and *LINC02708*. GO/KEGG pathway enrichment analyses were also conducted based on the common genes.

2.6 | Infiltration status of immune cells

We evaluated the tumour purity using the ESTIMATE R package [24] and analysed the statistical differences in the STROMAL, IMMUNE, and ESTIMATE scores between the high and low expression groups of *FAM99A* using a Wilcoxon test. Spearman correlations between *FAM99A* and the three indices of tumour purity were also determined. Furthermore, the correlation between *FAM99A* expression and tumour infiltration levels of a group of immune cells was analysed using the single-sample GSEA (ssGSEA) with the GSVA R package [25, 26]. Wilcoxon test was performed to explore the differences in the infiltration of immune cells between *FAM99A* high and low expression groups. Spearman's correlation analysis was used to determine the correlation between *FAM99A* expression and degree of immune cell infiltration.

In this study, some R-language analyses were conducted with the help of the Xiantao academic platform (<https://www.xiantao.love/products>). We provided the R scripts in Table S1.

3 | RESULTS

3.1 | Flow chart of analysis

Figure 1 presents the flowchart. Briefly, based on the datasets from TCGA, GTEx, GSE76427, and GSE64041, the expression profile of *FAM99A* in 33 types of cancers was analysed. Furthermore, survival analyses (Kaplan–Meier and Cox) were conducted for CHOL and hepatic cancers. We also investigated the correlation between *FAM99A* expression and a group of factors targeting hepatic cancer, including clinical traits, DNA methylation, CNV, immune cell infiltration, and tumour purity. Additionally, we obtained the differential genes for GO, KEGG, GSEA, and PPI analyses. After interaction analysis of these differential genes and *FAM99A*-correlated genes, we obtained a total of 49 common genes for another enrichment analysis of GO-KEGG.

3.2 | Expression profile analysis

The expression profile of *FAM99A* in different cancers was analysed based on TCGA, GTEx, and GEO databases. Compared with normal tissues, *FAM99A* was expressed at low levels in LIHC and CHOL tumour tissues (Figure 2a, $p < 0.001$). After the inclusion of normal controls for GTEx, we observed similar results for LIHC and CHOL (Figure 2b, $p < 0.001$). In addition, we observed a statistically significant difference between TGCT tissues and GTEx controls (Figure 2b, $p < 0.001$). Furthermore, we extracted the datasets of TCGA-paired samples to determine the expression difference between carcinoma and adjacent non-carcinoma tissues. Similar positive results were obtained for both LIHC and CHOL (Figure 2c, $p < 0.001$). In addition, as shown in Figure 2d, the ROC analysis data of 'Normal versus Tumour' confirmed the classification effect of *FAM99A* expression for TGCT (AUC = 0.950), LIHC (AUC = 0.825), and CHOL (AUC = 1.000). We also demonstrated the interactive bodymap of lncRNA *FAM99A* (Figure S1). In addition to TCGA, the difference in *FAM99A* expression between the HCC and adjacent normal tissues was analysed in the GSE64041 and GSE76427 datasets. We also observed a lower expression of *FAM99A* in HCC tissues than in the normal controls (Figure 2e, all $p < 0.05$).

The statistical difference in *FAM99A* expression in the different groups of clinical traits of LIHC or CHOL was also determined. As shown in Figure 3, we observed the statistical difference of *FAM99A* expression for LIHC in the groups of 'T1 versus T2' ($p = 2.2 \times 10^{-3}$), 'T1 versus T3' ($p = 0.03$), 'Stage I versus Stage II' ($p = 8.9 \times 10^{-4}$), 'Stage I versus Stage III/IV' ($p = 4.0 \times 10^{-3}$), 'G1 versus G3' ($p = 4.4 \times 10^{-6}$), 'G2 versus G3' ($p = 2.0 \times 10^{-3}$), 'No versus Yes (Vascular invasion)' ($p = 3.6 \times 10^{-5}$), but not other groups. With regard to CHOL, there was a statistical difference in *FAM99A* expression for the sex groups (Figure S2, $p = 1.9 \times 10^{-3}$), but not for the other groups.

3.3 | Survival analysis

The relationship between *FAM99A* expression status and the clinical prognosis of hepatic cancer and CHOL was analysed. There were lower rates of OS (Figure 4a, HR = 0.56, $p = 0.0014$), PFS (HR = 0.62, $p = 0.0035$), RFS (HR = 0.63,

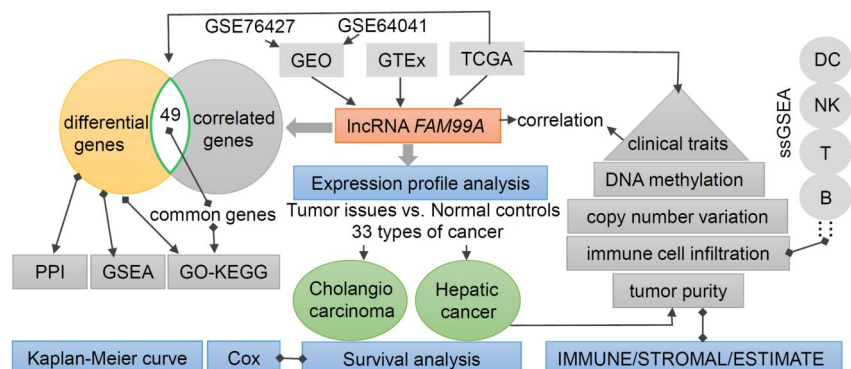


FIGURE 1 Flow chart of the study.

