Scientific Research Publishing

Open Journal of Gastroenterology, 2022, 12, 1-18
https://www.scirp.org/journal/ojgas
ISSN Online: 2163-9469
ISSN Print: 2163-9450

Identification of Driver Genes in Primary Liver Cancer by Integrating NGS and TCGA Mutation Data

Lin Li[1], Lin Niu[1], Na Guo[2], Luyang Cheng[2], Tengfei Hao[3], Ying Xu[2], Xiangling Li[4], Qian Xu[5], Lei Liu[2*], Songhe Yang[1*]

[1]Department of Human Anatomy, Chengde Medical University, Chengde, China
[2]Department of Immunology, Chengde Medical University, Chengde, China
[3]Department of Cardiology, Handan Yongnian Dist. No.1 Hospital, Handan, China
[4]Department of Pathology, Chengde Medical University, Chengde, China
[5]Basic Medical Institute, Chengde Medical University, Chengde, China
Email: *homingreceptor@hotmail.com, *yjsxyysh@126.com

How to cite this paper: Li, L., Niu, L., Guo, N., Cheng, L.Y., Hao, T.F., Xu, Y., Li, X.L., Xu, Q., Liu, L. and Yang, S.H. (2022) Identification of Driver Genes in Primary Liver Cancer by Integrating NGS and TCGA Mutation Data. *Open Journal of Gastroenterology*, **12**, 1-18.
https://doi.org/10.4236/ojgas.2022.121001

Received: December 7, 2021
Accepted: January 11, 2022
Published: January 14, 2022

Abstract

Background: This study is aimed towards an exploration of mutant genes in primary liver cancer (PLC) patients by using bioinformatics and data mining techniques. **Methods:** Peripheral blood or paraffin-embedded tissues from 8 patients with PLC were analyzed using a 551 cancer-related gene panel on an Illumina NextSeq500 Sequencer (Illumina). Meanwhile, the data of 396 PLC cases were downloaded from The Cancer Genome Atlas (TCGA) database. The common mutated genes were obtained after integrating the mutation information of the above two cohorts, followed by functional enrichment and protein-protein interaction (PPI) analyses. Three well-known databases, including Vogelstein's list, the Network of Cancer Gene (NCG), and the Catalog of Somatic Mutations in Cancer (COSMIC) database were used to screen driver genes. Furthermore, the Chi-square and logistic analysis were performed to analyze the correlation between the driver genes and clinicopathological characteristics, and Kaplan-Meier (KM) method and multivariate Cox analysis were conducted to evaluate the overall survival outcome. **Results:** In total, 84 mutation genes were obtained after 8 PLC patients undergoing gene mutation detection with next-generation sequencing (NGS). The top 100 most mutate gene data from PLC patients in TCGA database were downloaded. After integrating the above two cohorts, 17 common mutated genes were identified. Next, 11 driver genes were screened out by analyzing the intersection of the 17 mutation genes and the genes in the three well-known databases. Among them, RB1, TP53, and KRAS gene mutations were connected with clinicopathological characteristics, while all the 11 gene muta-

tions had no relationship with overall survival. **Conclusion:** This study investigated the mutant genes with significant clinical implications in PLC patients, which may improve the knowledge of gene mutations in PLC molecular pathogenesis.

Keywords

Primary Liver Cancer, Mutation, Next-Generation Sequencing, TCGA, Driver Genes

1. Introduction

As the second leading cause of cancer death, the incidence of primary liver cancer (PLC) has shown a significant increase in almost all countries, especially in Asia [1]. Despite advances in many aspects of PLC treatment, including surgical treatment, arterial embolization, and systemic chemotherapy, the 5-year average survival rate is less than 10% [2] [3]. The emergence of molecular targeted drugs provides new treatment options for patients, but currently, there is no effective target for PLC. Therefore, analysis of genetic mutations in malignant progression and identification of biomarkers that would predict tumor behavior to research and develop novel target drugs are urgently needed. It is also the core target of tumor genomics in information mining.

Next-generation sequencing (NGS), a mainstream technology in oncology, is an ability to produce millions of reads in a single run. Compared with traditional gene sequencing (known as Sanger sequencing), NGS makes abundant parallel sequencing with higher throughput and lower cost [4]. Hence, the gene expression profile, mutational genes, and hotspot mutations in pathological samples from PLC patients could be detected on a large scale by NGS technology, furthermore, through bioinformatics analysis, key genes related to the disease can be screened out, which might pave the way toward novel therapeutic targets and molecular targeted drugs [5]. Recently, an increasing number of databases have been developed based on the sequencing of cancer genome, among which The Cancer Genome Atlas (TCGA) database provides relevant data such as tumor gene expression, the copy number variation (CNV), gene mutation, DNA methylation, and clinical patient prognosis. It can also provide important clues for exploring the mechanism of PLC development and searching for therapeutic targets [6] [7].

In this study, a 551 panoramic cancer gene panel was designed and 84 genetic mutations have been identified using targeted NGS in 8 PLC patients. Seventeen common mutations were obtained, following integrating top 100 genes with the highest mutation frequency of 396 PLC cases from the TCGA database. Subsequently, the functional enrichment analysis was performed and a protein-protein interaction (PPI) network was constructed. After interaction with 3 driver gene databases, 11 driver mutants were screened and visualized, furthermore, their

correlations with clinical characteristics and survival were evaluated. The present study aimed to identify mutant genes related to clinical prognosis in PLC, searching for promising molecular targets in tumor progression, and develop more efficacious targeted agents in PLC therapy.

2. Patients and Methods

2.1. Patients and Samples

Eight patients with PLC admitted to the Affiliated Hospital of Chengde Medical University from 2018 to 2020 were recruited, including 5 males and 3 females with an average age of 59.500 ± 15.464 years. Of them, 7 had late-stage hepatocellular carcinoma (HCC) tumor and 1 had early-stage intrahepatic cholangiocarcinoma (ICC) tumor. This study was approved by the Ethics Review Committee of Chengde Medical University, and all patients provided written informed consent.

The mutation status of 396 PLC cases filtered for "primary tumor" and "liver and intrahepatic bile ducts" were downloaded from the official website of TCGA (https://portal.gdc.cancer.gov/). Of 396 patients, 254 with complete clinical data were included, detailed information of who was obtained to analyze the correlation between mutation and clinicopathological characteristics or overall survival.

Peripheral blood (5 ml) was drawn in EDTA tubes and processed immediately (centrifugation 1500 g, 5 min, 4˚C) to collect plasma and buffy coats, which were aliquoted and stored at −80˚C until further use.

Postoperative paraffin-embedded tissue sections were collected, 10 pieces with a thickness of 5 μm, and stored in a refrigerator at −20˚C for use. DNA was isolated from whole blood, plasma, and tissue using the QIAampDNA Mini Kit (Qiagen, Hilden, Germany) following the manufacturer instructions.

2.2. Library Construction and Target Capture Sequencing

The genomic DNA samples were broken into 150 - 200 bp fragments and ligated to Illumina sequencing adaptors to build a sequencing library using Accel-NGS 2S Plus DNA library kit (Swift Biosciences, Ann Arbor, Michigan) in strict accordance with the instructions. The captured DNA fragments were sequenced on the Illumina NextSeq500 Sequencer (Illumina, California, USA) using a cancer-related gene panel consisting of 551 widely known cancer-associated genes.

2.3. Identification of Mutant Genes

The mutation genes of 8 clinical cases were visualized using R software, in which each row represented a mutated gene, and column represented a patient. The mutation types of the genes were exhibited, including single nucleotide polymorphic (SNP), single nucleotide variants (SNV), and insertion/deletion (InDel).

