Up-regulated Expression and Related Gene Regulatory Networks of ANLN Predict Poor Prognosis and Correlate with Immune Infiltration in Hepatocellular Carcinoma

Bo Hu  
Peking Union Medical College Hospital

Xiao-Bo Yang  
Peking Union Medical College Hospital

Xinting Sang (✉ sangxt@pumch.cn)  
Peking Union Medical College Hospital  https://orcid.org/0000-0003-1952-0527

Primary research

Keywords: Anillin, Hepatocellular Carcinoma, Prognosis, Immune System

DOI: https://doi.org/10.21203/rs.3.rs-31170/v1

License: ☇️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: The aberrant Anillin (ANLN) expression is reported to be associated with carcinogenesis. In this study, sequencing data collected from the Cancer Genome Atlas database were utilized to analyze ANLN expression in hepatocellular carcinoma (HCC).

Methods: The relationships of clinicopathological features with ANLN were investigated, and gene set enrichment analysis (GSEA) was performed to reveal the ANLN-related functions. LinkedOmics was employed to identify the co-expressed genes of ANLN and to examine the target networks of kinases, microRNAs (miRNAs) and transcription factors (TFs). Besides, the correlation of ANLN expression with cancer immune infiltrates was analyzed by TIMER.

Results: ANLN over-expression predicted dismal prognosis, and GESA results revealed several functions that were related to cell cycle and mRNA binding. Moreover, functional network analysis indicated that, ANLN might regulate DNA replication and cell cycle signaling through pathways that involved several cancer-related kinases, miRNAs and E2F1. Additionally, ANLN was suggested to be associated with the infiltration of several immune cells, which was proved to be upregulated in both HCC cells and tissues.

Conclusion: Those efficiently mined data reveal information regarding ANLN expression, the potential regulatory networks and the relationship with immune infiltration in HCC, which lay a foundation for further study on the role of ANLN in carcinogenesis.

1. Background

Hepatocellular carcinoma (HCC) represents the second leading cause of cancer-related deaths worldwide, which affects about 800,000 cases annually[1]. Generally, treatments for HCC include surgical resection and chemotherapy, but the mortality rate remains quite high [2], and the prognosis for HCC patients is affected by various factors. Besides, the tumor frequently relapses in HCC patients even after they have received radical surgery. Therefore, it is helpful to further understand the HCC-related molecular mechanism and further investigate the HCC oncogenes, so as to develop the novel diagnostic and therapeutic strategies.

Anillin (ANLN), a phylogenetically conserved protein, is first identified in Drosophila, and it interacts with cytoskeletal components together with their regulators[3]. Located on chromosome 7p14.2, ANLN encodes an actin-binding protein consisting of 1125 amino acids, and it plays an important role in cytokinesis[4]. Remarkably, ANLN expression is also reported to be disordered during HCC development and progression. Lian et al. elaborated that ANLN was essential for HCC tumor growth, and that it was regulated by miR-15a/miR-16-1[5]. Meanwhile, another study also suggests that, ANLN expression in HCC is up-regulated, and the positive protein expression indicates dismal prognosis for the long-term survival of patients undergoing liver transplantation[6]. However, the underlying role of ANLN in human HCC has not been elucidated yet.
This study comprehensively analyzed \textit{ANLN} expression and its correlations with clinicopathological features as well as the prognosis for HCC patients based on electronic databases, including TCGA, HCCDB, UALCAN and Kaplan-Meier plotter. Using the multi-dimensional analysis methods, the functional networks related to \textit{ANLN} were analyzed in HCC cases. Furthermore, \textit{ANLN} expression in HCC cells and tissues was also detected. Thus, our results potentially revealed the novel targets and strategies for the diagnosis and treatment of HCC.

2. Materials And Methods

2.1. Acquisition of clinical samples and RNA-sequencing patient data

The level 3 transcriptomic RNA-sequencing data of HCC samples were downloaded from TCGA data portal (https://cancergenome.nih.gov/), which included data from 374 primary HCC and 50 non-tumor tissues, to extract the \textit{ANLN} expression levels. In addition, the clinicopathological characteristics were also collected from these patients, including age, gender, pathological stage (grade), clinical stage and other parameters. The \textit{ANLN} gene expression levels in various cancer types were identified in the Oncomine database (https://www.oncomine.org/resource/login.html)[23] according to the following thresholds, \textit{P}-value of 0.001, fold change of 2, and gene ranking of all. In addition, the HCCDB database (http://lifeome.net/database/hccdb/about.html) was also used for broader validation. HCCDB is a one-stop online resource to explore HCC gene expression, which contains 15 public HCC expression datasets with up to about 4000 clinical samples [24].

