Identification and Validation of Novel Biomarkers and Potential Targeted Drugs in Cholangiocarcinoma: Bioinformatics, Virtual Screening, and Biological Evaluation

Jiena Wang1, Weiwei Zhu1,2,3, Junxue Tu1, and Yihui Zheng1*

1Department of Pharmacy, The Second Affiliated Hospital and Yuying Children’s Hospital of Wenzhou Medical University, Wenzhou, Zhejiang 325000, P.R. China
2College of Pharmacy, Wenzhou Medical University, Wenzhou, Zhejiang 325035, P.R. China
3College of Pharmacy, Chonnam National University, Gwangju 61186, Republic of Korea

Introduction

Approximately 13% of global cancer-related mortality is currently caused by hepatobiliary malignancies [1]. Cholangiocarcinoma (CCA) captures 15-20% of the primary hepatobiliary malignancies, thereby making it the second most common primary liver cancer [2]. Although CCA is relatively rare in, it is much more aggressive than other common cancers such as liver hepatocellular carcinoma, pancreatic adenocarcinoma, and breast invasive carcinoma [3]. CCA arises from epithelial cells in the bile duct [4]. Based on its anatomical location, CCA can be classified into three subtypes: intrahepatic, perihilar, and distal CCA [5]. Relatively few studies of CCA diagnosis, treatment, and prognosis have been conducted [6]. CA19-9 is the only available tumor marker for CCA, but low specificity and false-positive are regularly reported in clinical use [7]. Surgery, radiotherapy, and chemotherapy are combined for treating CCA patients, but the outcomes are limited with a small extension to survival rate[8]. At the same time, the symptoms of early-stage CCA, such as jaundice and abdominal pain, similar to those of other common hepatobiliary diseases, so the majority of patients neglect early-mild symptoms and are diagnosed at an advanced stage, further increasing the difficulty of CCA treatment[9]. Therefore, more specific CCA biomarkers are urgently required. In addition, the targeted drug blocks the growth and spread of cancer cells in CCA treatment by targeting the specific genes, proteins, or tissue environment of the cancer that contributes to its growth and survival [10]. Therefore, the identification of new targeted drugs would be helpful for neoadjuvant and adjuvant therapy in CCA.

Keywords: Cholangiocarcinoma, miRNA, CDK1, virtual screening, molecular modeling
Bioinformatics is a field combining biology, medicine, mathematics, and computer science [11]. While omics, particularly genomics, have become increasingly popular in recent years, bioinformatics can help scientists easily find useful information among the large and complex amount of genome data [12]. Therefore, it is more reliable to analyze the underlying molecular pathway, genetic alternation, prognosis, potential target gene, and novel biomarkers in cancer through the utilization of bioinformatics for processing multi-omics biological data. Combined high-throughput technology and computer-aided analysis would be helpful for identifying new CCA markers and novel target molecules, which have huge clinical value for overcoming challenges during early diagnosis, proper treatment, and ideal prognosis, thereby decreasing CCA morbidity and mortality.

Virtual screening is a promising tool for drug discovery and development [13]. In comparison to random screening, virtual screening achieves a hundred to thousand-fold efficiency for drug hit rate enrichment [14, 15]. There are generally two different virtual screening strategies: docking-based virtual screening and pharmacophore-based virtual screening [16]. Researchers have found pharmacophore-based virtual screening to outperform docking-based virtual screening for the retrieval of actives from the databases, making pharmacophore-based virtual screening a powerful drug discovery method [17].

This study identifies and analyzes differentially expressed genes (DEGs) from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases using the bioinformatics approach. The potential key hub genes related to CCA are identified through the construction of a protein-protein interaction (PPI) network. For tumor down-regulated hub genes, the regulatory miRNA are predicted and the content change is confirmed in several GEO databases and biliary cell lines. For tumor up-regulated hub genes, the expression level in cells is validated and some compounds that exert potential gene-targeted activities are screened using pharmacophore-based virtual screening. Molecular dynamics simulations are adopted as a means of confirming the degree of binding between the key target and drug candidate. Finally, in vitro experiments are utilized to provide further confirmation of the anti-tumor activity of the drug candidate in cholangiocarcinoma cells.

Materials and Methods

Data Collection

Two gene expression profiles (GSE26566 and GSE45001) and three miRNA expression profiles (GSE113740, GSE113486, and GSE112264) were retrieved from the GEO (https://www.ncbi.nlm.nih.gov/geo/) database [18]. GSE26566 contains 104 CCA tumor samples and 6 normal samples, while GSE45001 contains 10 CCA tumor samples and 10 normal samples. GSE113740, GSE113486, and GSE112264 contain 25, 40, and 50 CCA tumor samples and 10, 100, and 41 normal samples respectively. The RNA sequencing datasets of 33 CCA tumor samples and eight normal samples, including clinicopathological information, were downloaded from the TCGA (TCGA-CHOL, https://tcga-data.nci.nih.gov/) database [19].