TCGA PLC cancer data were screened by ticking "primary tumor" and "liver

and intrahepatic bile ducts" options from the official TCGA website. In the "Exploration" pattern, the "Mutations" and "OncoGrid" were selected sequentially, and then the waterfall maps of the top 50 mutated genes in 200 most mutated patients were downloaded.

Intersection of the mutation genes in 8 clinical cases and the top 100 mutated genes from the TCGA dataset was taken using Venn online website (http://jvenn.toulouse.inra.fr/app/example.html) [8].

2.4. Functional Enrichment Analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) database (http://david.ncifcrf.gov, version 6.8) was performed for Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the mutated genes. GO is a biological model framework, which describes gene functions and divides them into three parts: biological process (BP), cellular component (CC), and molecular function (MF). KEGG is an information resource for understanding biological systems and genomic functions at the molecular level. The ggplot2 program package of R software was loaded to visualize the enrichment analysis results obtained from the DAVID database.

2.5. Construction of PPI Network

PPI network was constructed after importing the name list of mutated genes from the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) online database (https://www.string-db.org/), and at the same time, the species source was limited to Homo sapiens [9].

PPI network data was imported into Cytoscape software (version 3.7.2). A functional module was identified using the Molecular Complex Detection application (MCODE) plug-in. Next, using the cytoHubba plug-in, the top 10 hub genes were selected according to the maximal clique centrality (MCC) method.

2.6. Screening and Visualization of Driver Genes

An intersection between the identified mutant genes and the driver genes from three databases, including Vogelstein's list [10], the Network of Cancer Genes (NCG) [11], the Catalog of Somatic Mutations in Cancer database (COSMIC) [12] was taken using the Venn online website.

2.7. Statistical Analysis

All statistical results were conducted using SPSS version 25.0. The Chi-square test and Fisher's exact test were used to compare the differences in categorical variables. Factors with $p < 0.100$ in the Chi-square test were included in logistic regression to analyze the interactions between mutation status and clinicopathological characteristics. Kaplan-Meier (KM) survival analysis and the log-rank test were used to estimate patient survival and quantify the differences between

groups. Factors with $p < 0.100$ in KM survival analysis were included in multivariate Cox analysis to identify independent prognostic factors. The results are presented as estimated odds ratios (OR) or hazard ratios (HR) with respective 95.000% confidence interval (95% CI) and p-values. The value of $p < 0.050$ was regarded as statistically significant.

3. Results

3.1. Identification of 17 Mutated Genes

Eighty-four mutated genes were identified in the blood and tissue samples from 8 PLC patients through the high-throughput detection of 551 panoramic cancer genes. As shown in **Figure 1(A)**, the most common mutation type was SNP (73.362%), followed by SNV (20.524%) and InDel (6.114%). The five most frequently mutated genes were CDA, ERCC1, SLC19A1, NOS3, ABCB1. A waterfall map, exhibiting the top 50 mutated genes in 200 most mutated PLC cases, was downloaded from TCGA of the Genomic Data Commons (GDC) data portal

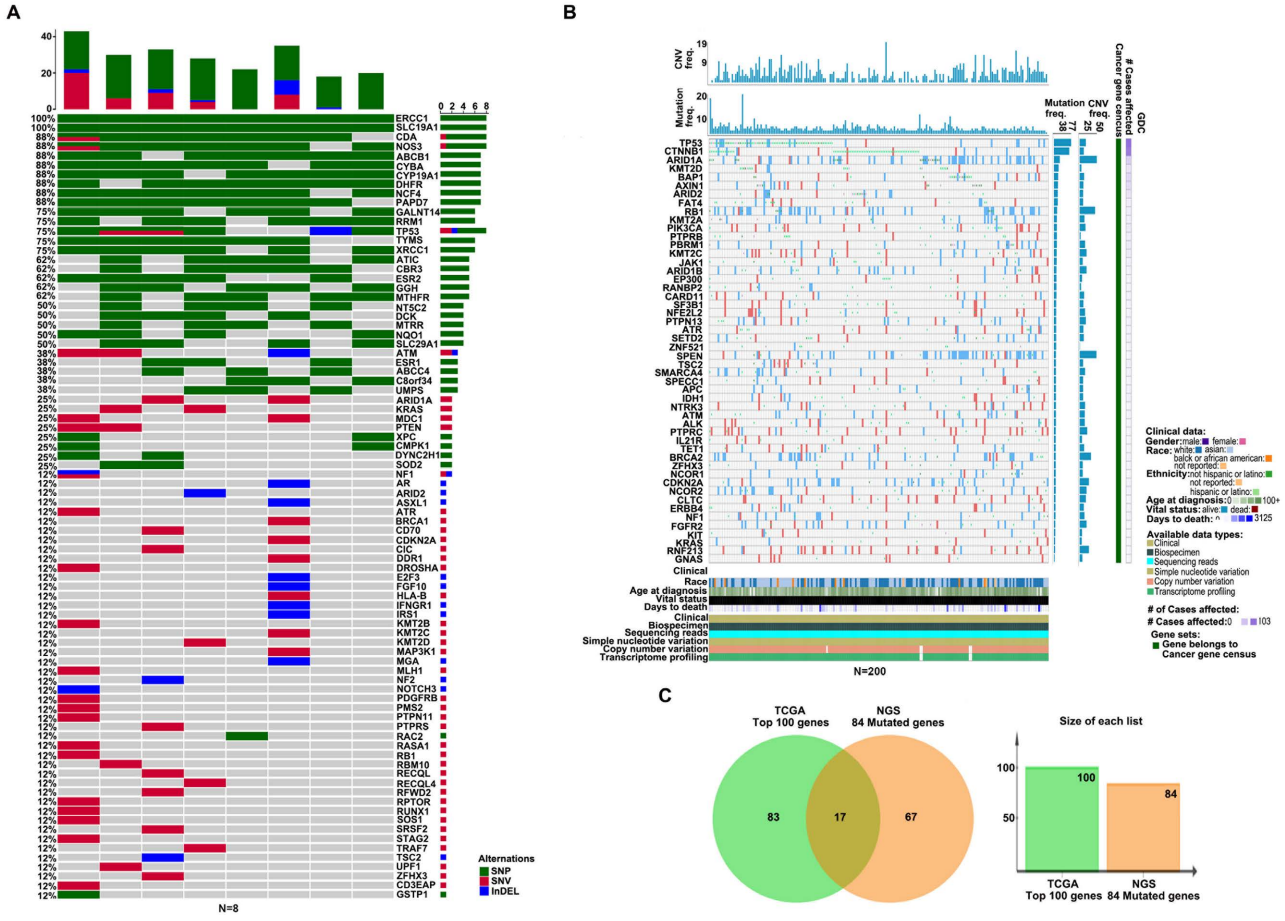

Figure 1. Visualization and screening for common mutants in PLC. (A) Genetic waterfall plot of 8 PLC patient mutations identified by NGS technology. (B) Waterfall plot showing the top 50 mutant genes in 200 PLC patients from TCGA database. The two rows at the top were CNV and mutation frequencies of each case, and the middle panel represents gene mutation patterns with different mutation types color coded differently. (C) Two-set Venn diagram showing the 17 common mutated genes from 8 clinical cases and TCGA patients.

(Figure 1(B)). Among them, the most commonly mutated gene was TP53 (36.500%), CTNNB1 (33.500%), and ARID1A (12.000%). Other genes had a mutation prevalence of <10.000%. Afterwards, the intersection of 84 mutated genes from 8 patients and the top 100 most frequently mutated genes from the TCGA database were analyzed by Venn software. Ultimately, 17 common mutated genes were obtained (Figure 1(C)).