2.2. Survival curve plotting and clinical correlation analysis

Clinical data downloaded from TCGA data portal were collected to extract the overall survival (OS), and then the survival curves were further plotted using the “survival” function of R package, whereas the time-dependent prognostic value of the specific gene was investigated using the “survivalROC” function of R package[25]. Afterwards, both univariate and multivariate Cox proportional hazard regression analyses were performed to confirm whether the specific gene predicted patient prognosis in a conventional clinical factors-independent manner. Moreover, the Kaplan-Meier plotter, which assesses the impacts of 54 k genes on survival in 21 cancer types (http://kmplot.com/analysis/), was also carried out to assist in estimating the effect of \textit{ANLN} on survival[26]. Typically, the largest datasets in the Kaplan-Meier plotter include breast cancer (n=6,234), ovarian cancer (n=2,190), lung cancer (n=3,452), and gastric cancer (n=1,440). The gene chip and RNA-seq data sources for GEO, EGA, and TCGA databases are included in the Kaplan-Meier plotter. In this study, the hazard ratios (HRs), together with the corresponding 95% confidence intervals (CIs) and log-rank \textit{P}-value, were also determined. There were few patients at stage IV, as a result, patients at stage III and stage IV were added up for analysis. Moreover, HCCDB was adopted to verify the effect of \textit{ANLN} expression from the International Cancer Genome Consortium (ICGC)-liver cancer (LIRI)-JP on the patient OS.

Additionally, the relationships of clinicopathological parameters with \textit{ANLN} expression were further examined to reveal the underlying clinical significance of \textit{ANLN} via the UALCAN database. In particular,
UALCAN, which uses the TCGA level 3 RNA-seq and clinical data from 31 cancer types, is an interactive web-portal to perform in-depth analyses on TCGA gene expression data, and it is also utilized to assist in analysis[27]. One of the user-friendly features of the portal is that it allows to analyze the relative expression of a query gene(s) across tumor and normal samples, as well as in various tumor sub-groups stratified based on individual cancer stage, tumor grade or other clinicopathological features. UALCAN is publicly available at http://ualcan.path.uab.edu.

2.3. Gene set enrichment analysis (GSEA)

GSEA is a computational approach to determine whether a previously defined set of genes show statistically significant and concordant differences between two biological states[28]. In our study, GSEA was conducted to illustrate the differences in gene function between high- and low-ANLN groups, with ANLN expression being used as a phenotype label. The gene set permutations were performed 1000 times in each analysis. Besides, the nominal p-value, false discovery rate (FDR) and enrichment score (ES) were adopted for sorting the pathways enriched in each phenotype.

2.4. LinkedOmics and GeneMANIA analyses

The LinkedOmics database (http://www.linkedomics.org/login.php) is a Web-based platform to analyze 32 TCGA cancer-associated multi-dimensional datasets[29]. In this study, the LinkFinder module of LinkedOmics was adopted to investigate the ANLN-associated differentially expressed genes (DEGs) in the TCGA LIHC cohort (n=371). Results were statistically analyzed using the Pearson correlation coefficient, and the statistical plots for individual genes were also drawn. All results were graphically presented in volcano plots, heat maps or scatter plots. Further, the LinkInterpreter module of LinkedOmics was utilized to analyze the pathways and networks related to the DEGs. Afterwards, data from LinkFinder results were signed and ranked, and GSEA was performed to analyze the kinase-target enrichment, miRNA-target enrichment and TF-target enrichment, among which, the latter two network analyses were carried out based on the Molecular Signatures Database (MSigDB)[30]. The rank criterion was set at \( P < 0.05 \), and 1000 simulations were run. On the other hand, GeneMANIA (http://www.genemania.org) is a flexible and user-friendly web interface to construct the protein-protein interaction (PPI) network, which generates hypotheses about gene function, analyzes gene lists, and prioritizes genes for functional assays[31]. The website is able to set the source of network edge, and it features several bioinformatics methods, including physical interaction, gene co-expression, gene co-location, gene enrichment analysis, and website prediction. Therefore, GeneMANIA was conducted to visualize the gene networks and to predict the functions of genes identified to be enriched in HCC by GSEA.