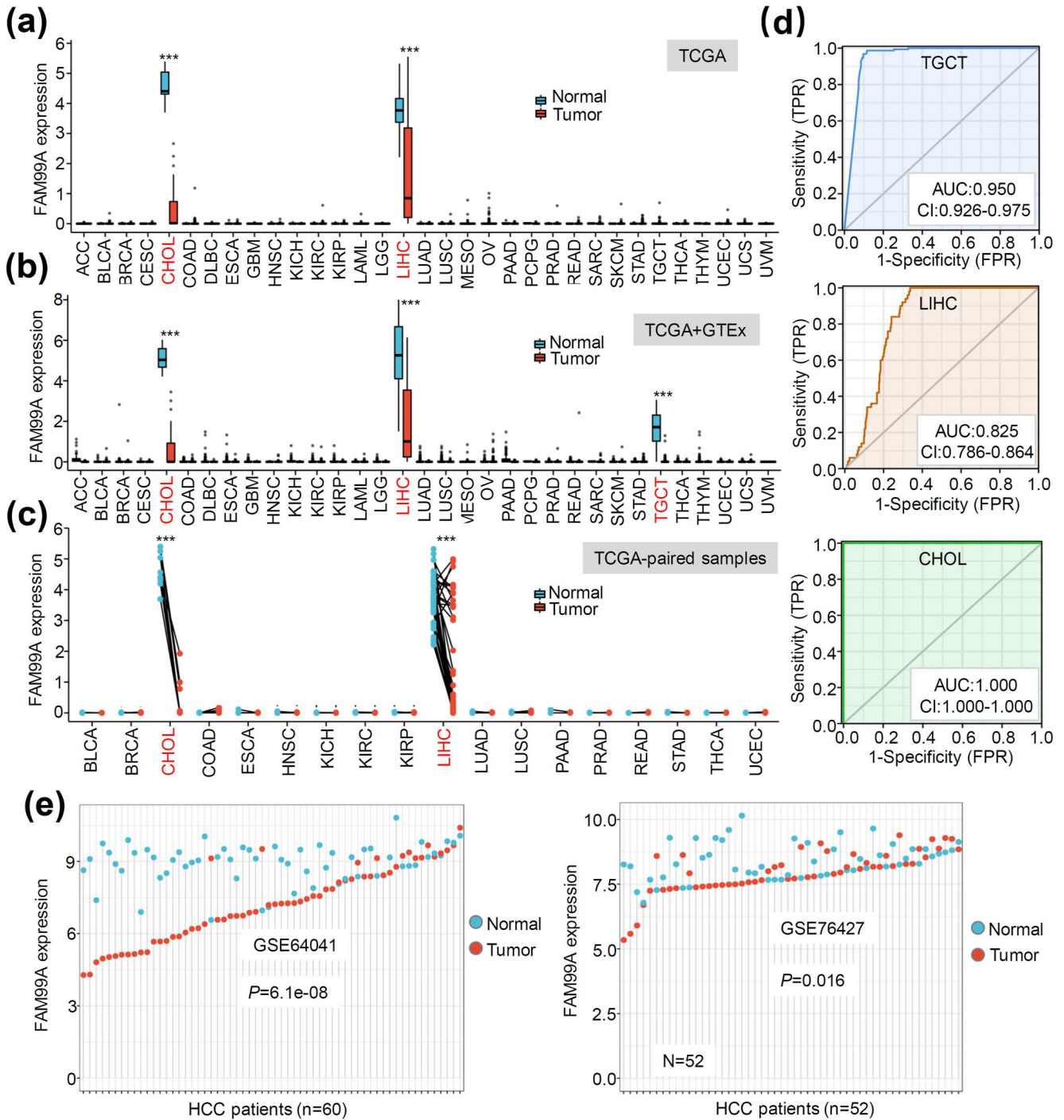


FIGURE 2 Expression profile analysis of lncRNA family with sequence similarity 99 member A (*FAM99A*). (a) Expression profiles of *FAM99A* gene in different cancer and corresponding control tissues in TCGA project were analysed. (b) Data of normal tissues within the GTEx database were combined with those of controls within TCGA databases. Wilcoxon test was used for statistical analysis. (c) Paired expression dataset of carcinoma and adjacent non-carcinoma tissues on TCGA was extracted to perform the Wilcoxon signed-rank test. ****p* < 0.001. (d) ROC analyses of 'Normal versus Tumour' tissues for *FAM99A* expression levels in LIHC, CHOL, and TGCT were conducted using a pROC R package. (e) Expression differences of *FAM99A* between the carcinoma and adjacent non-carcinoma tissues within GSE64041 and GSE76427 datasets were determined using the Wilcoxon signed-rank test. CHOL, cholangiocarcinoma; GTEx, genotype-tissue expression; LIHC, liver hepatocellular carcinoma; lncRNA, long non-coding RNA; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas; TGCT, testicular germ cell tumours.

p = 0.011), and DSS (HR = 0.56, *p* = 0.015) in patients with hepatic cancer in the *FAM99A* low expression group than in the high expression group. However, we did not obtain a

positive conclusion for the cholangio carcinoma (Figure 4b). Therefore, in this study, we focussed on hepatic cancer. We performed further survival analyses after grouping the samples