3.2. Functional and Interactional Analysis of 17 Mutant Genes

To extensively investigate the function and mechanism of the 17 mutated genes, GO and KEGG analyses were performed using the DAVID online application (Figures 2(A)-(D)). The GO analysis results demonstrated that the 17 mutated genes were significantly enriched for regulation of nucleobase-containing compound metabolic process, regulation of nitrogen compound metabolic process, and cellular macromolecule biosynthetic process in the BP category. With regards to CC analysis, 17 mutated genes were markedly enriched in nucleoplasm, chromosome, and nucleoplasm part. The top 3 significantly enriched MF terms included heterocyclic compound binding, organic cyclic compound binding, and DNA binding. Furthermore, 17 mutated genes were mainly enriched in P53 signaling pathway, cell cycle, and HTLV-I infection in the KEGG analysis.

To explore the interaction between the 17 mutant genes, a PPI network was constructed using the STRING database. As shown in Figure 2(E), a total of 17 nodes and 76 edges were mapped in the PPI network, with a local clustering coefficient of 0.795 and a PPI enrichment p value < $1.0e^{-16}$. Then, a significant module, consisting of 10 nodes and 44 edges, with a score of 9.778, was identified by the MCODE plugin of Cytoscape (Figure 2(F)). Using the cytoHubba plugin and MCC algorithm, the top 10 hub genes were identified from the network (Figure 2(G)).

3.3. Screening of 11 Driver Genes

The three databases, Vogelstein's list, NCG database, and COSMIC database, including 125, 711 and 576 driver genes, respectively, are commonly studied for the analysis of tumor driver genes. After analyzing the intersection of 17 mutant genes and driver genes in the above 3 well-known databases, 11 driver genes were screened out, including ARID1A, ARID2, ATM, CDKN2A, KRAS, NF1, NF2, PTEN, RB1, STAG2, TP53 (Figure 3(A)). Mutant status of the 11 driver genes in 8 PLC patients and 396 cases from the TCGA database were visualized using the MafTools routine package of the R language in Figure 3(B) and Figure 3(C), respectively. It is worth mentioning that the mutation rates of TP53 gene were 75.000% and 29.000% in the above two cohorts, respectively, with the highest mutation frequency.

3.4. Correlation of Driver Gene Mutations with Clinicopathological Characteristics

After integrating 8 clinical patients and 254 TCGA cases with completed clini-

copathological data, the correlations between mutations for 11 driver genes and clinicopathological features were analyzed using the Chi-square test. The results

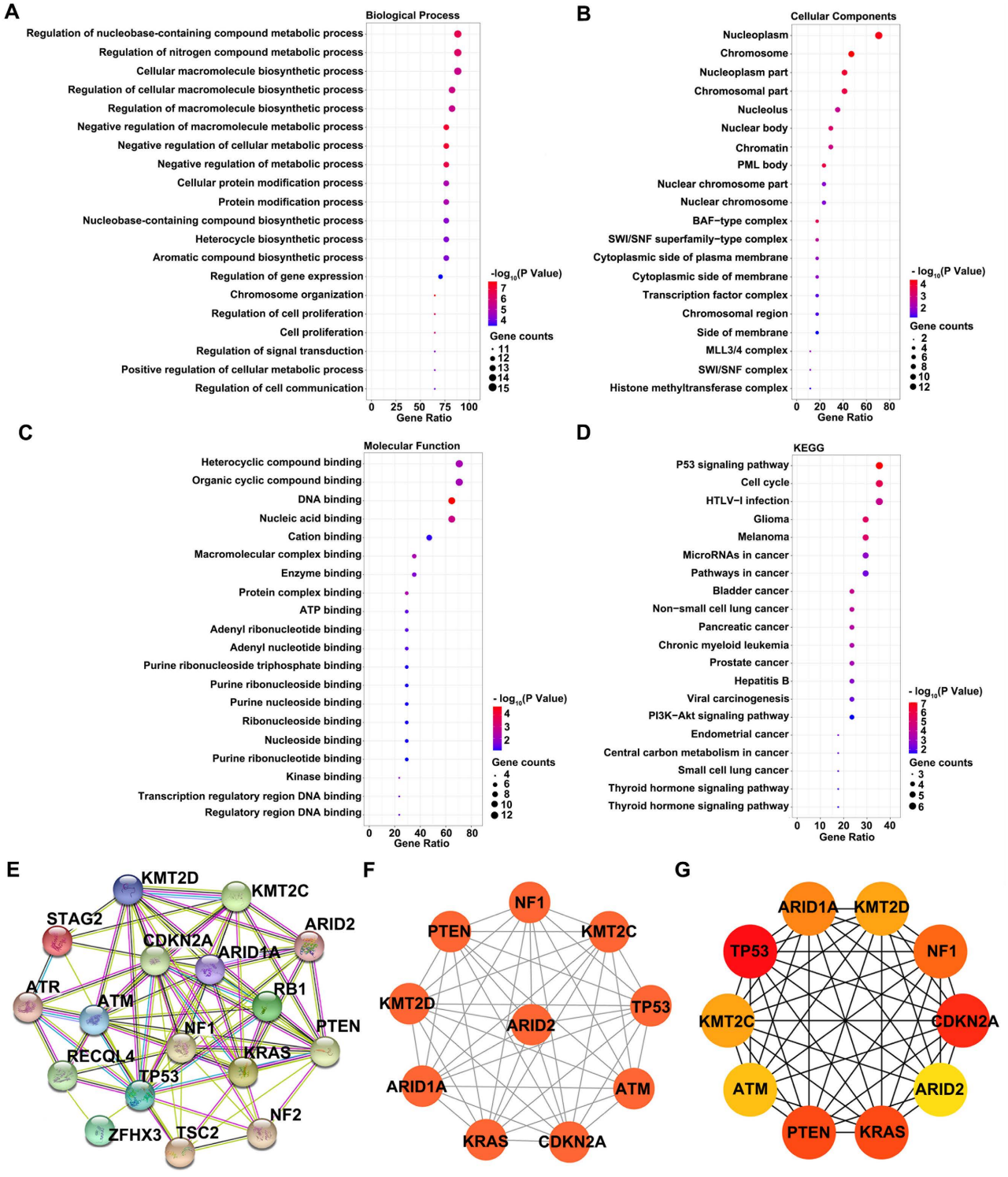

Figure 2. Functional enrichment analysis and PPI network construction of 17 mutated genes. (A) The top 20 significantly enriched biological process. (B) The top 20 significantly enriched cell component. (C) The top 20 significantly enriched molecular function. (D) The top 20 significantly enriched KEGG pathway. (E) PPI network of 17 mutant genes. (F) A significant module consisting of 10 nodes and 44 edges in the PPI network. (G) The top 10 hub genes in the PPI network.