Identification of Differentially Expressed Genes and miRNA

GSE26566, GSE45001, and TCGA-CHOL data were processed using RStudio software (version 4.1.0, https://www.rstudio.com/). For GSE26566 and GSE45001, the limma package was used to screen differentially expressed genes (DEGs) between tumor samples and normal samples [20]. With the TCGA-CHOL data, the edgeR package was used to screen DEGs [21]. GEO2R online analysis tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) was used for evaluating differentially expressed miRNA between CCA tumors and normal tissue samples [22]. All genes or miRNA with log2FC > 1 and adj.p < 0.05 were considered to be of statistical significance. Venn diagram analysis was used as a means of obtaining the overlapping DEGs in the GSE26566, GSE45001, and TCGA-CHOL data.

GO and KEGG Enrichment Analysis

Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) online web server was used for performing Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis [23]. Terms with p < 0.05 were considered to be significant. Analytic results in DAVID were plotted using the Hiplot (https://hiplot.com.cn/) data visualization webserver.

Construction of Protein-protein Interaction Network and Analysis of Clusters

Protein-protein interactions (PPI) of the DEGs were constructed by using the STRING (https://string-db.org/) database [24]. A confidence score of ≥ 0.4 was considered to be significant. The results of the PPI network were then downloaded and embellished using Cytoscape (version 3.8.2) [25]. The plugin Molecular Complex Detection (MCODE) was utilized for screening key gene clusters with haircut on and degree cut-off, node score cut-off, k-core, max depth set as 2, 0.2, 2, 100.

Hub Gene Analysis

CytoHubba, which is a Cytoscape plugin, was used for identifying the targets with higher degrees in different computation methods [26]. 10 Cytohubba algorithms, including BottleNeck, EPC, Betweenness, Closeness, Degree, MCC, MNC, Radiality, EcCentricity, and Stress were used for gene ranking. The common genes ranking in the top 10 for different algorithms were listed and those with an occurrence rate of ≥50% were considered to be hub genes. The expression of hub genes in 23 different cancers was analyzed by the GEPIA (http://gepia.cancer-pku.cn/) webserver based on information that was derived from the TCGA database [27].
Prediction of miRNA Regulating Hub Gene Expression

The TargetScan (human7.2, http://www.targetscan.org/vert_72/) webserver was used for predicting hub gene-regulatory miRNA [28]. Only broadly conserved miRNA, which presents greater reliability and prediction accuracy, was recorded for further analysis.

Survival Analysis

The GEPIA webserver was used for performing prognostic analysis of hub genes in CCA [27]. Patients were split using auto selection, and the hazard ratio with 95% confidence intervals and log-rank $p$ values were then computed.

Pharmacophore-Based Virtual Screening

Pharmacophore-based screening was performed using the online server Pharmit (https://pharmit.csb.pitt.edu/) [30]. The pharmacophore models were constructed based on CDK1/Cks2 complex (PDB: 6GU7), which was co-crystallized with AZD5438 [31]. For the virtual screening program, most pre-installed Pharmit parameters were left unchanged. One hydrogen donor and five hydrogen acceptors were locked for compound screening in the Molport library. The top 10 hit compounds were listed based on their scores.

Molecular Dynamics Simulation

The crystal structure of 6GU7 was derived from the Protein Data Bank. The basic structure of Molport-003-703-800 was derived from the Molport library and the geometry structure was further optimized using the Gaussian09 program with B3LYP/6-31G (d,p) basis sets. Molecular docking between 6GU7 and Molport-003-703-800 was performed by using AutoDock Vina software. The minimum energy conformation from 100 docking results was chosen for the subsequent molecular dynamics simulation.

The Gromacs2019.4 program was used to simulate the molecular dynamics of the selected docked poses. The force field parameters for small molecules were calculated using the online webserver LigParGen (http://zarbi.chem.yale.edu/ligpargen/). The entire system was solvated in TIP3P water molecules in a box with 10Å×10Å×10Å dimensions. Before the MD simulation commenced, 2,500 steps of energy minimization were performed using the steepest descent and conjugate gradient method. The constraints were subsequently released and the same 5,000 steps of energy minimization were run for the entire system, including 2,500 steepest descent method steps and 2,500 conjugate gradient method steps. During the molecular dynamics simulations, long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method. The time step was set as 2 fs, and LINCS was applied as a means of constraining the bonds that connect hydrogen atoms. A nonbonded interaction cut-off of 10Å was utilized. Temperature and pressure were maintained at 300 K and 1 atm through use of the V-rescale temperature and Berendsen pressure coupling method. The system was finally submitted to 100 ns molecular dynamics simulations. Simulation results were visualized by using the Gromacs embedded program and PyMOL (https://pymol.org/2/).

Reagents and Cell Culture

Molport-003-703-800 (M078) was purchased from Molport (https://www.molport.com/, Riga, Latvia). The human cholangiocarcinoma cell lines RBE cells and human intrahepatic biliary epithelial cells HiBEC were acquired from the Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences (China). The cells were cultured in RPMI-1640 (Gibco, Germany) medium, which was supplemented with 10% fetal bovine serum (FBS) and 1× antibiotic/antimycotic (Gibco).

Real-Time qPCR Assay

Total RNA was extracted from cells by using TRizol reagent (Molecular Research Centre, USA). RNA was reverse transcribed by using the PrimeScript RT reagent Kit (Takara, Japan). Real-time qPCR was performed using TB Green Premix Ex Taq II (Takara), on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Target gene sequences can be seen in Table S1. Target transcript levels were normalized to Actb levels.