2.5. TIMER database analysis

TIMER is a web resource to systemically evaluate the clinical impacts of different immune cells in diverse cancer types, which covers 10,879 samples across 32 cancer types from TCGA to estimate the abundances of six tumor-infiltrating immune cell subtypes, including B cells, CD4 T cells, CD8 T cells, macrophages, neutrophils, and dendritic cells (https://cistrome.shinyapps.io/timer/)[32]. In this study, the
correlations of \textit{ANLN} expression with the abundances of immune infiltrates in HCC, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells, were analyzed based on the gene modules.

2.6. Experimental validation of the molecular biological function of \textit{ANLN}

2.6.1. Cell Culture

The HCC cell line (Hep3B) and normal hepatic cell line (LO2) were purchased from the Cell Resource Center of the Shanghai Institutes for Sciences (Shanghai, China). Then, the Hep3B cells were cultured in the Dulbecco’s modified Eagle’s medium (DMEM, Gibco BRL, MD, USA) supplemented with 10% fetal bovine serum (FBS, Gibco BRL, MD, USA) and incubated under 5% CO2 and at 37°C conditions.

2.6.2. Quantitative real-time polymerase chain reaction (qRT-PCR)

The total cellular RNA was extracted using the Trizol reagent (Beijing Solarbio Science & Technology Co. Ltd). Afterwards, the concentration and purity of total RNA were determined by an Eppendorf instrument, and the degradation of RNA was detected through agarose gel electrophoresis (AGE). Then, cDNA was prepared through reverse transcription of RNA using the Universe RT-PCR Kit (Beijing Solarbio Science & Technology Co. Ltd) according to manufacturer instructions. Table 1 shows the primer sequences of \textit{ANLN} and \textit{ACTB}. Subsequently, 2X Master Mix, the forward and reverse primers for the gene, cDNA, and Rnase Free ddH$_2$O, were added and mixed via the Arraystar Real-Time PCR Kit in accordance with manufacturer protocols. The PCR amplification program was set as follows: 95 °C for 3 min, followed by 40 cycles of 95 °C for 30 s and 55 °C for 20 s, and finally 72 °C for 20 s. After completing the amplification reaction, a melt curve of PCR product was plotted (95 °C for 15 s; 60 °C for 15 s; and 95 °C for 15 s). The \textit{ANLN} expression level was normalized to the internal control β-actin and calculated according to the $2^{-\Delta \Delta CT}$ method.

2.6.3. Western blotting

Total cellular protein was extracted and the protein concentration was measured using the BCA (bicinchoninic acid) Protein Quantitation Kit (Beyotime Biotechnology Co. Ltd) following the manufacturer protocols. Thereafter, the protein samples were separated by 10% SDS-PAGE and transferred onto the polyvinylidene difluoride (PVDF) membranes (Millipore, Bedford, MA, USA) at a constant current of 100 mA. Later, the PVDF membranes were blocked with 3% BSA on a shaker at room temperature for 2h and incubated with anti-\textit{ANLN} antibody (1:500 dilution) or anti-β-actin antibody (1:1000 dilution) at 4 °C overnight. After washing with TBST, the membranes were incubated with the appropriate anti-IgG antibody (1:1000 dilution) on a shaker at room temperature for 2 h. After washing with TBST, the bands were detected using an enhanced chemiluminescence reagent and visualized based on the X-ray films in dark. The ImageJ software was employed to determine the grayscale value of the protein band.
2.6.4. Immunohistochemistry (IHC)

The 3 mm tumor sections were incubated with the commercial anti-ANLC mouse polyclonal antibodies (Abcam, 1:200 dilution) for 2 h at room temperature. After rinsing with PBS buffer, the HRP-conjugated goat anti-mouse IgG antibody (Biosharp Biotech Co. LTD) was added dropwise to incubate at room temperature for 20 min. After rinsing, the diaminobenzidine substrate solution (DAB kit) was added to the tumor sections for visualization. Specific staining of tissue sections was observed under a microscope (Olympus), and the number of positive cells was counted from at least five fields of view (FOV) of every region.