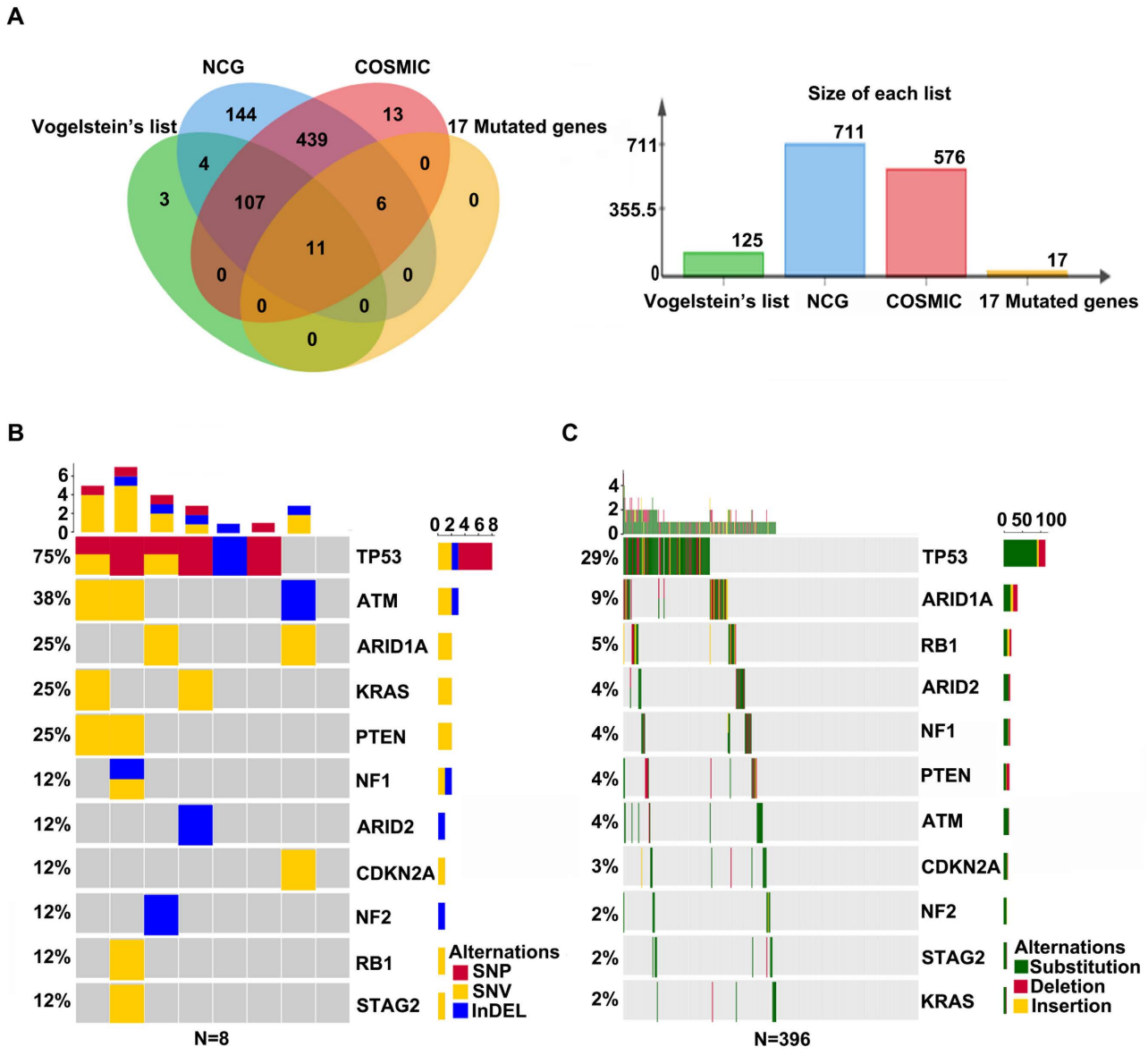

Figure 3. Screening and visualization of 11 driver genes. A: The driver genes in the three databases of Vogelstein's list, NCG, and COSMIC were intersected with 17 mutated genes, and 11 driver genes were obtained by Venn diagram. (B) and (C) depict the distributions of 11 driver genes in 8 clinical patients and 396 TCGA cases, respectively.

showed that 6 driver genes, including ARID2, CDKN2A, KRAS, NF1, RB1, TP53, were statistically significantly associated with clinical or pathological features (Table 1) ($p < 0.100$), while other 5 driver genes, including ARID1A, ATM, NF2, PTEN, STAG2, had no correlation with clinicopathological parameters (Supplementary Table S1) ($p > 0.100$).

To further explore the role of the 6 driver genes with statistical significance, logistic regression analysis was used to assess the relationship between mutation status of driver genes and clinicopathological characteristics (Table 2). First, we assessed the influence of age, sex, and race on ARID2, CDKN2A, RB1, and TP53 gene mutations. Patients under the age of 60 were prone to RB1 mutations ($p =$

Table 1. Six statistically significant driver genes with clinical characteristics of 262 PLC patients.

Clinical characteristics	No. of patients	ARID2			CDKN2A			KRAS			NF1			RB1			TP53		
		Mut	χ^2	p	Mut	χ^2	p	Mut	χ^2	p	Mut	χ^2	p	Mut	χ^2	p	Mut	χ^2	p
Total	262																		
Sex																			
Male	169	10	0.439	0.508	8	0.068	0.795	6	0.000	1.000	8	1.566	0.211	8	2.416	0.120	55	5.192	0.023#
Female	93	3			3			3			8			9			18		
Age																			
<60	134	7	0.040	0.842	4	0.157	0.692	4	0.005	0.944	8	0.009	0.925	13	4.666	0.031#	49	10.340	0.001#
≥60	128	6			6			5			8			4			24		
Race																			
Asian	156	11	3.570	0.059#	11	6.147	0.013#	3	1.650	0.199	9	0.077	0.782	13	2.163	0.141	62	27.080	<0.001#
non-Asian	106	2			0			6			7			4			11		
Pathological typing																			
HCC	225	13			9			5			11			16			69		
ICC	34	0	2.020	0.338	2	1.130	0.685	4	7.027	0.028#	4	6.138	0.038#	0	5.632	0.047#	3	8.015	0.012#
cHCC-ICC	3	0			0			0			1			1			1		
Primary site																			
Liver	228	13	1.010	0.315	9	0.004	0.947	5	5.541	0.019#	12	0.417	0.151	17	1.621	0.203	70	7.046	0.008#
Intrahepatic bile duct	34	0			2			4			4			0			3		
Treatment type																			
Pharmaceutical therapy, NOS	132	7	0.066	0.798	4	0.903	0.342	5	0.000	1.000	8	0.001	0.975	8	0.080	0.777	42	2.071	0.150
Radiation therapy, NOS	130	6			7			4			8			9			31		
T stage																			
T1 - T2	183	10	0.068	0.795	9	0.301	0.584	5	0.338	0.561	12	0.033	0.855	10	1.049	0.306	51	0.000	0.997
T3 - T4	79	3			2			4			4			7			22		
M stage																			
M0	248	13	0.061	0.805	10	0.000	1.000	7	2.363	0.124	15	0.000	1.000	16	0.000	1.000	68	0.135	0.713
M1	14	0			1			2			1			1			5		
N stage																			
N0	252	13	0.000	1.000	10	0.017	0.897	8	0.077	0.782	16	0.022	0.882	17	0.038	0.846	71	0.042	0.837
N1	10	0			1			1			0			0			2		
Stage																			
I - II	177	10	0.190	0.663	9	0.494	0.482	5	0.177	0.674	12	0.431	0.512	10	0.633	0.426	50	0.040	0.841
III - IV	85	3			2			4			4			7			23		

Categorical variables were compared using the Chi-square test or Fisher's exact test, #$p < 0.1$. Mut, mutated type; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; cHCC-ICC, combined hepatocellular carcinoma and intrahepatic cholangiocarcinoma.