Cell Viability Assay

The RBE cell line was placed into 96-well plates at a density of 5,000 cells/well. Different concentrations of the test compounds were then added to the cells. Following 72 h of incubation, 20 μl MTT solution (Solarbio Life Science, China) was added to each well, and the cells were then incubated for a further 4 h. The medium was then removed, and 150 μl DMSO solution was added to each well. Cell viability was calculated by taking measurements of the OD490 using a SpectraMax M5 Multi-Mode Microplate Reader.

Western Blot

Proteins were isolated using RIPA lysis and extraction buffer (catalog number: 89900; Thermo Fisher), and protein concentrations were measured with a Pierce BCA Protein Assay Kit. Protein lysates were separated through the use of sodium dodecyl sulfate-polyacrylamide gel electrophoresis before being transferred to polyvinylidene fluoride membranes. The membranes were blocked with skim milk for 1 h at room temperature, before being incubated overnight with primary antibodies at 4°C. Secondary antibodies were applied at room temperature.
temperature for one hour. Immunoreactivity was visualized by using an enhanced chemiluminescence reagent (Bio-Rad, USA).

**Scratch-Wound Migration Assay**

RBE cells were seeded in 6-well plates where they were allowed to grow until confluent. The confluent cells were wounded by being scraped with a sterile 0.2-ml-pipette tip. The culture medium was replaced with fresh medium containing 0, 2, 5, 10, 25 μM of Molport-003-703-800 and cultured for 48 h. The rate of wound closure was monitored by images that were captured using a phase-contrast microscope.

**Cell Apoptosis Analysis with Flow Cytometry**

A FITC Annexin V Apoptosis Detection Kit (BD Biosciences, USA) was used to conduct apoptosis analysis. The RBE cells were treated with Molport-003-703-800 for 24 h in gradient concentration. The cells were then harvested before being resuspended in 1× binding buffer and stained with propidium iodide (PI)/FITC-Annexin V. Staining was terminated after 15 min with 400 μl of 1× binding buffer. The cells were finally analyzed using BD Accuri C6 Cytometer (BD Biosciences, USA).

**Tunel Staining Assay**

RBE cells were permeabilized with 4% paraformaldehyde fixation for 10 min and 0.1% Triton X-100 for 10 min. They were then blocked with 5% bovine serum albumin for 30 min. 50 μl TUNEL detection solution was added to the sample and then incubated in the dark at 37°C for one hour. Images were obtained via a fluorescent microscope (200 × amplification; Nikon, Japan).

**Statistical Analysis**

Kolmogorov–Smirnov test was used for validating the normal distribution of all data. Student’s t-test was used as a means of comparing two groups of data. One-way ANOVA and Dunnett’s post-hoc test were used for comparing more than two groups of data. All data was expressed as mean ± SEM. Prism 8.0 software (GraphPad, USA) was used for statistical analysis and statistical significance was set as \( p < 0.05 \).

**Results**

**Identification of DEGs in Cholangiocarcinoma**

DEGs in cholangiocarcinoma (CCA) were retrieved from both the TCGA and GEO databases. In the GEO database, the GSE26566 dataset contained 104 tumor cases and 6 normal cases, while GSE45001 contained 10 tumor cases and 10 normal cases. In the TCGA database, there were 33 tumor cases and 8 normal cases that were associated with CCA (TCGA-CHOL). Based on the parameters of an adjust. \( p \) value < 0.05 and \(|\log_{2} FC|\geq1\), 1,692 DEGs were identified in GSE26566, 1,793 DEGs were identified in GSE45001, and 8,640 DEGs were identified from TCGA data. In addition, Venn diagram analysis revealed there to be 301 common DEGs in the three different datasets (Fig. 1A). The overlapped DEGs were considered to be of great significance and were recorded for further analysis.

**Functional and Pathway Enrichment Analysis of the DEGs**

GO and KEGG analyses were performed on the 301 overlapping DEGs. With KEGG pathway enrichment, the top 10 involved pathways were listed and it was noted that several pathways, including metabolism pathway, bile secretion, cell cycle, PPAR signaling pathway, and complement and coagulation cascades, were mainly affected by CCA (Fig. 1B). In addition, GO analysis revealed CCA-induced alternation in three terms. Regarding the biological process, the overlapped DEGs were significantly enriched in the oxidation-reduction process, cell adhesion, metabolic process, cell proliferation, and extracellular matrix organization (Fig. 1C). Regarding the molecular function, the overlapped DEGs were significantly enriched in ATP binding, calcium ion binding, receptor binding, enzyme binding, and lipid binding (Fig. 1D). Regarding the cellular component, the overlapped DEGs were significantly enriched in the extracellular environment and cytosol (Fig. 1E).

**Construction of PPI Network and Analysis of Clusters**

A PPI network is propitious to visualizing the relationship between the DEGs. Following analysis of the PPI by the STRING web server, the network was displayed using Cytoscape. PPI network interaction revealed 219 nodes and 1,421 edges on the map (Fig. 2A). Node color is proportional to the degree of connectivity with other targets. The MCODE plugin was then used for analyzing the entire PPI network and distinguishing between key gene clusters. The top three clusters were shown in the results based on their scores, and these clusters contained 38, 13, and 18 DEGs respectively (Figs. 2B-2D). In addition, KEGG analysis was performed on all 3 clusters. As the pathway enrichment results reveal, cluster 1 was significantly involved in cell proliferation, which is characteristic for almost all types of cancer, while clusters 2 and 3 were both found to have a strong correlation to complement and coagulation cascades (Fig. 2E). These results prove that CCA mainly results in the alternation of cell proliferation and complement and coagulation cascades.