2.7. Statistical analysis

The R (v.3.6.1) software was employed for all statistical analyses. The results produced from Oncomine were expressed as \( P \)-values, fold changes, and ranks. The results of Kaplan-Meier plots and GEPIA were presented as HR and \( P \)-value or Cox \( P \)-values from a log-rank test. Pearson chi-square test or Fisher’s exact test was adopted for exploring the qualitative variables as appropriate. \( P < 0.05 \) was considered as statistically significant, unless otherwise specified.

3. Results

3.1. The \( ANLN \) mRNA expression levels in HCC and other human cancers

The \( ANLN \) mRNA levels in different tumors and in the normal tissues of multiple cancer types were analyzed using the Oncomine (Fig. 1A) and TIMER (Fig. 1B) databases, with an aim to reveal the aberrant expression of \( ANLN \) from a holistic perspective. To determine the differences in \( ANLN \) expression between HCC and normal samples, the differential scatter plot (Fig. 1C) and paired plot (Fig. 1D) were drawn. Also, \( ANLN \) expression in different GEO datasets was also analyzed via the HCCDB database (Fig. 1E and Table 2). Our results suggested that, \( ANLN \) expression was upregulated in HCC among different datasets, as well as in bladder, breast, biliary, colorectal, head and neck, esophageal, gastric, kidney, lung, prostatic and rectal cancers compared with that in normal tissues.

3.2. Prognostic value of \( ANLN \) in HCC

As displayed in Fig. 2A, results of survival analysis showed that HCC with high \( ANLN \) expression was associated with worse prognosis than that with low \( ANLN \) expression (\( p < 0.001 \)). Similar results were also obtained from the analysis of HCC samples in ICGC-LIRI-JP (Fig. 2B). In addition, the relationships among \( ANLN \) expression, clinical characteristics and survival prognosis for HCC patients were also investigated in the Kaplan-Meier plotter databases (Table 3). According to the results, \( ANLN \) over-expression was related to the worse OS in male and female patients, white and Asian patients, as well as patients with different clinical stages and T stages according to the American Joint Committee on Cancer (AJCC), none or micro-vascular invasion status, alcohol consumption, and hepatitis virus infection (\( P < 0.05 \)). Specifically, the effects of high \( ANLN \) expression on OS were particularly significant in Asian
(hazard ratio [HR] =5.66, P<0.001), male (HR=2.43, P<0.001) and G1 (HR=6.52, P<0.001) patients.
Besides, the values of area under the receiver operating characteristic (ROC) curve (AUC) for 1-, 3- and 5-year OS were 0.679, 0.587 and 0.602, respectively (Fig. 2C).

Moreover, univariate analysis revealed that high ANLN expression was remarkably correlated with poor OS (HR: 1.305; 95% confidence interval [CI]: 1.095–1.556; p < 0.01) (Table 4). Additionally, race was another clinicopathological variable associated with poor survival. Meanwhile, ANLN (HR: 1.584; CI: 1.158–2.167, p < 0.01) and race (HR: 4.651; CI: 1.778–12.152, p < 0.01) remained the factors associated with OS upon multivariate analysis.

3.3. Associations of ANLN expression with clinicopathological variables

Altogether 371 HCC samples with available ANLN expression data stratified based on multiple patient characteristics were analyzed from TCGA. As shown in Figure 3A–3G, the transcription levels of ANLN were notably higher in HCC patients than in healthy people upon subgroup analyses stratified according to gender, age, weight and obese, ethnicity, disease stage, tumor grade and nodal metastasis status. In addition, the logistic regression univariate analysis demonstrated that ANLN expression, which served as a categorical dependent variable, was linked with poor prognostic clinicopathological characteristics (Table 5). Besides, the up-regulated ANLN expression in HCC was evidently associated with high grade (4.82 for G3 vs. G1, 4.88 for G4 vs. G1), stage (OR = 1.68 for II vs. I, 2.14 for III vs. I), and T stage (OR = 1.73 for T2 vs. T1, 1.95 for T3 vs. T1). These findings demonstrated that HCC with high ANLN expression was more inclined to progress to a more advanced and malignant stage than that with low ANLN expression.