Table 2. Analysis of correlation between mutated genes and clinicopathological characteristics.

clinical characteristic	ARID2			CDKN2A			KRAS			NF1			RB1			TP53		
	P	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI
Age (<60 vs. ≥60)		–			–			–			–		0.040*	3.331	1.056 - 10.501	0.002*	2.498	1.418 - 4.400
Sex (Male vs. Female)		–			–			–			–			–		0.024*	2.010	1.096 - 3.688
Race (Asian vs. non-Asian)	0.078	3.945	0.856 - 18.173	0.060	7.310	0.922 - 57.810		–			–			–		<0.001*	5.696	2.824 - 11.491
Primary site (Liver vs. Bile duct)		–			–		0.011*	0.168	0.043 - 0.661		–			–		0.014*	4.578	1.354 - 15.476
Pathological typing																		
HCC								Reference			Reference			Reference			Reference	
ICC		–			–		0.011*	0.170	0.043 - 0.670	0.122	0.386	0.115 - 1.288	–	1.950E+07	1.950E+07	0.015*	4.571	1.351 - 15.458
cHCC-ICC							–	5.450E+06	5.450E+06	0.072	0.103	0.009 - 1.222	0.134	0.153	0.013 - 1.781	0.921	0.885	0.079 - 9.920

HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; cHCC-ICC, combined hepatocellular carcinoma and intrahepatic cholangiocarcinoma; OR, odds ratio; CI, confidence interval. *$p < 0.05$.

0.040), while those over 60 years of age were likely to have TP53 mutations (p = 0.002). Male patients are more likely to have TP53 mutations compared with female patients (p = 0.024). With regard to, Asian cases had a higher incidence of TP53 mutations than non-Asians cases ($p < 0.001$). In addition, we also analyzed the influence of KRAS and TP53 driver gene mutations on primary site and pathological typing. Compared to wild-type cases, patients with KRAS mutations had a 5.952-fold greater risk of primary site locating bile duct (p = 0.011) and had a 5.882-fold higher risk of developing HCC (p = 0.011). Patients with TP53 gene mutations had a 4.578-fold greater risk of the primary site locating liver (p = 0.014) and had a 4.571-fold higher risk of developing ICC, than wild-type patients (p = 0.015).

3.5. Survival Analysis of Driver Gene Mutations in PLC

As shown in **Figure 4(A)** and **Figure 4(B)**, KM survival analysis showed that among the 11 driver genes, NF2 (p = 0.092) and RB1 (p = 0.094) mutations associated with poor patient prognoses. In addition, the PLC patients with T3 - T4 stage ($p < 0.001$) or late clinical stage ($p < 0.001$) have relatively short survival time (**Figure 4(C)**, **Figure 4(D)**). Based on a criterion of $p < 0.100$ in the KM analysis, multivariate Cox analysis was carried out. The results revealed that all the above 4 parameters, including NF2 (p = 0.117), RB1 (p = 0.185), T stage (p = 0.873), clinical stage (p = 0.365), were not the independent prognosis factors in PLC patients (**Figure 4(E)**).

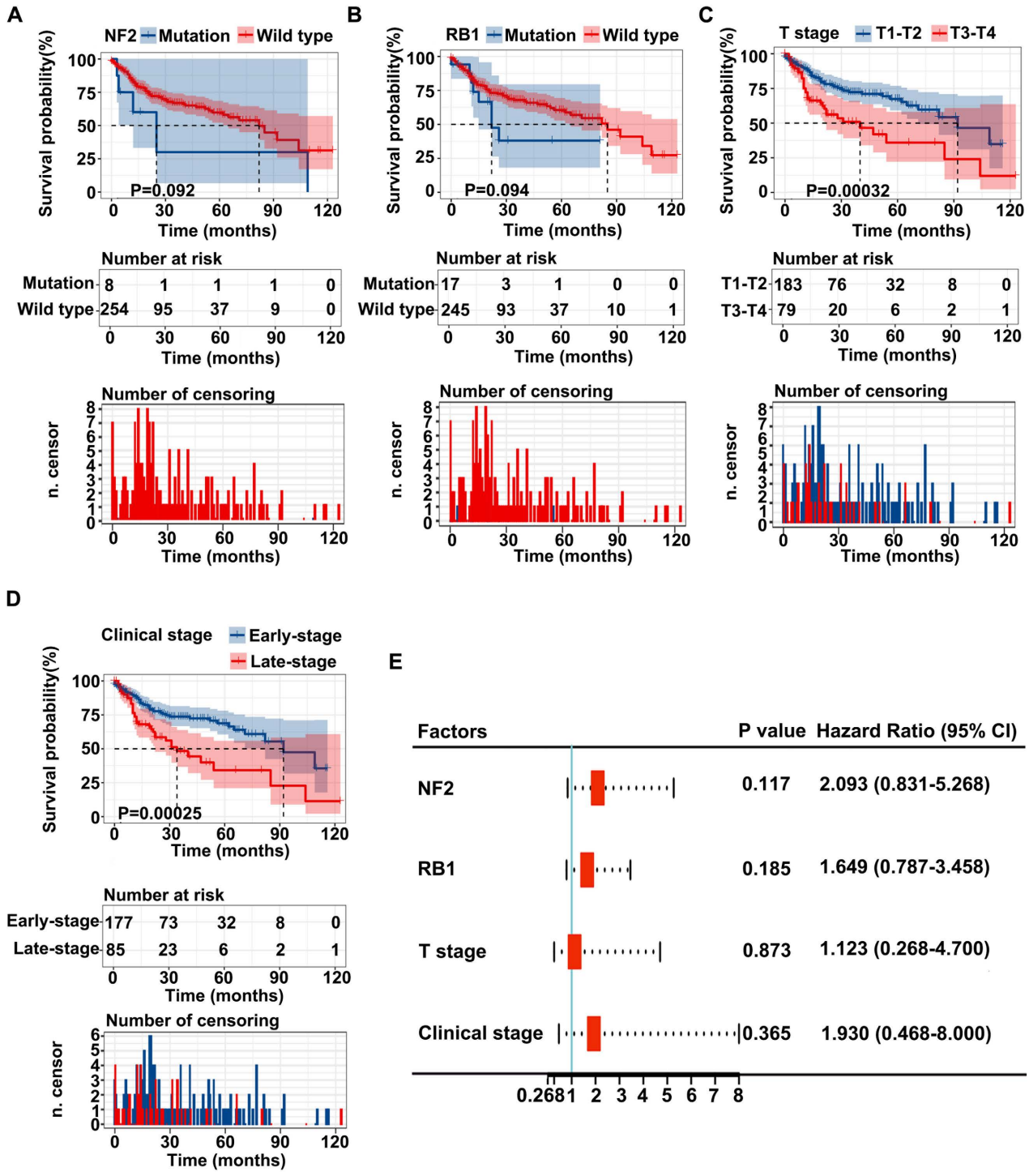

Figure 4. Overall survival analysis for driver genes and clinical characteristics. The KM curves of (A) NF2 mutation, (B) RB1 mutation, (C) T stage, and (D) clinical stage. E: Multivariate Cox analysis for NF2 mutation, RB1 mutation, T stage, and clinical stage.