**Hub Gene Identification**

Hub genes were chosen from the PPI network, in addition to using the cytoHubba following the construction of the PPI network. 10 different algorithms, including BottleNeck, EPC, Betweenness, Closeness, Degree, MCC, MNC, Radiality, EcCentricity, and Stress, were selected for gene ranking. The top 10 genes were constructed into a
network for each algorithm (Fig. 3A). As tremendous differences were observed between the different algorithms, gene occurrence frequency was further calculated. The top 10 genes with high frequency were listed, and seven genes, CDK1, AHSG, F2, TTR, CCNB1, KIAA0101, and KNG1, with a frequency of ≥50% were considered to be hub genes in CCA (Fig. 3B). Interestingly, AHSG, F2, TTR, and KNG1 were all mapped in cluster 2, whereas CDK1, CCNB1, and KIAA0101 were all mapped in cluster 1. As clusters 1 and 2 were enriched in totally different pathways, the seven hub genes were then divided into two groups for further discussion.

Expression of AHSG, F2, TTR, and KNG1 in CCA

Firstly, the expression level of the four hub genes mapped in cluster 2 was analyzed through utilization of the GEPIA web service. Heat map results found AHSG, F2, TTR, and KNG1 all to be significantly downregulated in CCA (Fig. 4A). Furthermore, CCA also downregulated AHSG, F2, TTR, and KNG1 in GSE26566, GSE45001, and TCGA-CHOL datasets (Fig. 4B). These transcript differences between normal biliary cell line HiBEC and cholangiocarcinoma cell line RBE were also validated, and it was found that all four genes were silent in cholangiocarcinoma cells (Fig. 4C). These results indicate that AHSG, F2, TTR, and KNG1 could all be considered as novel CCA biomarkers. It should also be noted that TTR and KNG1 were also expressed and suppressed in...
other types of cancer, while AHSG and F2 were only expressed in CCA and liver hepatocellular carcinoma and only suppressed in CCA (Fig. 4A), which indicates that AHSG and F2 have higher specificity.

Fig. 2. The PPI network construction of total DEGs. (A) The complete PPI network contains 219 nodes and 1421 edges. (B-D) The top 3 clusters seek out by the MCODE plugin. (E) The KEGG pathway enrichment of three clusters in panel B–D.
miRNA Prediction and Expression Analysis
miRNA was considered the underlying mechanism that is responsible for AHSG, F2, TTR, and KNG1 downregulation in CCA. In order to verify this hypothesis, miRNA prediction was performed to confirm the potential regulatory miRNA of AHSG, F2, TTR, and KNG1 using the TargetScan web service. Eight miRNAs were found that broadly conserved with AHSG, F2, and TTR (Table 1). Conversely, no KNG1 conserved miRNA was found. In addition, the expression level of eight miRNAs was explored in three GEO datasets, GSE113740, GSE113486, and GSE112264. Results found all eight miRNAs to have increased significantly in the tumor sample of all three GEO datasets, while only hsa-miR-206 in GSE113740 presented an adjust.p value that was greater than 0.05 (Fig. 4D). The upregulated expression of these miRNA was found to be identical to the downregulated expression of AHSG, F2, and TTR, thereby proving the potential miRNA-dependent regulation of AHSG, F2, and TTR expressions in tumor samples.

Expression and Prognosis of CDK1, CCNB1, and KIAA0101 in CCA
Secondly, CDK1, CCNB1, and KIAA0101 were mapped in cluster 1 where they played a key role in cell cycle regulation. Heat map results from GEPIA found CDK1, CCNB1, and KIAA0101 all to be significantly upregulated

Table 1. Highly conserved human miRNA list predicted by TargetScan web server.

| Symbol | Consequence | Position | Site type | Predict miRNA                  |
|--------|-------------|----------|-----------|--------------------------------|
| AHSG   | ACUGUGA     | 293-299 of 3' UTR | 7mer-m8 | hsa-miR-27b-3p, hsa-miR-27a-3p |
| TTR    | CAGUGUU     | 78-84 of 3' UTR | 7mer-m8 | hsa-miR-200a-3p, hsa-miR-141-3p |
|        | ACAUUCC     | 129-135 of 3' UTR | 7mer-m8 | hsa-miR-206, hsa-miR-1-3p, hsa-miR-613 |
| F2     | GGCAGCU     | 126-132 of 3' UTR | 7mer-m8 | hsa-miR-22-3p                  |
| KNG1   | Not found   | 126-132 of 3' UTR |          |                                |
Fig. 4. Verifying the mRNA/predicted miRNA expression of hub genes including AHSG, F2, TTR, and KNG1. (A) The heatmap of AHSG, F2, TTR, and KNG1 transcription in different cancer types and paired normal samples. T represents tumor tissues and N represents normal tissues. (B) The mRNA expression of AHSG, F2, TTR, and KNG1 in GSE26566, GSE45001, and TCGA-CHOL. (C) The mRNA expression differences of AHSG, F2, TTR, and KNG1 between RBE cells and HiBEC cells. (D) The miRNA expression in different GEO databases. Data are represented as mean ± SEM; n = 3 per group; *p < 0.05.