3.4. GSEA identified the ANLN-related signaling pathways and Gene Ontology (GO) terms

GSEA was carried out to further investigate the differences between the high and low ANLN expression groups. As revealed by our findings, the GO terms, including mRNA binding, G1/S phase transition in the cell cycle, and Atpase activity, were the top 3 with the highest ES scores, and they were differentially enriched in the high ANLN expression phenotype (FDR=0, NOM=0) (Fig. 4A -4C); in the meantime, the high density lipoprotein particle, protein activation cascade and protein lipid complex were closely correlated with the low ANLN expression phenotype (FDR<0.01, NOM=0) (Fig. 4D – 4F). Furthermore, results of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed that the genes in high ANLN expression group were mainly enriched in the cell cycle, ubiquitin-ediated proteolysis and oocyte meiosis (FDR=0.01, NOM=0) (Fig. 5A -5C); in addition, the complement and coagulation cascades, primary bile acid biosynthesis, and fatty acid metabolism were primarily enriched in low ANLN expression group (FDR<0.05, NOM=0) (Fig. 5D -5F). Fig.4G and Fig.5G exhibit the comprehensive diagrams showing the above items. More details on the above pathways and GO terms are summarized in Supplementary Table 1-4.

3.5. Co-expression analysis and ANLN networks of kinases, miRNAs and TF targets in HCC
The Function module of LinkedOmics was utilized to analyze the mRNA sequencing data of 371 HCC patients from TCGA database. As shown in the volcano plot (Fig. 6A), 129,555 genes (dark red dots) were significantly positively correlated with \( \text{ANLN} \), whereas 6,965 genes (dark green dots) were exhibited significant negative correlations (FDR < 0.01). Thereafter, those 50 genes that showed significant positive and negative correlations with \( \text{ANLN} \) were exhibited in the heat map (Fig. 6B and 6C, respectively). Supplementary Fig. 1A–1F display the statistical scatter plots for individual genes. To further explore the \( \text{ANLN} \) targets in HCC, the kinases, miRNAs and TFs target networks of the positively correlated gene sets generated by GSEA were analyzed, respectively. Of them, the top 5 most significant target networks were the kinase-target networks mainly related to the polo-like kinase 1 (PLK1), cyclin-dependent kinase 1 (CDK1), aurora kinase B (AURKB), ATM serine/threonine kinase, and cyclin-dependent kinase 2 (CDK2) (\( P = 0 \), FDR = 0) (Table 6). Besides, the miRNA-target network was associated with (AGTCTTA) miR-499, (CTGTTAC) miR-194, (GTAGGCA) miR-189, miR-380-3P and MIR-452 (\( P = 0 \), FDR < 0.1). The TF-target network was mainly related to the E2F (E2F) TF family, including E2F1_Q6, E2F4DP1_01, E2F_Q6, E2F1DP1_01, and E2F1DP2_01 (\( P = 0 \), FDR = 0). Besides, the protein-protein interaction (PPI) network constructed by GeneMANIA revealed correlations among genes for kinases PLK1 (Fig. 7), MIR499 (Supplementary Fig. 2), and TF E2F1_Q6 (Supplementary Fig. 3). Typically, the gene set enriched for kinases PLK was mainly responsible for mitosis, nuclear division and organelle fission, while that enriched for TF E2F1_Q6 was related to DNA replication.

3.6. \( \text{ANLN} \) expression was correlated with the immune infiltration levels in HCC

The correlations of \( \text{ANLN} \) expression with immune infiltration levels in HCC samples collected from TIMER were analyzed, and the results suggested that \( \text{ANLN} \) expression was significantly correlated with the infiltration levels of B cells (\( r = 0.429, \ P<0.001 \)), CD8+ T cells (\( r = 0.273, \ P<0.001 \)), CD4+ T cells (\( r = 0.342, \ P<0.001 \)), macrophages (\( r = 0.42 \ P<0.001 \)), neutrophils (\( r = 0.374, \ P<0.001 \)) and dendritic cells (\( r = 0.422, \ P<0.001 \)) (Fig. 8A). These findings suggested that \( \text{ANLN} \) might play a specific role in immune infiltration within the HCC microenvironment, especially for those of B cells, macrophages and dendritic cells.

3.7. \( \text{ANLN} \) expression in Hep3B cells was dramatically upregulated compared with that in LO2 cells

The cell culture results are presented in Supplementary Fig. 4. As observed, \( \text{ANLN} \) expression was significantly upregulated in Hep3B cells relative to that in normal liver LO2 cells (Fig. 8B) (\( P<0.05 \)). The amplification plot and melt curve were also displayed in Fig. 8C and Fig. 8D, respectively.

