

RESEARCH

Open Access



The effect of SNPs in lncRNA as ceRNA on the risk and prognosis of hepatocellular carcinoma

Han Mo^{1,2†}, Xi Wang^{3†}, Guohua Ji^{1,2}, Xiao Liang^{1,2}, Yi Yang^{1,2}, Wenjing Sun^{1,2}, Xueyuan Jia^{1,2}, Lidan Xu^{1,2}, Yuandong Qiao^{1,2}, Henan Zhou⁴, Wenhui Zhao⁵, Songbin Fu^{1,2} and Xuelong Zhang^{1,2*}

Abstract

Background: Most susceptible loci of hepatocellular carcinoma (HCC) identified by genome-wide association studies (GWAS) are located in non-coding regions, and the mechanism of action remains unclear. The objective of this study was to explore the association of single nucleotide polymorphisms (SNPs) on long non-coding RNAs (lncRNAs) that affect competing endogenous RNAs (ceRNA) regulation mechanism with the risk and prognosis of HCC.

Methods: Based on a set of bioinformatics strategies, eight lncRNA genes that affect HCC through the mechanism of lncRNA-mediated ceRNA were systematically screened, and 15 SNPs that affect microRNA (miRNA) binding in these lncRNA genes were annotated. Genotyping was performed in 800 HCC cases and 801 healthy controls to examine associations of these SNPs with HCC in a northeastern Chinese Han population.

Results: The GG, GC and GG + GC genotypes of *HOTAIR* rs7958904 were associated with a 0.65, 0.59 and 0.63-fold decreased HCC risk, respectively. In addition, HCC patients with *PVT1* rs3931282 AA + GA genotypes were less prone to develop late-stage cancers in a stratified analysis of clinical characteristics. When stratified by clinical biochemical indexes, rs1134492 and rs10589312 in *PVT1* and rs84557 in *EGFR-AS1* showed significant associations with aspartate aminotransferase (AST), alanine aminotransferase (ALT) or AST/ALT ratio in HCC patients. Furthermore, we constructed potential ceRNA regulatory axes that might be affected by five positive SNPs to explain the causes of these genetic associations.

Conclusions: *HOTAIR* rs7958904, *PVT1* rs3931282, rs1134492 and rs10589312, and *EGFR-AS1* rs84557 might be predictors for HCC risk or prognosis. Our results provide new insights into how SNPs on lncRNA-mediated ceRNAs confer interindividual differences to occurrence and progression of HCC.

Keywords: HOTAIR, PVT1, EGFR-AS1, HCC, Risk, Prognosis, SNP, ceRNA

Background

According to the Global Cancer Statistics 2020, primary liver cancer is the sixth most common cancer and the third leading cause of cancer mortality worldwide (new cases, 906,000/y; deaths, 830,000/y) [1]. The most common histology (about 80%) of primary liver cancer is hepatocellular carcinoma (HCC) [2]. Notably, East Asia is the region with the highest risk of HCC, with more than 50% of the world's HCC cases coming from China

[†]Han Mo and Xi Wang contributed equally to this work.

*Correspondence: zhangxuelong@hrbmu.edu.cn

¹Laboratory of Medical Genetics, Harbin Medical University, Harbin 150081, China

Full list of author information is available at the end of the article



[3]. Hepatocarcinogenesis is a multi-step process involving multiple risk factors in its occurrence, promotion and development [4]. Genetic variation plays an important role in these risk factors. In the past few decades, more and more genome-wide association studies (GWAS) have been applied to identify HCC susceptibility single nucleotide polymorphisms (SNPs) [5], most of which are located in non-coding regions [6]. Mutations in regulatory regions may lead to subtle changes in gene expression in cell type or tissue-specific manner, which may predispose mutation carriers to changes in cancer susceptibility throughout their life cycle [7–9]. Therefore, the mechanism of non-coding SNPs on HCC can be explained from the perspective of expression regulatory.

Long non-coding RNAs (lncRNAs) are defined as a class of transcripts with more than 200 nucleotides, which participate in transcription, post transcription, post-translational regulation and intercellular signal transduction [10]. lncRNAs serve as critical regulators of regulators of tumorigenesis and metastasis [11, 12]. MicroRNAs (miRNAs) are another small non-coding RNA molecules with a nucleotide length of about 22 that mainly mediate the process of post-transcriptional gene silencing [13]. As competing endogenous RNAs (ceRNAs), lncRNAs interact with miRNAs through complementary sequences and become the bait or sponge of miRNAs, which has been widely confirmed by functional studies [14–16]. With the development of high-throughput technologies, an increasing number of cancer-related studies have been published using The Cancer Genome Atlas (TCGA). HCC related lncRNA-mediated ceRNA networks have also been reported in recent TCGA-based articles [17].