in CCA (Fig. 5A). In addition, these upregulating trends were further confirmed in the GSE26566, GSE45001, and TCGA-CHOL datasets (Fig. 5B). Similarly, the mRNA levels of the three genes were found to be richer in RBE cells...
than HiBEC cells (Figs. 5C-5E). Prognostic analysis, which is supported by the GEPIA web service, found that lower expression levels of CDK1, CCNB1, and KIAA0101 have better overall survival in CCA, despite the $p$-value of $>0.05$ presenting no statistically significant differences (Fig. 5F). Considering the $p$-values were still approximately 0.2, a relatively lower value that indicates potential variability, it was believed that there could be a false negative caused by the small sample size ($n = 36$) for survival analysis. It should be noted that infinite

Fig. 5. Verifying the mRNA expression and survival outcomes of hub genes including CDK1, CCNB1, and KIAA0101. (A) The heatmap of CDK1, CCNB1, and KIAA0101 transcription in different cancer types and paired normal samples. T represents tumor tissues and N represents normal tissues. (B) The mRNA expression of CDK1, CCNB1, and KIAA0101 in GSE26566, GSE45001, and TCGA-CHOL. (C-E) The mRNA expression differences of CDK1 (C), CCNB1 (D), and KIAA0101 (E) between RBE cells and HiBEC cells. (F) Correlation between the expression of CDK1, CCNB1, KIAA0101, and survival time in CCA. Data are represented as mean ± SEM; $n = 3$ per group; *$p < 0.05$. 
proliferation potential is a common feature for most types of cancer, so CDK1, CCNB1, and KIAA0101 were broadly expressed and upregulated in a variety of cancers, making them unsuitable for use as CCA biomarkers but suitable for therapeutical targets.

**Potential Targeted Drug**

Currently, of CDK1, CCNB1, and KIAA0101, only CDK1 had been reported with the 3D crystal structure, thereby attracting attention from researchers and resulting in the subsequent development of several small molecular inhibitors. However, none of them were approved by FDA, so there was a requirement for more CDK1-targeted molecules. In order to achieve this, pharmacophores-based virtual screening was performed based on the crystal structure of CDK1/Cks2 in a complex with AZD5438 (PDB: 6GU7). In all pharmacophores, five hydrogen acceptors and one hydrogen donor were locked for searching in the Molport library (Fig. 6A). 480 hit compounds were found and the top 10 were listed (Table 2). The best hit compound was found to be Molport-003-703-800 (M078), and M078 structure was derived online (Fig. 6B).

In addition, molecular dynamics simulation was performed to provide further validation of the interactions between M078 and CDK1. In 100 ns molecular dynamics simulation, the root means standard deviation (RMSD) was calculated through a comparison with the initial position of complexes and the CDK1-M078 interaction displayed an RMSD of ~0.4 at 100 ns (Fig. 6C). It is notable that CDK1 had a small fluctuation before 10 ns and then had a tendency to stabilize, potentially due to interaction between protein and solvent water. For exploring binding affinity, the MM/PBSA approach was used to perform binding free energy calculations of the complexes (Table 3). The results found van der Waals interaction energy (ΔEvdW) to be a major interacting force between ligands and receptors that reached ~42.47 kcal/mol. At the same time, electrostatic interaction (ΔEle) contributed -42.93 kcal/mol to the entire system. Generally, the total binding free energy (ΔGTot) reached -52.48 kcal/mol, representing the high affinity between M078 and CDK1. In addition, the 3D conformation was extracted from the final frame in molecular dynamics simulation (Fig. 6D). At the same time, protein-ligand interaction analysis that

| Rank | CDK1/Cks2 (6GU7) | Score | RMSD |
|------|------------------|-------|------|
| 1    | Molport-003-703-800 | -8.61 | 4.358 |
| 2    | Molport-002-942-764 | -8.60 | 2.171 |
| 3    | Molport-002-696-165 | -8.60 | 6.768 |
| 4    | Molport-007-552-096 | -8.59 | 1.363 |
| 5    | Molport-008-779-098 | -8.57 | 3.074 |
| 6    | Molport-023-277-074 | -8.57 | 3.260 |
| 7    | Molport-002-685-121 | -8.57 | 4.044 |
| 8    | Molport-007-951-239 | -8.57 | 4.861 |
| 9    | Molport-002-694-975 | -8.56 | 4.649 |
| 10   | Molport-016-638-322 | -8.54 | 3.059 |
was performed by Discovery Studio Visualizer 2020 showed that Ile10, Asp86, and Gln132 of CDK1 formed hydrogen bonds with M078 to assist with protein-ligand binding and stabilize the entire complex (Fig. 6D). Therefore, M078 has the potential to be used as the CDK1-targeted drug.