3.8. The protein levels corresponding to \( \text{ANLN} \) mRNA levels in cell lines detected by Western blotting

Fig. 9 shows the protein expression levels corresponding to \( \text{ANLN} \) mRNA levels in Hep3B and L-O2 cells. Clearly, the protein level corresponding to \( \text{ANLN} \) mRNA level was lower in LO2 cells than in Hep3B cells (\( P< 0.05 \)).

3.9. IHC results regarding \( \text{ANLN} \) expression in HCC tissues
IHC was carried out to examine the ANLN protein expression in HCC tissues and their counterparts, as well as ANLN expression in HCC. It was discovered that, the ANLN protein expression was upregulated in HCC tissues (Fig. 10A and 10B) relative to that in normal tissues (Fig. 10C and 10D).

4. Discussion

The differential expression of ANLN has been reported in various cancers[7-11]. According to previous studies, the high ANLN level is associated with the dismal prognosis for HCC patients[5]. Using the MYC-induced, DEN-induced liver tumorigenesis and tumor engraftment mouse models, Zhang et al. concluded that ANLN knockdown in liver cells inhibited the development of liver cancer [12]. It was found in this study that, compared with normal tissues, ANLN was highly expressed in bladder, breast, biliary, head and neck, colorectal, esophageal, gastric, kidney, lung, uterus, prostatic, rectal cancers and HCC based on TCGA database. Using the HCCDB database, the differential expression of ANLN was further observed in multiple independent datasets with regard to HCC, among which, obvious expression differences were observed in GSE25097 (LogFC=11.852, \(P<0.001\)) and ICGC-LIRI-JP (LogFC=5.559, \(P<0.001\)) datasets. Meanwhile, high ANLN expression was significantly associated with the poor OS for TCGA-LIHC and ICGC-LIRI-JP patients. Besides, data from the Kaplan-Meier Plotter dataset suggested that, high ANLN expression was correlated with the high HR for poor OS of LIHC. Specifically speaking, the up-regulated ANLN expression was shown to be correlated with the poor prognosis for liver cancer at stage I to IV, grade 1 to 3 and T1 to T3. Patients with micro-vascular invasion, alcohol consumption and hepatitis virus infection also had worse OS in the presence of high ANLN expression. In addition, as far as HR was concerned, the prognosis for male and Asia patients with high ANLN expression was worse than that for female and white cases. Of interest, G1 patients with high ANLN expression had higher HR values than those in G2 and G3 patients, which might reflect that the high ANLN expression had a greater impact on the survival for patients with lower-grade tumor. Emphatically, the results generated from the UALCAN dataset also indicated that, the up-regulated ANLN expression in HCC was associated with the advanced clinicopathological characteristics (such as high grade, clinical stage, and T stage). Afterwards, logistic regression analysis suggested the HRs between different tumor grades and stages. Taken together, these findings strongly suggested that ANLN was a prognostic biomarker in HCC.

To further investigate the functions of ANLN in HCC, GSEA was carried out using TCGA data, and the findings revealed that ANLN might be implicated in regulating mRNA binding, G1/S phase transition in the cell cycle, and atpase activity. Moreover, cell cycle, ubiquitin-mediated proteolysis, and oocyte meiosis were proved to be the pathways with the highest NES scores that were differentially enriched in the high ANLN expression phenotype. Also, the interactions between ANLN and other functional partners in LinkedOmics database were also investigated. The results revealed that, many genes were co-expressed with ANLN in LIHC, among which, the BUB1 mitotic checkpoint serine/threonine kinase B (BUB1B) encoding a kinase involved in spindle checkpoint function, had the highest Pearson's correlation coefficient of 0.9317. Noteworthily, BUB1B up-regulation in tumor tissues also predicted the worse OS and disease-free survival (DFS) for HCC patients[13]. Moreover, among the highly ANLN-related kinases studied in this study, PLK1, which belongs to the conserved serine/threonine kinases family involved in
multiple mitotic processes (including functional maturation of centrosomes and response to DNA damage), is emphasized by previous literature to be a determinant factor of the SN38 sensitivity of tumor cells, and its over-expression is inversely correlated with the survival rate of HCC[14-16]. Other kinases, such as CDK1 and AURKB involved in our analysis, also regulate the genomic stability, mitosis and cell cycle[17, 18]. In terms of the TFs targeted by ANLN, E2F1 is one of the key links in the cell cycle regulation network, and its aberrant expression is actively involved in HCC genesis and development of HCC, which predicts the unfavorable prognosis for patients[19, 20]. Besides, miR499, which may have a certain regulating effect on ANLN, is demonstrated to be associated with the susceptibility to HBV-related HCC among the Chinese population[21]. Furthermore, our results also revealed that, the infiltration levels of macrophages, DCs, B cells, CD8+ T cells, CD4+ T cells and neutrophils were significantly positively correlated with ANLN expression in HCC. A recent study has recognized ANLN as an immune-related gene for lung adenocarcinoma[22]. Although few reports are available to examine the effect of ANLN on the immune microenvironment of HCC, our research provides novel clues for understanding the role of ANLN in HCC.