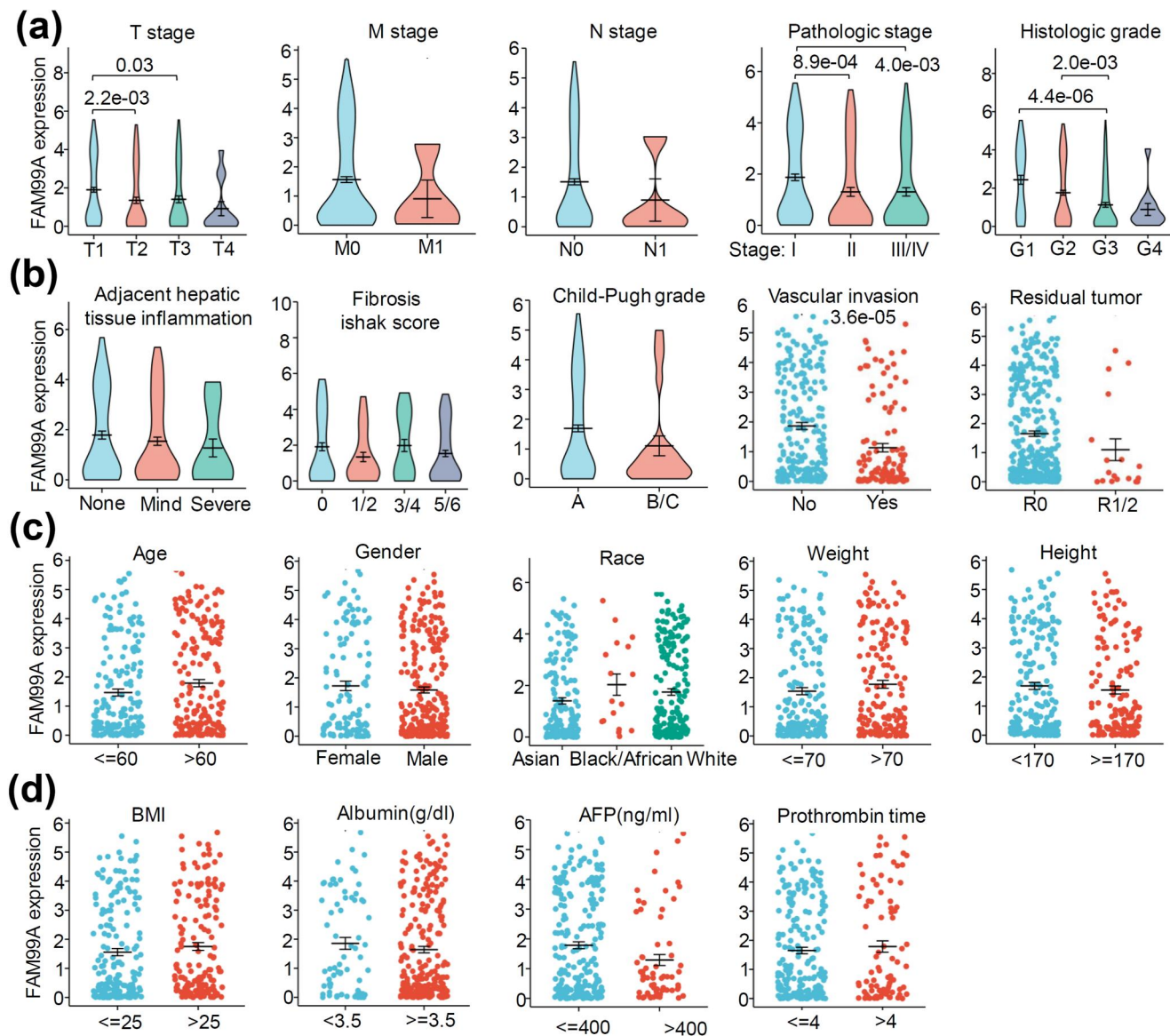


FIGURE 3 Expression correlations between lncRNA *FAM99A* expression and clinical traits in LIHC. Wilcoxon test or Kruskal–Wallis test/Dunn's test was performed to analyse the statistical differences between *FAM99A* expression levels and the clinical traits in LIHC, including (a) T/M/N pathologic stage, histologic grade; (b) adjacent hepatic tissue inflammation, fibrosis ishak score, Child–Pugh graph, vascular invasion, residual tumour; (c) age, gender, race, weight, height; (d) BMI, albumin (g/dL) levels, AFP (ng/mL), and prothrombin time. AFP, alpha fetoprotein; BMI, body mass index; LIHC, liver hepatocellular carcinoma; lncRNA, long non-coding RNA.

according to different clinical factors. The results in Tables S2–S5 suggested that the lowly expressed *FAM99A* was linked to the worse survival in all the subgroups of ‘male’, ‘pathologic stage 3’, ‘grade 3’, and ‘American Joint Committee on Cancer (AJCC)_T3’ (all HR <1, $p < 0.05$), but not subgroup of ‘female’.

Moreover, univariate and multivariate Cox analyses regarding *FAM99A* expression and clinical traits were conducted using the data from GSE76427 and TCGA-LIHC respectively. For the GSE76427 cohort, we failed to observe positive conclusions in both univariate and multivariate Cox analyses (Table S6, all $p > 0.05$). However, it should be noted that only 95 HCC cases and four clinical traits were included in the multivariate Cox analysis. The

result of the multivariate Cox analysis also indicated a negative association between the pathologic stage and poor clinical outcome of HCC patients within the GSE76427 cohort (Table S6, $p > 0.05$). The results of the Cox analysis based on a larger sample size and clinical information can yield more objective and effective conclusions. Thus, a total of 346 hepatic cancer cases and 12 types of clinical traits (e.g. gender, age, pathologic stage, histologic grade, tumour status, adjacent hepatic tissue inflammation, residual tumour, race, fibrosis ishak score etc.) were included in the Cox analyses of TCGA-LIHC cohort. As shown in Table S7, there was a link between low *FAM99A* expression and worse survival outcomes of hepatic cancer patients within TCGA-LIHC cohorts in both univariate (HR = 0.888, $p = 0.027$) and multivariate

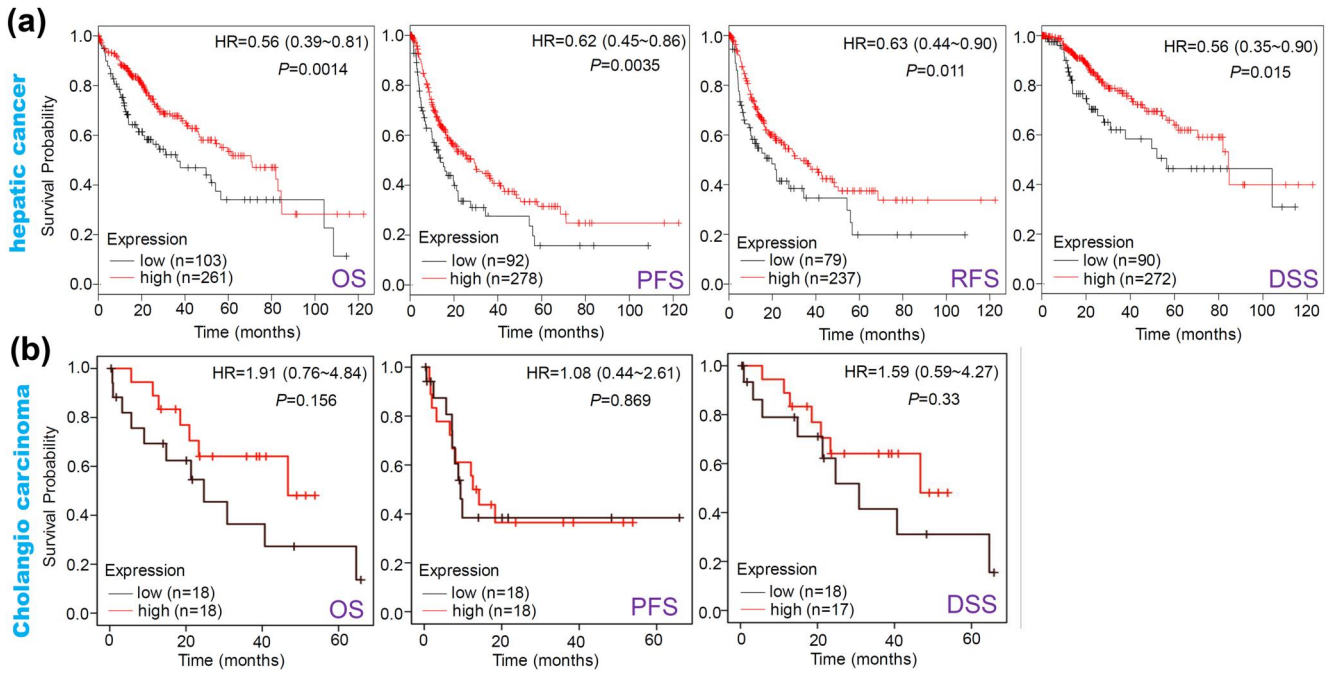


FIGURE 4 Survival curve analysis of lncRNA *FAM99A* in hepatic cancer and cholangiocarcinoma. (a) Analyses of OS, RFS, PFS, and DSS in hepatic cancer cases according to the expression levels of *FAM99A* gene using Kaplan–Meier Plotter. (b) Analyses of OS, PFS, and DSS in cholangiocarcinoma cases using the survival and survminer R packages. DSS, disease-specific survival; HR, hazard ratio; lncRNA, long non-coding RNA; OS, overall survival; PFS, progression-free survival; RFS, relapse-free survival.