**Validation of the Anti-tumor Activity of Molport-003-703-800 In Vitro**

Following the identification of M078 as the potential targeted drug for CDK1, it was hypothesized that M078 could inhibit the activation of CDK1, thereby restraining the proliferation of cancer cells. As anticipated, in the in vitro study, the CDK1 phosphorylation, which is the active form of CDK1, was inhibited by M078 treatment of the RBE cells (Fig. 7A). In addition, M078 treatment lowered the viability of RBE cells in a dose-dependent manner (Fig. 7B). In addition, according to the Annexin V/PI staining assay and Tunel staining assay results, M078 treatment significantly promoted the apoptosis of RBE, which was indicated by an increase in Annexin V−/PI+ fractions and the Tunel positive area (Figs. 7C and 7D). The scratch-wound assay results also found M078 treatment to reduce RBE migration (Fig. 7E). These results indicate that M078 regulated the cholangiocarcinoma cell proliferation and apoptosis as a means of disrupting cell growth by blocking CDK1 activity.

### Table 3. The results of MM/PBSA free energy calculation (kcal/mol).

| Energy Component   | ΔE_vdw | ΔE_elec | ΔG_tot |
|--------------------|--------|---------|--------|
| Protein-ligand     | -42.47 | -42.93  | -52.48 |

**Fig. 7. In vitro validation of the anti-tumor activity of Molport-003-703-800.** (A) The protein level of CDK1 and phosphorylated CDK1 in RBE cells treated with 0, 2, 5, 10, 25 μM M078 for 1 h. The right panel showed the densitometric quantifications. (B) Cell viability of RBE cells treated with 0, 2, 5, 10, 25, 50 μM M078 for 72 h. (C) Flow cytometry analysis of RBE cells (treated with 0, 2, 5, 10, 25 μM M078 for 24 h) stained with PI/Annexin V. Q2 area (Annexin V−/PI+) represents the cell apoptosis. The right panel showed the quantification of Q2 area. (D) Tunel staining of RBE cells treated with 0, 2, 5, 10, 25 μM M078 for 12 h. The red arrow represents the positive focus. PC means phase contrast. (E) Scratch-wound migration image of RBE cells treated with 0, 2, 5, 10, 25 μM M078 for 48 h. Data are represented as mean ± SEM; n = 3 per group; *p < 0.05.
Discussion

In this study, differential RNA expression data for CCA patients was collected from the TCGA and GEO databases. Over 300 common DEGs were identified by Venn analysis. By using KEGG analysis, GO analysis, and PPI network visualization, seven hub genes (including AHSG, F2, TTR, KNG1, CDK1, CCNB1, and KIAA0101) mapped in two clusters were identified. AHSG, F2, TTR, and KNG1, are focused on the complement and coagulation cascades and play a different role in the regulation of coagulation response. AHSG, also known as fetuin-A, has been proven to present a negative correlation with coagulation activation in preeclampsia [32]. F2 encodes the thrombin, and this is a major component in coagulation cascades that converts fibrinogen into insoluble fibrin strands [33]. TTR and TTR amyloidosis can potentially activate the coagulation and fibrinolytic systems [34]. KNG1 has two different preferences in splicing, thereby generating either high molecular weight kininogen (HMWK) or low molecular weight kininogen (LMWK) [35]. HMWK is essential in blood coagulation and helps position optimally prekallikrein and factor XI next to factor XII [36]. Although these four genes have an opposite function in coagulation systems, they were all found to coincidently decrease in CCA in this study, further confusing the relationship between coagulation response and CCA. In addition, disseminated intravascular coagulopathy (DIC) is a common complication of solid tumors that is accompanied by the overproduction of coagulation factors [37]. However, there is a scarcity of clinical cases and the relationship between DIC and CCA is not yet fully understood. In this study, the relationship between AHSG, F2, TTR, and KNG1 and DIC was confirmed through molecular docking and dynamics simulation.

In this study, differential RNA expression data for CCA patients was collected from the TCGA and GEO databases. Over 300 common DEGs were identified by Venn analysis. By using KEGG analysis, GO analysis, and PPI network visualization, seven hub genes (including AHSG, F2, TTR, KNG1, CDK1, CCNB1, and KIAA0101) mapped in two clusters were identified. AHSG, F2, TTR, and KNG1, are focused on the complement and coagulation cascades and play a different role in the regulation of coagulation response. AHSG, also known as fetuin-A, has been proven to present a negative correlation with coagulation activation in preeclampsia [32]. F2 encodes the thrombin, and this is a major component in coagulation cascades that converts fibrinogen into insoluble fibrin strands [33]. TTR and TTR amyloidosis can potentially activate the coagulation and fibrinolytic systems [34]. KNG1 has two different preferences in splicing, thereby generating either high molecular weight kininogen (HMWK) or low molecular weight kininogen (LMWK) [35]. HMWK is essential in blood coagulation and helps position optimally prekallikrein and factor XI next to factor XII [36]. Although these four genes have an opposite function in coagulation systems, they were all found to coincidently decrease in CCA in this study, further confusing the relationship between coagulation response and CCA. In addition, disseminated intravascular coagulopathy (DIC) is a common complication of solid tumors that is accompanied by the overproduction of coagulation factors [37]. However, there is a scarcity of clinical cases and the relationship between DIC and CCA is not yet fully understood. In this study, the relationship between AHSG, F2, TTR, and KNG1 and DIC was confirmed through molecular docking and dynamics simulation.