In this study, ANLN is also observed to be upregulated in HCC tissues and cell lines, which validated ANLN expression from another aspect. Nevertheless, more in-depth experiments, such as overexpression studies, are required to further confirm the role of ANLN in HCC progression. The relationships of ANLN with the miRNAs revealed in this study, including miR499 and miR194, should also be further verified. Another limitation of this study is that, few stage IV LIHC samples are enrolled, yet the clinical reality is that most HCC patients are first diagnosed at the advanced stage, with extremely poor prognosis. Consequently, more comprehensive patient information and a larger sample size are needed to complete the more specific analysis.

Conclusions

In conclusion, ANLN, which may be positively co-expressed with BUB1B, serves as a potential prognostic molecular marker to predict the poor survival for HCC. Moreover, ANLN expression is linked with multiple clinicopathological parameters, also, it is involved in cell cycle progression and mRNA binding. Besides, PLK1, miR499 and E2F1 are identified to be the candidate targets of ANLN. Further, ANLN is found to be associated with the infiltration of immune cells in HCC.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.
Acknowledgements

The authors would like to thank the HCCDB, UALCAN, Kaplan-Meier plotter, TIMER, LinkedOmics and TCGA databases for the availability of the data.

Consent for publication

All authors agree to submit for consideration for publication in the journal.

Availability of data and materials

All data generated or analyzed during this study are included in this article and its supplementary information files.

Funding

Not applicable.

Authors' contributions

Xin-Ting Sang and Bo Hu created the idea for the paper. Bo Hu performed the selection of literature, drafted the manuscript, and prepared the figures. Xiao-Bo Yang and Xin-Ting Sang revised the manuscript. All authors read and approved the final manuscript.

References


**Supplementary Figure Legends**

**Supplementary Figure 1.** Gene expression correlation analysis for *ANLN* (LinkedOmics). The scatter plot shows Pearson correlation of *ANLN* expression with expression of *BUB1B* (A), *KIF23* (B), *TOP2A* (C), *CCl14* (D), *NDUFAF1* (E) and *CLEC3B* (F).

**Supplementary Figure 2.** Protein-protein interaction network of miR-499 microRNA-target networks (GeneMANIA). Protein-protein interaction (PPI) network and functional analysis indicating the gene set that was enriched in the target network of miR499. Different colors of the network edge indicate the bioinformatics methods applied: co-expression, website prediction, pathway, physical interactions, shared protein domains, genetic interaction and co-localization.

**Supplementary Figure 3.** Protein-protein interaction network of E2F1_Q6 networks (GeneMANIA). Protein-protein interaction (PPI) network and functional analysis indicating the gene set that was enriched in the target network of E2F1_Q6. Different colors of the network edge indicate the bioinformatics methods applied: co-expression, website prediction, pathway, physical interactions, shared protein domains, genetic interaction and co-localization. The different colors for the network nodes indicate the biological functions of the set of enrichment genes.