4. Discussion

The occurrence and development of liver cancer is a complex process involving multiple genetic events, diverse etiologies, and genomic heterogeneity, which manifests not only in different patients but also in individual tumor nodules

from a single case. NGS technologies have allowed a high-throughput, comprehensive characterization of cancer genomes at unprecedented rates, which could improve the cancer genetic map and our understanding of the genetic landscape in liver cancer [13]. In Morishita's study, 50 genes associated with tumor development were targeted, and the relationship between the genetic mutations and the clinical characteristics of HCC patients was investigated using an NGS platform [14]. Lu *et al.* [15] provided a mutation spectrum of HCC tissue in 12 western Chinese cases using NGS with a panel of 372 cancer-associated genes, assisting in the investigation of the mechanism of liver carcinogenesis. Kan *et al.* [16] conducted whole-genome sequencing analysis on 88 HCC patients by the NGS technology, which not only verified multiple gene mutation sites that had been reported but also found new sites of BRCA2 and IGF1R gene mutations. In this study, a panel containing 551 cancer-associated genes was used in the NGS platform to analyze the gene mutations of 8 PLC cases, which is a relatively large panel up to now, providing a comprehensive genetic landscape survey of PLC patients.

A total of 84 mutant genes were identified, corroborating the progression of PLC development is the accumulation of multiple genetic events. Given the number of our cases analyzed is low, a PLC TCGA dataset was included in this study. After integrating the two cohorts, 8 clinical PLC cases and 396 PLC TCGA patients, 17 common mutated genes were obtained. Among them, 12 genes, including TP53, ARID1, ARID2, ATM, ATR, CDKN2A, KMT2C, KRAS, NF1, PTEN, RB1, and RECQL4, has been reported previously in NGS-based study of the liver cancer genome [17] [18] [19], while other 5 gene mutations, including KMT2D, NF2, STAG2, TSC2, ZFHX3, have never been verified in PLC, which remain to be further explored. Of the above 17 mutant genes, TP53 has a high mutation rate of 29.455% (119/404). Consistent with our data, the NGS result in 59 liver cancer tissues showed the most mutated gene was also TP53 (35.600%) [14]. In addition, TP53 was the most frequently mutated gene in 12 HCC patients studied by Lu *et al.* [15], with mutation rates reported up to 41.670%.

In order to further investigate the biological function and potential pathways of the 17 mutated genes, GO and KEGG analysis were performed using David online analysis platform. Notably, the result of the GO analysis showed that the 17 mutation genes were mainly enriched in metabolic process and macromolecule biosynthetic process. As we know, the liver is an important metabolic organ of the human body, and liver dysfunction induces intracellular redox imbalance, leading to the damage of intracellular biomolecules. Recent studies have reported that metabolic rearrangement contributes to the increased risk of PLC. Ikeno *et al.* [20] demonstrated GLUT-1 expression was significantly higher in tumors with mutated KRAS than in tumors with wild-type KRAS. High metabolic tumor volume is associated with KRAS mutation and poor postoperative outcomes in ICC patients. To meet the metabolic requirements for cancer cell growth, the de novo nucleotide synthetic pathway is activated to support the biology activities of cancer cells, including nucleic acid and protein synthesis,

energy preservation, signaling activity, glycosylation mechanisms, and cytoskeletal function. Both oncogenes and tumor suppressors have been identified as key molecular determinants for de novo nucleotide synthesis. Oncogenic KRAS maintains high intracellular nucleotide levels by enhancing de novo synthesis of purines and pyrimidines through upregulating MYC-mediated transcriptional activation of ribose-5-phosphate isomerase A (RPIA), which has been shown to play a crucial role in the development of HCC [21]. Additionally, the inactivation of tumor suppressor TP53 has been shown to fuel nucleotide synthesis in tumor cells that contributes to the maintenance of homeostasis and the proliferation of cancer cells. Moreover, the top 2 significantly enriched KEGG terms were p53 signaling pathway and cell cycle. The tumor suppressor p53 plays a major role in cell cycle arrest and/or apoptosis, and p53 mutations and functional inactivation are linked to the pathology of PLC. He's results showed that abnormal expression of p53 and cyclinD1 can lead to the progression of HCC by regulating G_1/S transformation [22] [23].

Genes with acquired mutations or abnormal expression that are causally associated with cancer progression are called driver genes. Identifying and understanding genetic driver mutations dramatically facilitates the development of targeted cancer therapies. For this reason, we screened 11 driver genes by comparing the 17 mutated genes with the data from 3 databases, including Vogelstein's list, NCG database, and COSMIC database, which have been often utilized to identify driver genes. In 2013, Bert Vogelstein reviewed 125 driver genes containing 71 tumor suppressor genes and 54 oncogenes, which had been defined by the 20/20 rule in a total of 294,881 mutations [9]. Since then, in many researches, the driver genes in Vogelstein's list have been regarded as well-known oncogenes and tumor suppressor genes undergoing copy number alterations in the common solid tumors, such as colon, lung, prostate, breast, etc. [24] [25] [26]. The NCG is an open-access database of 2372 genes, consisting of 1661 predicted driver genes in cancer and 711 known cancer-driving genes which contains data on gene mutations [11]. Bioinformatic analysis of an HCC dataset from the Gene Expression Omnibus (GEO) database was performed to identify cancer-related genes which were followed by imported into the NCG database and identified several driver genes, including ATC, CCND1, CREBBP, FTCD, MDH2, PPPARG, and TP53 [27]. The COSMIC database is currently the most comprehensive database of mutations in cancer, containing 576 securable driver genes [12]. El-Ayadi *et al.* [28] firstly reported a medulloblastoma case with concurrent IDH1 and SMARCB1 mutations after searching the catalog of somatic mutations in the COSMIC database. In this study, the above 3 databases were combined to screen driver genes from 17 mutation genes, which significantly improved accuracy and reliability.

The correlations between the 11 driver genes and clinicopathological characteristics of 262 PLC patients were performed by logistic regression analysis. The results showed that only 3 mutant genes, RB1, TP53, and KRAS had statistical

significance. The RB1 gene, which was the first tumor suppressor gene identified, has a negative regulation function on cell growth and can promote cell death. In this study, we found that patients under the age of 60 were prone to RB1 mutations, and the same result was also mentioned in Chaudhary's study [29]. To reveal the possible link between 10 driver genes and age, they conducted Mann-Whitney-Wilcoxon tests for the continuous age variable in 6 HCC cohorts. The results confirmed that RB1 is the driver gene significantly and preferably prevalent in younger patients. In addition, Chaudhary *et al.* also demonstrated that TP53 was preferred in males, and in terms of associations with race, TP53 showed higher relative risk in Asians. This coincides with our study, in which male or Asian patients were more likely to have TP53 mutations. TP53, as a well-known tumor suppressor gene, seems to have a higher risk of developing HCC in this study, however, Hill *et al.* [30] reported TP53 loss enhanced reprogramming of hepatocytes to biliary cells, which may be a mechanism facilitating the formation of hepatocyte-derived ICC. Another mutant gene with statistical significance, KRAS, increased the risk of ICC, which has been supported by Levi S. In their study, all the 15 cholangiocarcinomas patients showed a KRAS mutation at codon 12, and 9 of them contained 2 or more mutations [31].