(HR = 0.885, $p = 0.042$) Cox analyses. Positive conclusions were obtained for two other clinical factors, namely pathologic stage (Table S7, HR > 1, $p < 0.001$) and tumour status (Table S7, HR > 1, $p < 0.01$).

3.4 | CNV analysis

Next, the CNV status of *FAM99A* and another family member, *FAM99B* (family with sequence similarity 99 member B) was examined. As shown in Figure 5a,b, we did not observe copy number variations in the majority of hepatic cancer cases, and heterozygous amplification/heterozygous deletion in the limited cancer cases. Furthermore, we did not detect a strong correlation between CNV and expression of lncRNA *FAM99A* (Figure 5c). Thus, copy number variations in *FAM99A* may not play an essential role in hepatic tumorigenesis.

3.5 | DNA methylation analysis

We attempted to exploit the potential molecular mechanism of *FAM99A* DNA methylation for LICH, CHOL, and TGCT. Based on TCGA-LIHC methylation data, we found that *FAM99A* gene expression was negatively correlated with the methylation signal values of five methylation probe sites: cg24218935 (Figure 6, $r = -0.397$, $p < 0.001$), cg01745044 ($r = -0.359$, $p < 0.001$), cg04353359 ($r = -0.564$, $p < 0.001$), cg04938738 ($r = -0.421$, $p < 0.001$), and cg25356611 ($r = -0.395$, $p < 0.001$). The similar result was detected for

CHOL (Figure S3, cg04353359, $r = -0.513$, $p < 0.001$; cg04938738, $r = -0.465$, $p < 0.01$; cg25356611, $r = -0.435$, $p < 0.01$). However, we found a positive correlation between the expression level and the four methylation probe sites of *FAM99A* for TGCT (Figure S4, $r > 0$, $p < 0.01$). This suggests a potential role of *FAM99A* DNA methylation in hepatobiliary tumorigenesis.

3.6 | Enrichment analysis

We screened for differential genes between *FAM99A* high and low expression groups. As shown in Figure 7a, 173 differential genes were obtained. Of these, *FAM99B* and *PCSK1* (pro-protein convertase subtilisin/kexin type 1) showed a statistically significant difference (Figure 7a). Based on these genes, we performed the GO (Figure 7b) and KEGG (Figure 7c) enrichment analysis and identified the presence of a group of cellular issues or pathways, such as ‘receptor ligand activity’, ‘carbohydrate binding’, ‘response to metal ion’, ‘protein digestion and absorption’, ‘calcium signalling pathway’. Further, the hallmark gene set-based GSEA data (Figure 7d) indicated the enriched cell cycle-related tissues, including ‘G2M checkpoint’ and ‘MYC targets’. The data of the PPI network (Figure 7e) further suggested that these *FAM99A*-associated differential genes, such as CPA1 (carboxypeptidase A1), PRSS1 (serine protease 1), and PRSS3P2 (PRSS3 pseudogene 2), may interact with each other at the protein level to form an interaction network and participate in the regulation mechanism of *FAM99A*.