It is well known that SNPs located in lncRNA-miRNA binding regions may affect their interactions, thereby altering their functions and leading to genome-wide butterfly effects [18]. Several evidences have indicated that these SNPs are involved in the occurrence and development of various cancers. For example, one study showed that rs2147578 CG and GG genotypes were significantly associated with an increased risk of colorectal cancer by influencing the combination of lnc-LAMC2-1:1 and miR-128-3p [19]. Another study confirmed that the LINC00673 polymorphism created a miR-1231 binding site and affected the risk of pancreatic cancer by interfering with PTPN11 degradation [9]. Encouragingly, a large number of SNPs located in lncRNA-miRNA binding regions have recently been identified by the lncRNASNP2 database, which is also supported by experimental evidence [20]. These SNPs are more likely to play a role by interfering with the ceRNA function of lncRNAs.

Thus, we suggest that SNPs affecting lncRNA-miRNA interaction may cause HCC risk. To further test the

hypothesis, we systematically screened out such potentially functional SNPs with a set of bioinformatics strategies. Next, a case-control study was conducted in a northeastern Chinese Han population to assess the relationship of these candidate SNPs with the occurrence and development of HCC. Finally, we discussed how positive SNPs affect HCC through the potential regulatory axis of ceRNA.

Results

Differential expression analysis

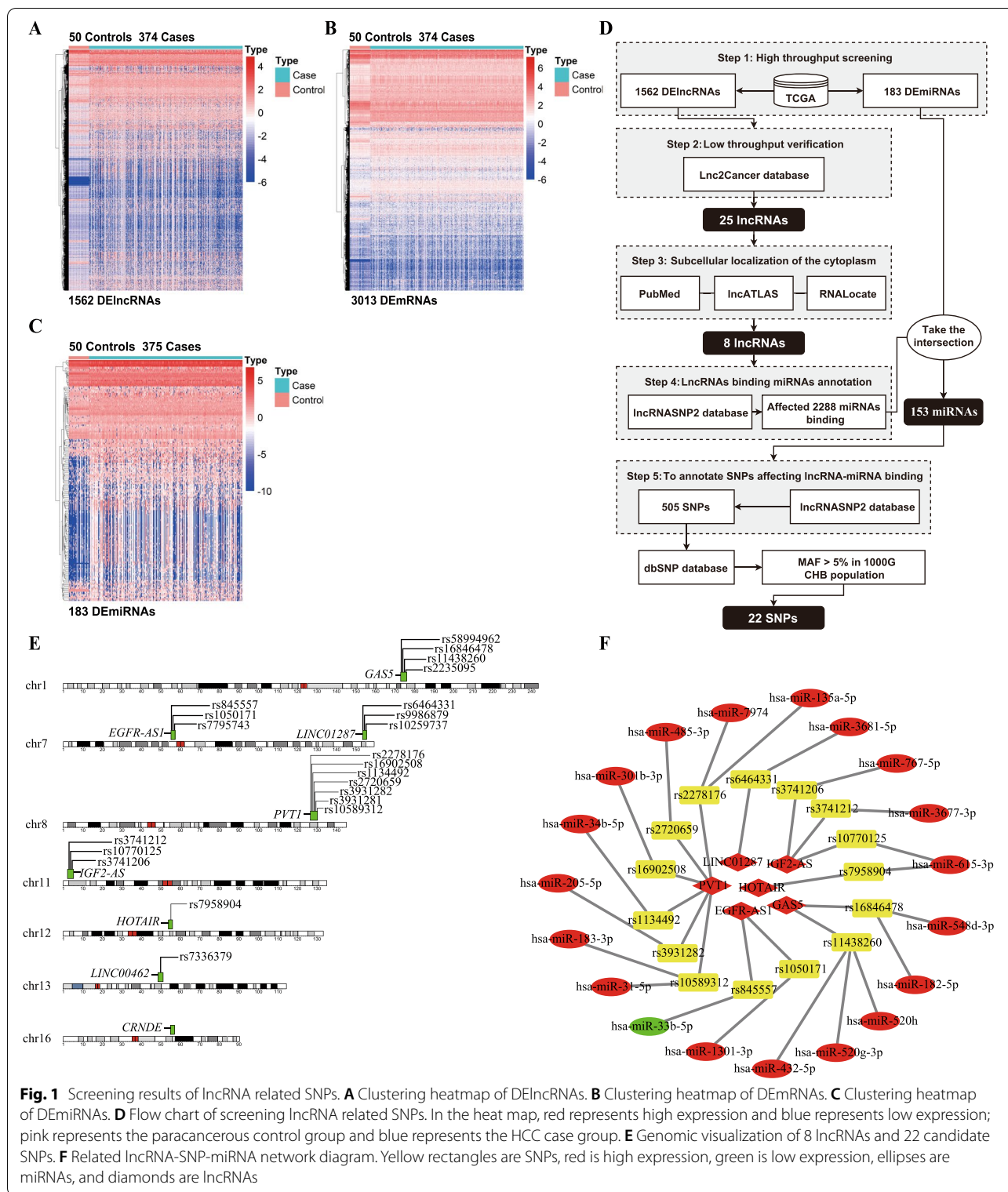
The gene expression heatmaps of differentially expressed lncRNAs (DElncRNAs), differentially expressed mRNAs (DEmRNAs) and differentially expressed miRNAs (DEmiRNAs) were shown in Fig. 1A, Fig. 1B and Fig. 1C, respectively. Based on differential analysis, we obtained 1562 DElncRNAs (1415 up-regulated and 147 down-regulated), 3013 DEmRNAs (2533 up-regulated and 480 down-regulated) and 183 DEmiRNAs (175 up-regulated and 8 down-regulated).