5. Conclusion

The pattern of genetic alterations in cancer driver genes in PLC patients is highly diverse, which partially explains the low efficacy of available therapies. Therefore, it may be a new option to try to use NGS technology to find the driver genes and carry out targeted drug delivery. In this study, 17 mutated genes and 11 mutation driver genes in PLC were identified through bioinformatic analysis of TCGA PLC data and NGS detection of clinical samples. Among them, RB1, TP53, and KRAS have relationships with clinicopathological characteristics, including age, gender, race, primary sites, and pathological type. This study provided significant clues and basis for further understanding the molecular pathogenesis, drug development, and treatment of PLC.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant number 81703001), Hebei Province Medical Science Research Project (grant number 20210247), Chengde Medical University Scientific Research Major Projects (grant number KY2020005), Key Laboratory of Family Planning and Eugenics of National Health and Family Planning Commission (grant number 201502), Hebei Province Key Research and Development Projects (grant number 19277783D) and Project for Science and Technology Innovation Guidance Fund of Hebei Provincial Department of Science and Technology.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] Lin, D.C., Mayakonda, A., Dinh, H.Q., Huang, P., Lin, L., Liu, X., Ding, L.W., Wang, J., Berman, B.P., Song, E.W., Yin, D. and Koeffler, H.P. (2017) Genomic and Epigenomic Heterogeneity of Hepatocellular Carcinoma. *Cancer Research*, **77**, 2255-2265. https://doi.org/10.1158/0008-5472.CAN-16-2822

[2] Greten, T.F., Lai, C.W., Li, G. and Staveley-O'Carroll, K.F. (2019) Targeted and Immune-Based Therapies for Hepatocellular Carcinoma. *Gastroenterology*, **156**, 510-524. https://doi.org/10.1053/j.gastro.2018.09.051

[3] Wang, H., Lu, Z. and Zhao, X. (2019) Tumorigenesis, Diagnosis, and Therapeutic potential of Exosomes in Liver Cancer. *Journal of Hematology & Oncology*, **12**, Article No. 133. https://doi.org/10.1186/s13045-019-0806-6

[4] Wu, K., Huang, R.S., House, L. and Cho, W.C. (2013) Next-Generation Sequencing for Lung Cancer. *Future Oncology*, **9**, 1323-1336. https://doi.org/10.2217/fon.13.102

[5] Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nature Reviews Genetics*, **10**, 57-63. https://doi.org/10.1038/nrg2484

[6] Gao, J., Ciriello, G., Sander, C. and Schultz, N. (2014) Collection, Integration and Analysis of Cancer Genomic Profiles: From Data to Insight. *Current Opinion in Genetics & Development*, **24**, 92-98. https://doi.org/10.1016/j.gde.2013.12.003

[7] Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge. *Contemporary Oncology*, **19**, A68-A77. https://doi.org/10.5114/wo.2014.47136

[8] Bardou, P., Mariette, J., Escudié, F., Djemiel, C. and Klopp, C. (2014) Jvenn: An Interactive Venn Diagram Viewer. *BMC Bioinformatics*, **15**, Article No. 293. https://doi.org/10.1186/1471-2105-15-293

[9] Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering. C. and Jensen, L.J. (2013) STRING v9.1: Protein-Protein Interaction Networks, with Increased Coverage and Integration. *Nucleic Acids Research*, **41**, D808-D815. https://doi.org/10.1093/nar/gks1094

[10] Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz Jr., L.A. and Kinzler, K.W. (2013) Cancer Genome Landscapes. *Science*, **339**, 1546-1558. https://doi.org/10.1126/science.1235122

[11] Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S.K., Tourna, A., Yakovleva, A., Palmieri, T. and Ciccarelli, F.D. (2019) The Network of Cancer Genes (NCG): A Comprehensive Catalogue of Known and Candidate Cancer Genes from Cancer Sequencing Screens. *Genome Biology*, **20**, Article No. 1. https://doi.org/10.1186/s13059-018-1612-0

[12] Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., Teague, J.W., Campbell, P.J., Stratton, M.R. and Futreal, P.A. (2011) COSMIC: Mining Complete Cancer Genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research*, **39**, D945-D950. https://doi.org/10.1093/nar/gkq929

[13] Alekseyev, Y.O., Fazeli, R., Yang, S., Basran, R., Maher, T., Miller, N.S. and Remick, D. (2018) A Next-Generation Sequencing Primer—How Does It Work and What Can It Do? *Academic Pathology*, **5**, 1-11. https://doi.org/10.1177/2374289518766521

[14] Morishita, A., Iwama, H., Fujihara, S., Watanabe, M., Fujita, K., Tadokoro, T., Ohura, K., Chiyo, T., Sakamoto, T., Mimura, S., Nomura, T., Tani, J., Yoneyama, H., Okano, K., Suzuki, Y., Himoto, T. and Masaki, T. (2018) Targeted Sequencing of

Cancer-Associated Genes in Hepatocellular Carcinoma Using Next-Generation Sequencing. *Oncology Letters*, **15**, 528-532. https://doi.org/10.3892/ol.2017.7334

[15] Lu, J., Yin, J., Dong, R., Yang, T., Yuan, L., Zang, L., Xu, C., Peng, B., Zhao, J. and Du, X. (2015) Targeted Sequencing of Cancer-Associated Genes in Hepatocellular Carcinoma Using Next Generation Sequencing. *Molecular Medicine Reports*, **12**, 4678-4682. https://doi.org/10.3892/mmr.2015.3952

[16] Kan, Z., Zheng, H., Liu, X., Li, S., Barber, T.D., Gong, Z., Gao, H., Hao, K., Willard, M.D., Xu, J., Hauptschein, R., Rejto, P.A., Fernandez, J., Wang, G., Zhang, Q., Wang, B., Chen, R., Wang, J., Lee, N.P., Zhou, W., Lin, Z., Peng, Z., Yi, K., Chen, S., Li, L., Fan, X., Yang, J., Ye, R., Ju, J., Wang, K., Estrella, H., Deng, S., Wei, P., Qiu, M., Wulur, I.H., Liu, J., Ehsani, M.E., Zhang, C., Loboda, A., Sung, W.K., Aggarwal, A., Poon, R.T., Fan, S.T., Wang, J., Hardwick, J., Reinhard, C., Dai, H., Li, Y., Luk, J.M. and Mao, M. (2013) Whole-Genome Sequencing Identifies Recurrent Mutations in Hepatocellular Carcinoma. *Genome Research*, **23**, 1422-1433. https://doi.org/10.1101/gr.154492.113

[17] Janku, F., Kaseb, A.O., Tsimberidou, A.M., Wolff, R.A. and Kurzrock, R. (2014) Identification of Novel Therapeutic Targets in the PI3K/AKT/mTOR Pathway in Hepatocellular Carcinoma Using Targeted Next Generation Sequencing. *Oncotarget*, **5**, 3012-3022. https://doi.org/10.18632/oncotarget.1687

[18] Li, M., Zhao, H., Zhang, X., Wood, L.D., Anders, R.A., Choti, M.A., Pawlik, T.M., Daniel, H.D., Kannangai, R., Offerhaus, G.J., Velculescu, V.E., Wang, L., Zhou, S., Vogelstein, B., Hruban, R.H., Papadopoulos, N., Cai, J., Torbenson, M.S. and Kinzler, K.W. (2011) Inactivating Mutations of the Chromatin Remodeling Gene ARID2 in Hepatocellular Carcinoma. *Nature Genetics*, **43**, 828-829. https://doi.org/10.1038/ng.903

[19] Anjanappa, M., Hao, Y., Simpson, E.R., Bhat-Nakshatri, P., Nelson, J.B., Tersey, S.A., Mirmira, R.G., Cohen-Gadol, A.A., Saadatzadeh, M.R., Li, L., Fang, F., Nephew, K.P., Miller, K.D., Liu, Y. and Nakshatri, H. (2018) A System for Detecting High Impact-Low Frequency Mutations in Primary Tumors and Metastases. *Oncogene*, **37**, 185-196. https://doi.org/10.1038/onc.2017.322

[20] Ikeno, Y., Seo, S., Iwaisako, K., Yoh, T., Nakamoto, Y., Fuji, H., Taura, K., Okajima, H., Kaido, T., Sakaguchi, S. and Uemoto, S. (2018) Preoperative Metabolic Tumor Volume of Intrahepatic Cholangiocarcinoma Measured by ^{18}F-FDG-PET Is Associated with the *KRAS* Mutation Status and Prognosis. *Journal of Translational Medicine*, **16**, Article No. 95. https://doi.org/10.1186/s12967-018-1475-x

[21] Saliani, M., Jalal, R. and Ahmadian, M.R. (2019) From Basic Researches to New Achievements in Therapeutic Strategies of KRAS-Driven Cancers. *Cancer Biology & Medicine*, **16**, 435-461.