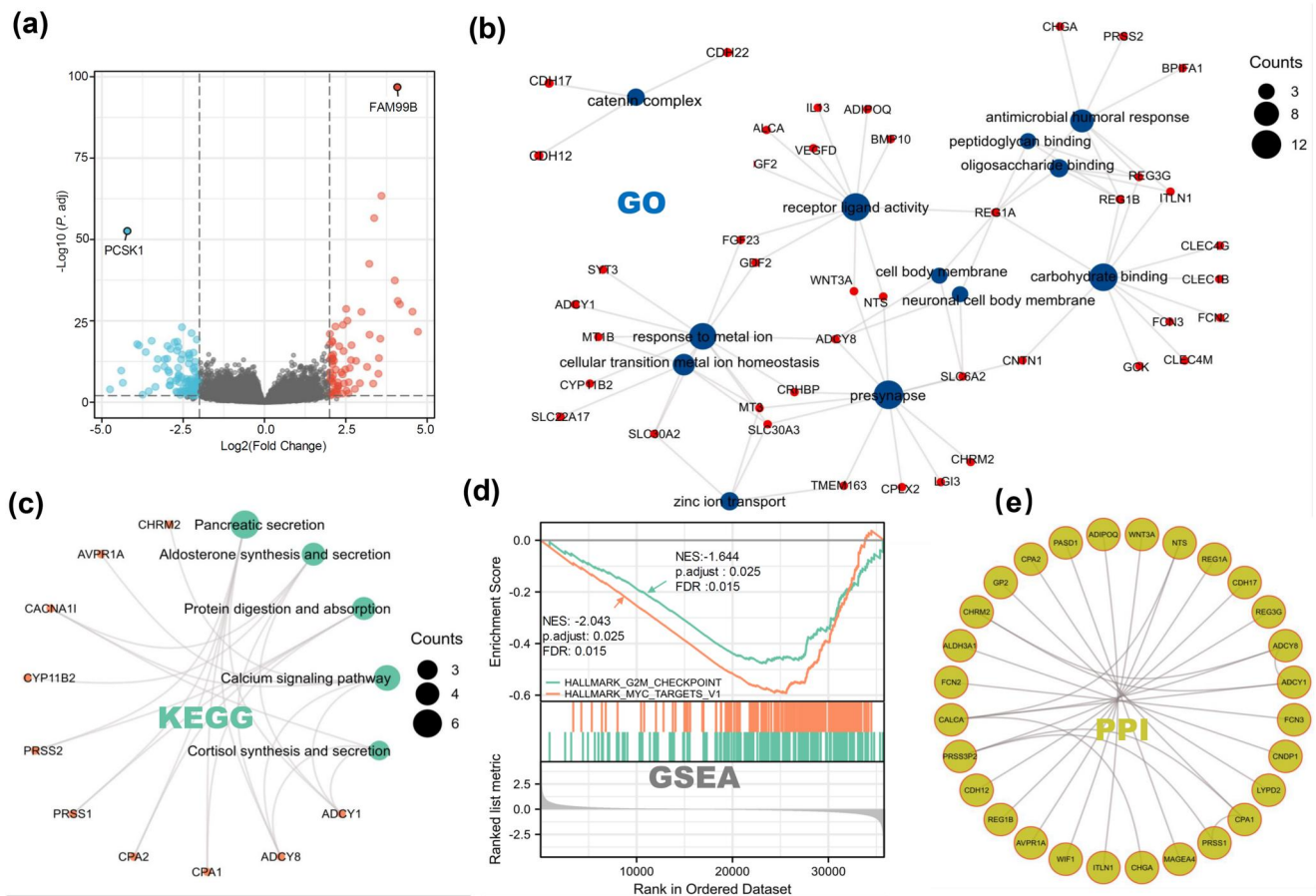


FIGURE 7 Cluster analysis of lncRNA *FAM99A*-associated differential genes. (a) lncRNA *FAM99A*-correlated differential genes for TCGA-LIHC cohort were obtained and volcano plot was created using the DESeq2 and ggplot2 R packages. GO (b), KEGG pathway (c) enrichment analyses and GSEA (d) of *FAM99A*-correlated significant genes were conducted using clusterProfiler and ggplot2 R packages. *FAM99A*-associated PPIs were analysed using the igraph and ggraph R packages (e). GO, Gene Ontology; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; lncRNA, long non-coding RNA; PPIs, protein-protein interactions; TCGA-LIHC, The Cancer Genome Atlas-liver hepatocellular carcinoma.

(Figure 8c). These results suggest a functional link between biochemical metabolism-related events and hepatic tumourigenesis.

3.7 | Immune cell infiltration analysis

Finally, we used the ESTIMATE R package to analyse the potential relationship between *FAM99A* expression and tumour purity. As shown in Figure 9a, we obtained a weak correlation between *FAM99A* expression and the STROMAL score ($r = 0.110$, $p = 0.041$), but not for ESTIMATE and IMMUNE scores (all $p > 0.05$). Furthermore, we observed a statistically significant difference in the STROMAL score between the *FAM99A* high and low groups (Figure 9b, $p = 5.5 \times 10^{-3}$). The ssGSEA data (Figure 9c) further showed a correlation between *FAM99A* expression and the infiltration levels of several immune cells, such as dendritic cell (DC), Tregs, and T helper cells. We observed a similar difference in the infiltration of some immune cells, including B cells, cytotoxic cells, DC, natural killer (NK) cells, neutrophils, Th17 cells, and Tregs (Figure 9d, $p < 0.05$). Detailed Spearman

correlation data are shown in Figure S5. A positive correlation was detected for DC ($p < 0.001$), Tregs ($p < 0.001$), Th17 cells ($p < 0.001$), B cells ($p = 0.002$), cytotoxic cells ($p = 0.003$), and NK cells ($p = 0.005$). These results suggest a potential link between *FAM99A* expression and the infiltration status of certain immune cells.

4 | DISCUSSION

Using the available cancer datasets from TCGA, GSE76427, and GSE64041, we discovered that lncRNA *FAM99A* is mainly expressed in liver-related cancers, namely HCC and CHOL. Compared to the adjacent controls, lncRNA *FAM99A* is expressed at low levels in hepatic cancer tissues, suggesting that *FAM99A* may be a liver-specific tumour suppressor gene. Nevertheless, only 36 CHOL tissues and five adjacent control tissues were considered in TCGA-CHOL project. In addition, we did not obtain positive results in the clinical prognostic analysis. For TGCT, there showed the lowly expressed *FAM99A* in the tumour tissues, compared with the normal tissues of GTEx. Normal controls were not available for