The screening of 22 candidate SNPs on HCC related ceRNAs

As shown in the flow chart (Fig. 1D), 22 candidate SNPs on eight HCC related lncRNA mediated ceRNAs were screened out following steps below. Step1 and step 2: 1562 DElncRNAs screened by high throughput were validated by low throughput experiments in HCC samples using lnc2Cancer database (<http://bio-bigdata.hrbmu.edu.cn/lnc2cancer/>), 25 of which had consistent expression trends and passed the validation. Step 3: Considering that the ceRNA mechanism of lncRNAs mainly occurred in the cytoplasm, based on PubMed database and web tools RNALocate (<https://www.rna-society.org/rnalocate/>) and lncATLAS (<https://lncatlas.org/eu/>), eight of 25 lncRNAs were retrieved with evidence of subcellular localization in the cytoplasm. Step 4 and step 5: The lncRNASNP2 database was used to annotate eight lncRNAs, and 2699 SNPs affecting the binding of lncRNA-miRNA were extracted. These 2699 SNPs affected the binding of 2288 miRNAs, but only 153 miRNAs were retained through DEmiRNAs filtering, involving 505 SNPs in total. Finally, based on the dbSNP database and the standard of $MAF > 0.05$ in the Han Chinese in Beijing (CHB) population of the 1000 Genomes Project, 22 of 505 SNPs were screened out. Genomic visualization of eight lncRNAs and 22 candidate SNPs were shown in Fig. 1E.

Identification of 15 SNPs for genotyping

As described in the Method section, the feasibility of genotyping for 22 candidate SNPs was evaluated. In the 48-Plex SNPscan™ typing system, 22 SNPs were classified into four levels: first, second, third and fail (Additional



file 1: Supplementary Fig. 1A). Subsequent genotyping was abandoned for unqualified grade SNPs (rs7336379, rs9986879 and rs10259737), which were unable to

design primers for genotyping. Haploview 4.2 software (<https://sourceforge.net/projects/haploview/>) was used to determine the linkage disequilibrium (LD) by the

standardized D' and r^2 values for the other 19 SNPs. Reference population is CHB of the 1000 Genomes Project. The results showed strong LD ($r^2 > 0.8$) between some SNPs (Additional file 1: Supplementary Fig. 1B). Specifically, rs1050171 on *EGFR-AS1* can represent rs7795743, rs16846478 on *GASS* can represent rs58994962 and rs2235095, and rs3931282 on *PVT1* can represent rs3931281 for subsequent genotyping. Finally, a total of 15 SNPs were genotyped, and their network diagram with corresponding lncRNAs and miRNAs was shown in Fig. 1F.

Characteristics of the study population

The demographic characteristics of 1601 subjects, clinical information and clinical test indexes of 800 HCC patients enrolled in our study are shown in Table 1. No significant differences were observed for age, gender and cigarette smoking between cases and controls (all $P > 0.05$). The P-P plot (Additional file 1: Supplementary Fig. 2) and Levene's test (Additional file 2: Supplementary Table 1) respectively show that the age conforms to the normal distribution and has homogeneity of variance. However, alcohol drinking was significantly different between the two groups ($P = 0.010$). Whether the difference was significant or not, these demographic characteristics were included in the statistical model as potential confounding factors for genetic association analysis.