In this study, differential RNA expression data for CCA patients was collected from the TCGA and GEO databases. Over 300 common DEGs were identified by Venn analysis. By using KEGG analysis, GO analysis, and PPI network visualization, seven hub genes (including AHSG, F2, TTR, KNG1, CDK1, CCNB1, and KIAA0101) mapped in two clusters were identified. AHSG, F2, TTR, and KNG1, are focused on the complement and coagulation cascades and play a different role in the regulation of coagulation response. AHSG, also known as fetuin-A, has been proven to present a negative correlation with coagulation activation in preeclampsia [32]. F2 encodes the thrombin, and this is a major component in coagulation cascades that converts fibrinogen into insoluble fibrin strands [33]. TTR and TTR amyloidosis can potentially activate the coagulation and fibrinolytic systems [34]. KNG1 has two different preferences in splicing, thereby generating either high molecular weight kininogen (HMWK) or low molecular weight kininogen (LMWK) [35]. HMWK is essential in blood coagulation and helps position optimally prekallikrein and factor XI next to factor XII [36]. Although these four genes have an opposite function in coagulation systems, they were all found to coincidently decrease in CCA in this study, further confusing the relationship between coagulation response and CCA. In addition, disseminated intravascular coagulopathy (DIC) is a common complication of solid tumors that is accompanied by the overproduction of coagulation factors [37]. However, there is a scarcity of clinical cases and the relationship between DIC and CCA is not yet fully understood. In this study, the relationship between AHSG, F2, TTR, and KNG1 and DIC was confirmed through molecular docking and dynamics simulation.

In conclusion, by performing progressive bioinformatics analysis, it was found that complement and coagulation cascades-related genes (including AHSG, F2, TTR, and KNG1) and cell cycle-related genes (including CDK1, CCNB1, and KIAA0101) were conspicuous in the process of CCA. AHSG, F2, TTR, and KNG1 were all specifically downregulated in CCA patients under the regulation of miRNA network, which could be the novel and specific biomarkers for CCA. In addition, CDK1, CCNB1, and KIAA0101 were found to be substantially overexpressed in CCA, which can promote the proliferation of tumor cells. Molport-003-703-800 is a new targeted drug that demonstrated a potent affinity with CDK1 and anti-tumor activity in vitro, making it a potential candidate drug for CCA chemotherapy. The new biomarkers and CDK1-targeted drug will make a significant contribution to improving CCA diagnosis.

Conflict of Interest

The authors have no financial conflicts of interest to declare.

References

1. Khan SA, Tavolari S, Brandi G. 2019. Cholangiocarcinoma: Epidemiology and risk factors. Liver Int. 39 Suppl 1: 19-31.
2. Blechacz R. 2017. Cholangiocarcinoma: Current knowledge and new developments. Gut Liver 11:13-26.
3. Rodrigues PM, Olaiola P, Pavia NA, Olaiola I, Agirre-Lizaso A, Landa A, et al. 2021. Pathogenesis of cholangiocarcinoma. Annu. Rev. Pathol. 16: 433-463.
4. Cardinale V, Bragazzi MC, Carpino G, Torrice A, Praveto A, Gentile R, et al. 2013. Cholangiocarcinoma: increasing burden of classifications. Hepatobiliary Surg. Nutr. 2: 272-280.
5. Valle JW, Kelley RK, Nervi B, Oh DY, Zha AX. 2021. Biliary tract cancer. 
   *Lancet* 397: 428-444.

6. Gigante E, Paradis V, Ronot M, Cauchy F, Sourbrune O, Ganne-Carrie N, et al. 2021. New insights into the pathophysiology and clinical care of rare primary liver cancers. 
   *JHEP* Rep. 3: 100174.

7. Coelho R, Silva M, Rodrigues-Pinto E, Cardoso H, Lopes S, Pereira P, et al. 2017. CA 19-9 as a marker of survival and a predictor of metastasis in cholangiocarcinoma. 
   *GE Port. J. Gastroenterol.* 24: 114-121.

8. Rizvi S, Khan SA, Hallemeyer CL, Kelley RK, Gores GJ. 2018. Cholangiocarcinoma - evolving concepts and therapeutic strategies. 
   *Nat. Rev. Clin. Oncol.* 15: 95-111.

9. Nalanes M, Marin JGC, Lamarca A, Rodrigues PM, Khan SA, Roberts LR, et al. 2020. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. 
   *Nat. Rev. Gastroenterol. Hepatol.* 17: 557-588.

10. Lee Y, Tan YJ, Oon CE. 2018. Molecular targeted therapy: Treating cancer with specificity. 
    *Eur. J. Pharmacol.* 834: 188-196.

11. Bayat A. 2002. Science, medicine, and the future: Bioinformatics. 
    *BMJ* 324: 1018-1022.

12. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, et al. 2018. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. 
    *Brief. Bioinform.* 19: 286-302.

13. Pereira CA, Saye M, Reigada C, Silber AM, Labadie GR, Miranda MR, et al. 2020. Computational approaches for drug discovery against trypanosomatid-caused diseases. 
    *Parasitology* 147: 611-633.

14. Iqbal D, Rehman MT, Bin Dukhyli A, Rizvi SMD, AI Ajmi MF, Alshehri BM, et al. 2021. High-throughput screening and molecular dynamics simulation of natural product-like compounds against Alzheimer’s disease through multitarget approach. 
    *Pharmaceuticals (Basel)* 14: 937.