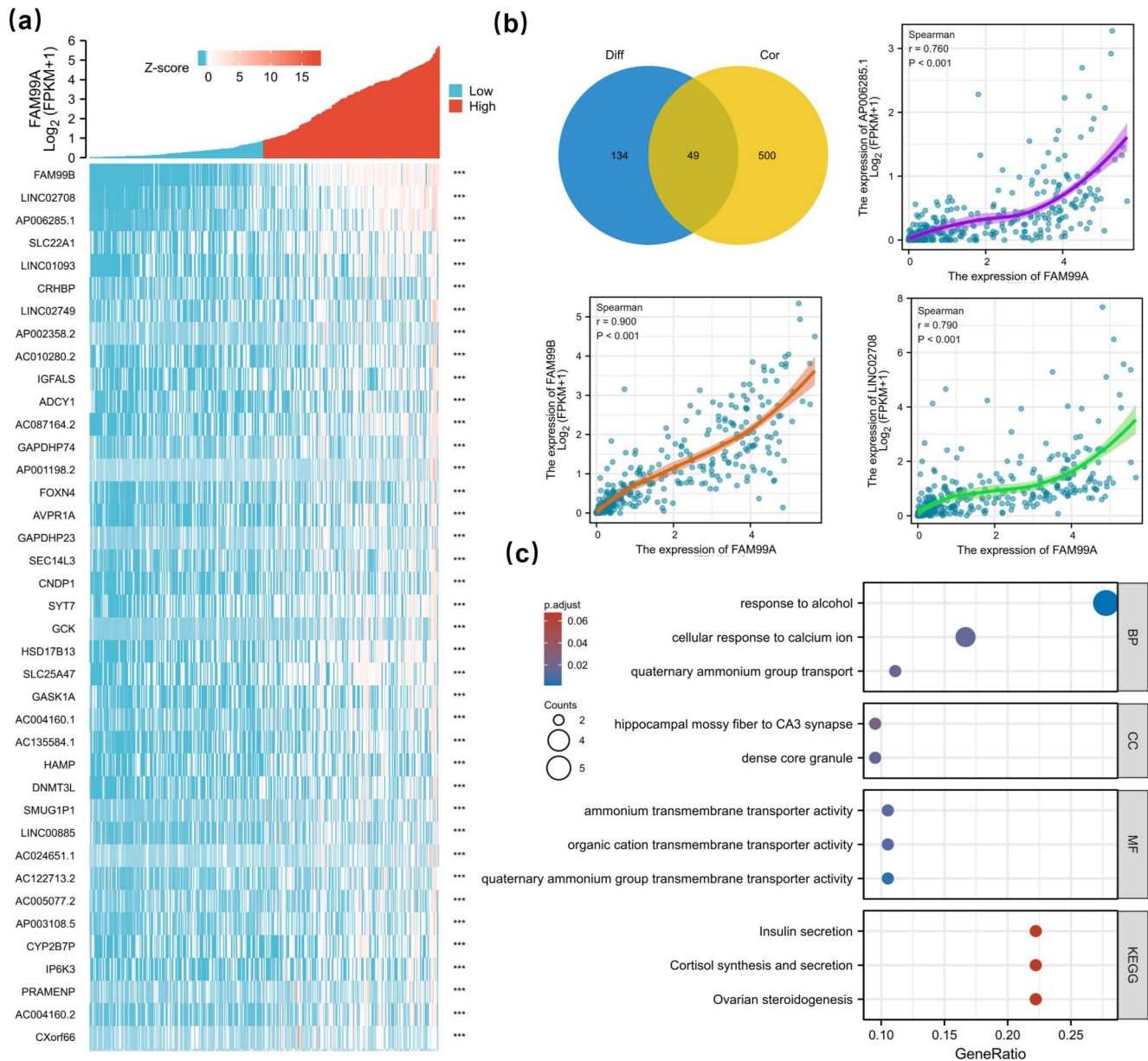


FIGURE 8 Cluster analysis of lncRNA *FAM99A*-correlated genes. We performed Spearman correlation analysis of *FAM99A* using the stat R package. Heat map targeting the *FAM99A*-positively or -negatively correlated genes was plotted using a ggplot2 R package (a). (b) Venn diagram between the above differential genes (Diff) and Spearman correlated genes (Cor) of *FAM99A* was plotted using the ggplot2 R package. Shapiro–Wilk normality test was performed to analyse the correlation between *FAM99A* and some target genes, including *FAM99B*, *AP006285.1*, and *LINC02708*. (c) Based on the common genes, the GO/KEGG pathway enrichment analysis was conducted using the clusterProfiler and ggplot2 R packages. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; lncRNA, long non-coding RNA.

TCGA project. Therefore, in this study, we focussed only on the correlation between lncRNA *FAM99A* and CHOL. Despite this, we cannot rule out the potential regulatory role of lncRNA *FAM99A* in the initiation and progression of CHOL, considering the link between non-coding RNAs and CHOL [27]. Our Kaplan–Meier and Cox survival analysis data indicated a statistical correlation between low expression of the *FAM99A* gene and poorer prognosis status of hepatic cancer patients, which is in line with the conclusion of one publication [12]. In addition, we found that there was a statistical expression reference for *FAM99A* among different pathologic stages

(stages I–IV). When hepatic cancer samples were grouped according to the clinical information, the positive association between lowly-expressed *FAM99A* and poor survival outcomes exists in the subgroups of ‘pathologic stage 3’, ‘grade 3’, ‘AJCC_T3’. It is important to note that we observed a correlation between *FAM99A* gene expression and the clinical prognosis of male hepatic cases, but not in female patients. This suggests that the prognostic warning ability of the lowly expressed *FAM99A* gene may increase with the tumour differentiation process or pathological state in male patients with hepatic cancer.