**Supplementary Figure 4.** (A-D) Cells of Hep3B cell line. (E-F) Cells of LO2 cell line.

**Figures**
Figure 1

Comparison of ANLN expression between normal tissue and hepatocellular carcinoma. (A) Expression of ANLN in different cancers in Oncomine database. (B) Human ANLN expression levels in different tumor types from TCGA database were determined by TIMER (*P < 0.05, **P < 0.01, ***P < 0.001). (C) and (D) indicates that ANLN expression levels in hepatocellular carcinoma is significantly higher than that in normal tissues. (E) Expression of ANLN in multiple hepatocellular carcinoma datasets.
Figure 2

Impact of ANLN expression on overall survival in hepatocellular carcinoma patients in The Cancer Genome Atlas (TCGA) cohort (A) and International cancer Genome Consortium (ICGC) cohort (B). The receiver operating characteristic (ROC) curve for 1-, 3- and 5-year OS was shown in (C) with regard to TCGA.
Figure 3

ANLN transcription in subgroups of patients with hepatocellular carcinoma, stratified based on gender, age and other criteria (UALCAN). (A) Boxplot showing relative expression of ANLN in normal individuals of either gender or male or female LIHC patients. (B) Boxplot showing relative expression of ANLN in normal individuals of any age or in LIHC patients aged 21–40, 41–60, 61–80, or 81–100 yr. (C) Boxplot showing relative expression of ANLN in normal individuals of any weight or in LIHC patients with normal weight, extreme weight, obese and extreme obese. (D) Boxplot showing relative expression of ANLN in normal individuals of any ethnicity or in LIHC patients of Caucasian, African-American or Asian ethnicity. (E) Boxplot showing relative expression of ANLN in normal individuals or in LIHC patients based on nodal metastasis status. (F) Boxplot showing relative expression of ANLN in normal individuals or in LIHC patients based on cancer stages. (G) Boxplot showing relative expression of ANLN in normal individuals or in LIHC patients with grade 1, 2, 3 or 4 tumors. Data are mean ± SE. *, P < 0.05; **, P < 0.01; ***, P < 0.001.
Figure 4

Enrichment plots of gene Ontology (GO) from gene set enrichment analysis (GSEA). GSEA results showing (A) mRNA binding, (B) Cell cycle G1 S phase transition, (C) Atpase activity were the top 3 by enrichment score that differentially enriched in high-ANLN phenotype, and (D) High density lipoprotein particle, (E) Protein activation cascade (F) Protein lipid complex were closely correlated with the low-ANLN phenotype. (G) summarizes the above six GO terms.
**Figure 5**

Enrichment plots of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways from gene set enrichment analysis (GSEA). GSEA results showing (A) Cell cycle, (B) Ubiquitin mediated proteolysis, (C) Oocyte meiosis were the top 3 by enrichment score that differentially enriched in high-ANLN phenotype. (D) Complement and coagulation cascades, (E) Primary bile acid biosynthesis, (F) Fatty acid metabolism were primarily enriched in low-ANLN group. (G) summarizes the above six KEGG pathways.
Figure 6

Genes differentially expressed in correlation with ANLN in hepatocellular carcinoma (LinkedOmics). (A) A Pearson test was used to analyze correlations between ANLN and genes differentially expressed in LIHC. (B–C) Heat maps showing genes positively and negatively correlated with ANLN in LIHC (TOP 50). Red indicates positively correlated genes and green indicates negatively correlated genes.
Figure 7

Protein-protein interaction network of PLK1 kinase-target networks (GeneMANIA). Protein-protein interaction (PPI) network and functional analysis indicating the gene set that was enriched in the target network of PLK1 kinases. Different colors of the network edge indicate the bioinformatics methods applied: co-expression, website prediction, pathway, physical interactions, shared protein domains, genetic interaction and co-localization. The different colors for the network nodes indicate the biological functions of the set of enrichment genes.
Figure 8

Correlation of ANLN expression with immune infiltration level in hepatocellular carcinoma was shown in (A). Expression has significant positive correlations with infiltrating levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells, other than tumor purity. (B). Expression of ANLN in Hep3B cell line compared to LO2 cell line. (C) and (D) displayed the amplification curve and melt curve, respectively. *, P < 0.05; **, P < 0.01; ***, P < 0.001.
Figure 9

The expression levels of the protein corresponding to ANLN mRNA in Hep3B cell line and L-O2 cell line.
Figure 10

Immunohistochemistry (IHC) results of ANLN protein expression in HCC tissues and their counterparts. ANLN proteins were more highly expressed in the HCC tissues (A and B) than in the normal tissues (C and D). All figures were shown in 200X.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.docx
- SupplementaryTable2.docx
- SupplementaryTable3.docx
- SupplementaryTable4.docx
- SuppFig1.jpg
• SuppFig2.jpg
• SuppFig3.jpg
• SuppFig4.jpg