[22] He, L., Fan, X., Li, Y., Chen, M., Cui, B., Chen, G., Dai, Y., Zhou, D., Hu, X. and Lin, H. (2019) Overexpression of Zinc Finger Protein 384 (ZNF 384), A Poor Prognostic Predictor, Promotes Cell Growth by Upregulating the Expression of Cyclin D1 in Hepatocellular Carcinoma. *Cell Death & Disease*, **10**, Article No. 444. https://doi.org/10.1038/s41419-019-1681-3

[23] Chen, S.L., Liu, L.L., Wang, C.H., Lu, S.X., Yang, X., He, Y.F., Zhang, C.Z. and Yun, J.P. (2020) Loss of RDM1 Enhances Hepatocellular Carcinoma Progression via p53 and Ras/Raf/ERK Pathways. *Molecular Oncology*, **14**, 373-386. https://doi.org/10.1002/1878-0261.12593

[24] Springer, S.U., Chen, C.H., Rodriguez, Pena. M.D.C., Li, L., Douville, C., Wang, Y., Cohen, J.D., Taheri, D., Silliman, N., Schaefer, J., Ptak, J., Dobbyn, L., Papoli, M., Kinde, I., Afsari, B., Tregnago, A.C., Bezerra, S.M., VandenBussche, C., Fujita, K.,

Ertoy, D., Cunha, I.W., Yu, L., Bivalacqua, T.J., Grollman, A.P., Diaz, L.A., Karchin, R., Danilova, L., Huang, C.Y., Shun, C.T., Turesky, R.J., Yun, B.H., Rosenquist, T.A., Pu, Y.S., Hruban, R.H., Tomasetti, C., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., Dickman, K.G. and Netto, G.J. (2018) Non-Invasive Detection of Urothelial Cancer through the Analysis of Driver Gene Mutations and Aneuploidy. *Elife*, **7**, Article No. e32143. https://doi.org/10.7554/eLife.32143

[25] Merid, S.K., Goranskaya, D. and Alexeyenko, A. (2014) Distinguishing between Driver and Passenger Mutations in Individual Cancer Genomes by Network Enrichment Analysis. *BMC Bioinformatics*, **15**, Article No. 308. https://doi.org/10.1186/1471-2105-15-308

[26] Tian, R., Basu, M.K. and Capriotti, E. (2014) ContrastRank: A New Method for Ranking Putative Cancer Driver Genes and Classification of Tumor Samples. *Bioinformatics*, **17**, i572-i857. https://doi.org/10.1093/bioinformatics/btu466

[27] Shangguan, H., Tan, S.Y. and Zhang, J.R. (2015) Bioinformatics Analysis of Gene Expression Profiles in Hepatocellular Carcinoma. *European Review for Medical and Pharmacological Sciences*, **19**, 2054-2061.

[28] El-Ayadi, M., Egervari, K., Merkler, D., McKee, T.A., Gumy-Pause, F., Stichel, D., Capper, D., Pietsch, T., Ansari, M. and Bueren. A.O. (2018) Concurrent IDH1 and SMARCB1 Mutations in Pediatric Medulloblastoma: A Case Report. *Frontiers in Neurology*, **9**, Article No. 398. https://doi.org/10.3389/fneur.2018.00398

[29] Chaudhary, K., Poirion, O.B., Lu, L., Huang, S., Ching, T. and Garmire, L.X. (2019) Multimodal Meta-Analysis of 1,494 Hepatocellular Carcinoma Samples Reveals Significant Impact of Consensus Driver Genes on Phenotypes. *Clinical Cancer Research*, **25**, 463-472. https://doi.org/10.1158/1078-0432.ccr-18-0088

[30] Hill, M.A., Alexander, W.B., Guo, B., Kato, Y., Patra, K., O'Dell, M.R., McCall, M.N., Whitney-Miller, C.L., Bardeesy, N. and Hezel, A.F. (2018) *Kras* and *Tp53* Mutations Cause Cholangiocyte- and Hepatocyte-Derived Cholangiocarcinoma. *Cancer Research*, **78**, 4445-4451. https://doi.org/10.1158/0008-5472.CAN-17-1123

[31] Levi, S., Urbano-Ispizua, A., Gill, R., Thomas, D.M., Gilbertson, J., Foster, C. and Marshall, C.J. (1991) Multiple K-ras Codon 12 Mutations in Cholangiocarcinomas Demonstrated with a Sensitive Polymerase Chain Reaction Technique. *Cancer Research*, **51**, 3497-3502.

Supplementary

Supplementary Table S1. Five driver genes that were not statistically significant with clinical characteristics of 262 PLC patients.

Clinical characteristics	No. of patients	ARID1A			ATM			NF2			PTEN			STAG2		
		Mut	χ^2	*p*	Mut	χ^2	*p*	Mut	χ^2	*p*	Mut	χ^2	*p*	Mut	χ^2	*p*
Total	262															
Sex																
Male	169	15	0.142	0.706	11	0.712	0.399	7	1.011	0.315	5	0.000	1.000	3	0.000	1.000
Female	93	7			3			1			3			2		
Age																
<60	134	14	1.500	0.221	7	0.008	0.930	5	0.086	0.769	2	1.307	0.253	4	0.725	0.394
≥60	128	8			7			3			6			1		
Race																
Asian	156	15	0.744	0.388	9	0.138	0.710	5	0.000	1.000	5	0.000	1.000	3	0.000	1.000
non-Asian	106	7			5			3			3			2		
Pathological typing																
HCC	225	19	0.270	1.000	13	0.584	0.747	7	0.921	1.000	6	2.200	0.347	3	4.294	0.179
ICC	34	3			1			1			2			2		
cHCC-ICC	3	0			0			0			0			0		
Primary site																
Liver	228	19	0.000	1.000	13	0.067	0.796	1	0.000	1.000	6	0.244	0.622	3	1.308	0.253
Intrahepatic bile duct	34	3			1			1			2			2		
Treatment type																
Pharmaceutical therapy, NOS	132	11	0.001	0.970	7	0.001	0.977	5	0.114	0.736	4	0.000	1.000	3	0.000	1.000
Radiation therapy, NOS	130	11			7			3			4			2		
T stage																
T1 - T2	183	13	1.319	0.251	8	0.586	0.444	3	0.669	0.102	5	0.005	0.945	3	0.000	1.000
T3 - T4	79	9			6			5			3			2		
M stage																
M0	248	19	1.721	0.190	12	0.843	0.358	7	0.013	0.908	8	0.000	1.000	4	0.219	0.640
M1	14	3			2			1			0			1		
N stage																
N0	252	21	0.000	1.000	13	0.000	1.000	8	0.000	1.000	7	0.133	0.715	5	0.000	1.000
N1	10	1			1			0			1			0		
Stage																
I - II	177	13	0.785	0.375	8	0.316	0.574	3	2.134	0.144	5	0.000	1.000	3	0.000	1.000
III - IV	85	9			6			5			3			2		

Categorical variables were compared using the Chi-square test or Fisher's exact test. Mut, mutated type; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; cHCC-ICC, combined hepatocellular carcinoma and intrahepatic cholangiocarcinoma.