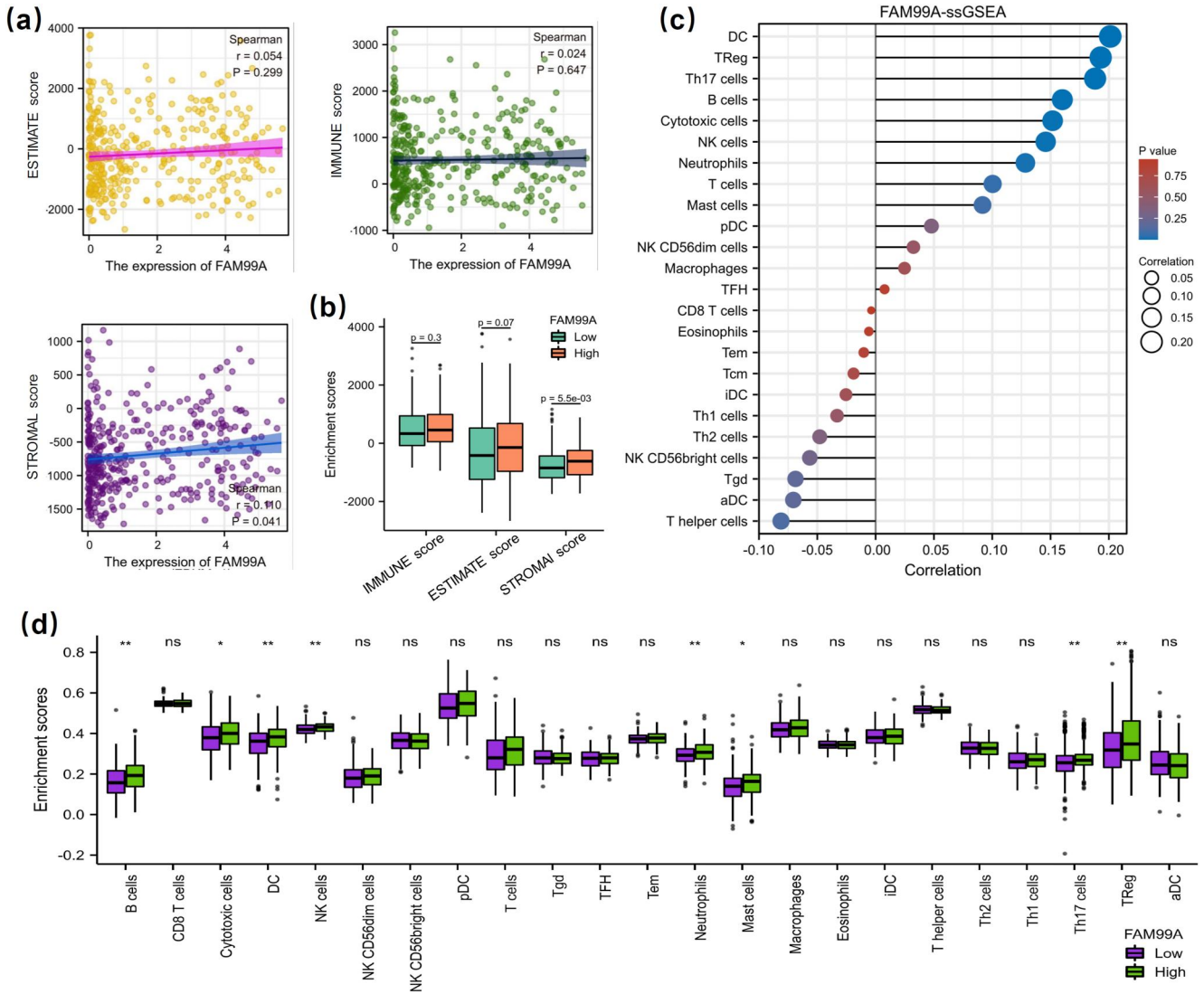


FIGURE 9 Correlation between lncRNA *FAM99A* expression and tumour purity/immune cell infiltration. We evaluated the tumour purity and determined the STROMAL, IMMUNE, and ESTIMATE scores using the ‘ESTIMATE’ R package. (a) Spearman correlations between *FAM99A* levels and three indices of tumour purity were determined. (b) Wilcoxon test was used to analyse the statistical differences between high and low expression groups of *FAM99A*. (c) Through ssGSEA using the GSEA R package, the *FAM99A* expression and tumour infiltration levels in the immune cells, including DCs, B cells, CD8 T cells, cytotoxic cells, eosinophils, iDCs, macrophages, mast cells, neutrophils, NK CD56 bright cells, NK CD56 dim cells, NK cells, pDC, T cells, T helper cells, Tcm cells, Tem cells, Tfh cells, Tgd cells, Th1 cells, Th17 cells, Th2 cells, and Tregs, were analysed. (d) Wilcoxon test was performed to analyse the infiltration differences of immune cells between the high and low expression groups of *FAM99A*. * $p < 0.05$, ** $p < 0.01$, ns $p \geq 0.05$. DCs, dendritic cells; NK, natural killer; ssGSEA, single-sample gene set enrichment analysis.

LncRNA *FAM99A* rs1489945 was reported to be linked to maternal mean arterial blood pressure in a Cambridge birth cohort [9]. Therefore, we explored the mutation and CNV status of *FAM99A* in this study. Our findings showed a very low genetic mutation frequency of *FAM99A* in cancer, which did not significantly correlate with the gene expression or clinical prognosis (data not shown). We also did not observe a high frequency of CNV or a strong correlation between *FAM99A* expression and CNV. In addition, considering the links between cellular immune responses and hepatic cancer [28], we also analysed the correlation between lncRNA *FAM99A* expression and the signatures of the following immune cells: central memory T cells, effector memory T cells,

effector T cells, effector Tregs, exhausted T cells, native T cells, Th1 like cells, and resting Tregs. However, no strong correlations were observed.

The DNA methylation status of RNA is closely related to gene expression and carcinogenesis in hepatic cancer [29, 30]. Eukaryotic lncRNAs also participate in the metastasis and prognosis of HCC through regulating chromatin remodelling and methylation [31, 32]. High methylation levels of the five methylated probe sites (cg24218935, cg01745044, cg04353359, cg04938738, and cg25356611) were negatively correlated with low expression levels of *FAM99A*. We found that the cg24218935 and cg04353359 sites were in the promoter region, whereas cg01745044, cg04938738, and cg25356611 were in the

non-promoter region. It is worthwhile to further explore the synergistic role of different methylation sites of *FAM99A* in the expression level and survival prognosis of patients with hepatic cancer.

As a downregulated gene in preeclampsia, *FAM99A* participates in the regulation of invasion, migration, and apoptosis of trophoblasts [10]. We analysed several genes related to *FAM99A* expression. Among these, we observed a high degree of expression consistency between *FAM99A* and *FAM99B*. The lncRNA *FAM99B* was reported to inhibit cellular proliferation, migration, and invasion in only one study [33]. Such cellular function attributes may also be involved in the role of *FAM99A* in hepatic tumourigenesis and progression. In addition, we performed enrichment analyses based on *FAM99A* expression-related genes. A set of biological events, including the ‘receptor ligand activity’, the ‘protein digestion and absorption’, ‘G2M checkpoint’, ‘response to alcohol’, and the ‘calcium signalling pathway’, were obtained. The molecular mechanism underlying the role of DNA methylation or competing endogenous RNA networks of *FAM99A* in the above biological activities require further investigation.

Our study has some limitations. We only conducted bioinformatic analysis on the available samples collected from TCGA and GEO databases, without further exploration of the underlying mechanisms in cell and animal experiments. Very few clinicopathological parameters were included in the Cox survival analysis of GSE76247. Although there were statistical differences, we only observed a weak relationship between *FAM99A* expression and the infiltration status of several immune cells, such as DCs, NK cells, and neutrophils. Additional experimental evidence is required to support this correlation. Moreover, we could not conduct a more in-depth analysis of *FAM99A* expression in TGCT and CHOL, which requires further elucidation.

To the best of our knowledge, this is the first study to perform an *in silico* investigation of the expression patterns of lncRNA *FAM99A* in 33 types of cancer. *FAM99A* was highly expressed in normal cholangiocytes and hepatocytes and downregulated in HCC and CHOL. The lowly expressed lncRNA *FAM99A* was further identified as a prognostic gene for HCC. Several methylation sites were significantly correlated with *FAM99A* expression. Mechanistically, various cellular biochemical metabolic or cell cycle-related pathways and the infiltration of immune cells, such as DCs, NK cells, and neutrophils, were correlated with *FAM99A* expression in hepatic tumourigenesis. However, these findings need to be validated using cell molecular experiments in the future.

AUTHOR CONTRIBUTIONS

Conceptualisation: Manyi Sun. Data curation: Shuhua Lv, Jin Zhong. Formal analysis: Manyi Sun, Shuhua Lv. Methodology: Manyi Sun, Jin Zhong. Software: Manyi Sun, Shuhua Lv, Jin Zhong. Supervision: Jin Zhong. Validation: Jin Zhong. Visualization: Manyi Sun, Shuhua Lv, Jin Zhong. Writing—original draft: Manyi Sun. Writing—review & editing: Shuhua Lv, Jin Zhong. Approval of final manuscript: all authors.

ACKNOWLEDGEMENTS

We would like to acknowledge TCGA, GEO, and GTEX databases for providing the datasets analysed in this study. We also appreciate GEPIA2, MEXPRESS, GSCALite, and XIANTAO Academic for their aid in data analysis. A previous version of this manuscript, which contained data on HCC, was presented as a preprint on Research Square (<https://www.researchsquare.com/article/rs-16068/v1>).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data used in this study were obtained from the open datasets of TCGA (<https://portal.gdc.cancer.gov/>), GEO (<https://www.ncbi.nlm.nih.gov/geo/>), and UCSC XENA (<https://xenabrowser.net/datapages/>). All data generated or analysed are included in this study and supplementary files.

PATIENT CONSENT STATEMENT

Not applicable.