15. Batoel M, Ahmad B, Choi S. 2019. A structure-based drug discovery paradigm. 
    *Int. J. Mol. Sci.* 20: 2783.

16. Alamri MA, Alamri MA. 2019. Pharmacophore and docking-based sequential virtual screening for the identification of novel Sigma 1 receptor ligands. 
    *Bioinformation* 15: 586-595.

17. Chen Z, Li HL, Zhang QJ, Bao XG, Hu B, Wu W, Li H, et al. 2019. Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets. 
    *Acta Pharmacol. Sin.* 30: 1694-1708.

18. Clough E, Barnett T. 2016. The gene expression omnibus database. 
    *Methods Mol. Biol.* 1418: 93-110.

19. Wang Z, Jensen MA, Zenklusen JC. 2016. A practical guide to the cancer genome atlas (TCGA). 
    *Methods Mol. Biol.* 1418: 111-141.

20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. 
    *Nucleic Acids Res.* 43: e47.

21. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. 
    *Bioinformatics* 26: 139-140.

22. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. 2012. NCBI GEO: archive for functional genomics data sets-update. 
    *Nucleic Acids Res.* 41: D991-995.

23. Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrated analysis of large gene lists using DAVID bioinformatics resources. 
    *Nat. Protoc.* 4: 44-57.

24. Sklarzecyzk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. 
    *Nucleic Acids Res.* 45: D362-D368.

25. Su G, Morris JH, Demchak B, Bader GD. 2014. Biological network exploration with Cytoscape 3. 
    *Curr. Protoc. Bioinformatics* 47: S1.1-3-24.

26. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. 
    *BMC Syst. Biol.* 8 Suppl 1: S11.

27. Tang Z, Li C, Kang B, Gao C, Li C, Zhang Z. 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. 
    *Nucleic Acids Res.* 45: W98-W102.

28. Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. 
    *Elife* 4: e05005.

29. Wong NW, Chen Y, Chen S, Wang X. 2018. OncomiR: an online resource for exploring pan-cancer microRNA dysregulation. 
    *Bioinformatics* 34: 713-715.

30. Sunseri J, Kees DR. 2016. Pharmit: interactive exploration of chemical space. 
    *Nucleic Acids Res.* 44: W442-448.

31. Wood DJ, Korolchuk S, Tatum NJ, Wang LZ, Endicott JA, Noble MEM, et al. 2019. Differences in the conformational energy landscape of CDK1 and CDK2 suggest a mechanism for achieving selective CDK inhibition. 
    *Cell. Chem. Biol.* 26: 121-130 e125.

32. Moller V, Kalab F, Lerszy Z, Sturka A, Halmos A, Stenczer B, et al. 2009. Preeclampsia is associated with decreased serum alpha(2)-HS glycoprotein (fetuin-A) concentration. 
    *Hypertens. Res.* 32: 665-669.

33. Falta S, Saroza R, Falta A. 2014. Overview of the coagulation system. 
    *Indian J. Anaesth.* 58: 515-523.

34. Weisgruber E, Czochraw A, 2021. Transferrin and structural stability of transferrin: from structural stability to cardiovascular and cardiovascular diseases. 
    *Cells* 10: 1768.

35. Peyrou M, Cereijo R, Quesada-Lopez T, Campderros L, Gavalda-Navarro A, Linares-Pose L, et al. 2020. The kallikrein-kinin pathway as a mechanism for auto-control of brown adipose tissue activity. 
    *Ann. Bot.* 1274: 1127-1140.

36. Zamolodchikov D, Duffield M, Macdonald LE, Alessandri-Haber N. 2022. Accumulation of high molecular weight kininogen in the brains of Alzheimer's disease patients may affect microglial function by altering phagocytosis and lysosomal cathepsin activity. 
    *Alzheimers Dement.* (Basel) 18: 397.

37. Palta S, Saroa R, Palta A. 2014. Overview of the coagulation system. 
    *Indian J. Anaesth.* 58: 515-523.

38. Liu H, Wu X, Lu Y, Wu W, Liu H, et al. 2019. Five hub genes can be the potential DNA methylation biomarkers for cholangiocarcinoma using bioinformatics analysis. 
    *Onco Targets Ther.* 12: 8355-8365.

39. Ling L, Cao J, Lin W, Chen H, Xiong X, Ao H, et al. 2020. The roles of cyclin-dependent kinases in cell-cycle progression and therapeutic strategies in human breast cancer. 
    *Int. J. Mol. Sci.* 21: 1960.

40. Strapaka W, Ziemienowicz A. 2011. Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation. 
    *Athen. Bot.* 107: 1127-1140.

41. De March M, Barrera-Vilarmau S, Crespan E, Mentegari E, Merino N, Gonzalez-Magana A, et al. 2018. p15PAF binding to PCNA modulates the DNA sliding surface. 
    *Nucleic Acids Res.* 46: 9816-9828.

42. Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. 
    *Cell* 144: 646-674.

43. Rokoskis R, Jr. 2019. Cyclin-dependent protein serine/threonine kinase inhibitors as anticancer drugs. 
    *Pharmaco. Res.* 139: 471-488.

44. Kommalapati A, Tella SH, Borad M, Javel M, Mahipal A. 2021. FGFR inhibitors in clinical practice: Insight on the management of toxicities in clinical practice. 
    *Cancers (Basel)* 13: 2968.