ORCID

Manyi Sun  <https://orcid.org/0000-0001-6647-0214>

REFERENCES

- Bhan, A., Soleimani, M., Mandal, S.S.: Long noncoding RNA and cancer: a new paradigm. *Cancer Res.* 77(15), 3965–3981 (2017). <https://doi.org/10.1158/0008-5472.can-16-2634>
- Mai, H., et al.: Molecular pattern of lncRNAs in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* 38(1), 198 (2019). <https://doi.org/10.1186/s13046-019-1213-0>
- Chi, Y., et al.: Long non-coding RNA in the pathogenesis of cancers. *Cells* 8(9), 1015 (2019). <https://doi.org/10.3390/cells8091015>
- Shi, H., et al.: Current research progress on long noncoding RNAs associated with hepatocellular carcinoma. *Anal. Cell Pathol.* 2019, 1534607 (2019). <https://doi.org/10.1155/2019/1534607>
- Liu, J., et al.: Identification of key genes and long non-coding RNA associated ceRNA networks in hepatocellular carcinoma. *PeerJ* 7, e8021 (2019). <https://doi.org/10.7717/peerj.8021>
- Wang, Z., Jensen, M.A., Zenklusen, J.C.: A practical guide to The Cancer Genome Atlas (TCGA). *Methods Mol. Biol.* 1418, 111–141 (2016)
- Tomczak, K., Czerwinska, P., Wiznerowicz, M.: The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* 19(1a), A68–A77 (2015)
- Mo, M., et al.: Liver-specific lncRNA *FAM99A* may be a tumor suppressor and promising prognostic biomarker in hepatocellular carcinoma. *BMC Cancer.* 22(1), 1098 <https://doi.org/10.1186/s12885-022-10186-2> (2022)
- Petry, C.J., et al.: Associations between the maternal circulating lipid profile in pregnancy and fetal imprinted gene alleles: a cohort study. *Reprod. Biol. Endocrinol.* 16(1), 82 (2018). <https://doi.org/10.1186/s12958-018-0399-x>
- He, T., et al.: lncRNA *FAM99A* is downregulated in preeclampsia and exerts a regulatory effect on trophoblast cell invasion, migration and apoptosis. *Mol. Med. Rep.* 20(2), 1451–1458 (2019)
- Petry, C.J., et al.: Associations between fetal imprinted genes and maternal blood pressure in pregnancy. *Hypertension* 68(6), 1459–1466 (2016). <https://doi.org/10.1161/hypertensionaha.116.08261>
- Zhao, B., et al.: HIF-1 α and HDAC1 mediated regulation of *FAM99A*-miR92a signaling contributes to hypoxia induced HCC metastasis. *Signal Transduct. Targeted Ther.* 5(1), 118 (2020). <https://doi.org/10.1038/s41392-020-00223-6>

13. Zheng, X., et al.: Icaritin-induced FAM99A affects GLUT1-mediated glycolysis via regulating the JAK2/STAT3 pathway in hepatocellular carcinoma. *Front. Oncol.* 11, 740557 (2021). <https://doi.org/10.3389/fonc.2021.740557>
14. Tang, Z., et al.: GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 47(W1), W556–w560 (2019). <https://doi.org/10.1093/nar/gkx430>
15. Tang, Z., et al.: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45(W1), W98–w102 (2017). <https://doi.org/10.1093/nar/gkx247>
16. Vivian, J., et al.: Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* 35(4), 314–316 (2017). <https://doi.org/10.1038/nbt.3772>
17. Menyhart, O., Nagy, A., Gyorffy, B.: Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. *R. Soc. Open Sci.* 5(12), 181006 (2018). <https://doi.org/10.1098/rsos.181006>
18. Liu, C.J., et al.: GSCALite: a web server for gene set cancer analysis. *Bioinformatics* 34(21), 3771–3772 (2018). <https://doi.org/10.1093/bioinformatics/bty411>
19. Koch, A., et al.: MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genom.* 16(1), 636 <https://doi.org/10.1186/s12864-015-1847-z>(2015)
20. Koch, A., et al.: MEXPRESS update 2019. *Nucleic Acids Res.* 47(W1), W561–W565 (2019). <https://doi.org/10.1093/nar/gkx445>
21. Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12), 550 (2014). <https://doi.org/10.1186/s13059-014-0550-8>
22. Yu, G., et al.: clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16(5), 284–287 (2012). <https://doi.org/10.1089/omi.2011.0118>
23. Szklarczyk, D., et al.: STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47(D1), D607–d613 (2019). <https://doi.org/10.1093/nar/gky1131>
24. Yoshihara, K., et al.: Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4(1), 2612 (2013). <https://doi.org/10.1038/ncomms3612>
25. Hänzelmann, S., Castelo, R., Guinney, J.: GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 14(1), 7 (2013). <https://doi.org/10.1186/1471-2105-14-7>
26. Bindea, G., et al.: Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39(4), 782–795 (2013). <https://doi.org/10.1016/j.immuni.2013.10.003>
27. Salati, M., Braconi, C.: Noncoding RNA in cholangiocarcinoma. *Semin. Liver Dis.* 39(1), 13–25 <https://doi.org/10.1055/s-0038-1676097>(2019)
28. Schmidt, N., Neumann-Haefelin, C., Thimme, R.: Cellular immune responses to hepatocellular carcinoma: lessons for immunotherapy. *Dig. Dis.* 30(5), 483–491 (2012). <https://doi.org/10.1159/000341697>
29. Raggi, C., Invernizzi, P.: Methylation and liver cancer. *Clin. Res. Hepatol. Gastroenterol.* 37(6), 564–571 (2013). <https://doi.org/10.1016/j.clinre.2013.05.009>
30. Yuan, S.X., et al.: Long noncoding RNA, the methylation of genomic elements and their emerging crosstalk in hepatocellular carcinoma. *Cancer Lett.* 379(2), 239–244 (2016). <https://doi.org/10.1016/j.canlet.2015.08.008>
31. Abbastabar, M., et al.: lncRNA involvement in hepatocellular carcinoma metastasis and prognosis. *EXCLI J.* 17, 900–913 (2018)
32. Lim, L.J., et al.: Roles and regulation of long noncoding RNAs in hepatocellular carcinoma. *Cancer Res.* 79(20), 5131–5139 (2019). <https://doi.org/10.1158/0008-5472.can-19-0255>
33. Mo, M., et al.: A liver-specific lncRNA, FAM99B, suppresses hepatocellular carcinoma progression through inhibition of cell proliferation, migration, and invasion. *J. Cancer Res. Clin. Oncol.* 145(8), 2027–2038 (2019). <https://doi.org/10.1007/s00432-019-02954-8>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sun, M., Lv, S., Zhong, J.: In silico analysis of the association between long non-coding RNA family with sequence similarity 99 member A (*FAM99A*) and hepatic cancer. *IET Syst. Biol.* 17(2), 83–94 (2023). <https://doi.org/10.1049/syb2.12062>