The association of 14 candidate SNPs in lncRNA-mediated ceRNAs with HCC risk

The genotype distribution of 15 SNPs is shown in Fig. 2A. Rs2278176 was excluded in subsequent analysis because it did not conform to Hardy–Weinberg equilibrium (HWE). Further analysis was conducted on the effect of the genotypes of 14 SNPs under different genetic models (Fig. 2 and Additional file 3: Supplementary Table 2). As shown in Fig. 2A, compared with the CC genotype, the distribution of GG, GC and GG+GC genotypes of rs7958904 was significantly different between the case and control group (GG vs CC, $P = 0.028$, codominant model 1; GC vs CC, $P = 0.008$, codominant model 2; GG+GC vs CC, $P = 0.014$, dominant model). Then binary logistic regression was used to explore the association between rs7958904 and HCC risk. Age, gender, cigarette smoking and alcohol drinking were included as covariates to adjust for confounders. As shown in Fig. 2B, compared with CC genotype, a protective effect of rs7958904 GG, GC and GG+GC genotypes was found for HCC risk ($P = 0.042$, OR = 0.65, 95% CI 0.43–0.98; $P = 0.013$, OR = 0.59, 95% CI 0.38–0.89; and $P = 0.023$, OR = 0.63, 95% CI 0.42–0.94, respectively). Furthermore, the association still stood after an FDR correction (corrected $P = 0.042$; $P = 0.035$ and $P = 0.035$, respectively).

Table 1 Characteristics of the study population

Variable	Cases (N = 800)	Controls (N = 801)	P value
Age (years)	Mean \pm S.D 56.87 \pm 10.25	Mean \pm S.D 56.97 \pm 10.26	0.851 ^a
Gender			
Male	652 (81.5%)	688 (86.0%)	
Female	148 (18.5%)	133 (16.6%)	0.319 ^b
Cigarette smoking			
No	415 (51.9%)	430 (53.8%)	0.469 ^b
Yes	385 (48.1%)	371 (46.4%)	
Alcohol consumption			
No	463 (57.9%)	514 (64.3%)	0.010^b
Yes	337 (42.1%)	287 (35.9%)	
Family history			
No	757 (94.6%)	-	-
Yes	43 (5.4%)	-	-
HBsAg			
Negative	171 (21.4%)	-	-
Yes	629 (78.6%)	-	-
Anti-HCV			
Negative	697 (87.1%)	-	-
Yes	103 (12.9%)	-	-
Liver cirrhosis			
No	230 (28.8%)	-	-
Yes	570 (71.3%)	-	-
Child–Pugh grade			
A	696 (87.1%)	-	-
B or C	104 (12.9%)	-	-
Tumor size			
< 5 cm	418 (52.3%)	-	-
\geq 5 cm	382 (47.8%)	-	-
Tumor number			
Solitary	400 (50.0%)	-	-
Multiple	400 (50.0%)	-	-
Vascular invasion			
No	541 (67.6%)	-	-
Yes	259 (32.4%)	-	-
Lymphatic metastasis			
No	614 (76.8%)	-	-
Yes	186 (23.3%)	-	-
Distant metastasis			
No	631 (78.9%)	-	-
Yes	169 (21.1%)	-	-
TNM staging			
I + II	275 (34.4%)	-	-
III + IV	525 (65.6%)	-	-

The bold and italicized value indicates statistical significance

^a Student's *t*-test

^b Chi-square test. S.D. standard deviation

For other SNPs, no association with HCC risk was observed in any genetic model (Fig. 2 and Additional file 3: Supplementary Table 2).

Association analysis between 14 SNPs and clinical characteristics of HCC patients

The association between 14 SNPs and clinical characteristics of HCC patients, including liver cirrhosis, Child–Pugh grade, tumor size, tumor number, vascular invasion, lymphatic metastasis, distant metastasis and TNM stage, was analyzed by binary logistic regression. And age, gender, cigarette smoking, alcohol drinking, family history, HBsAg and anti-HCV were used as covariates to adjust for confounding factors. FDR correction was performed for all *P* values. The HCC patients were staged according to the AJCC-TNM classification [21]. Early-stage patients included patients with stage I and II. Advanced stage patients included patients with stage III and IV. As shown in Fig. 3A, the distribution of rs3931282 genotypes showed a significant difference between early stage and advanced stage. In detail, compared with GG genotype, the frequency of AA + GA genotype was higher in early stage than that in advanced stage and showed a decreased risk for HCC progression (FDR = 0.042, OR = 0.576, 95% CI 0.402–0.825). In addition, the results of associations of other SNPs with clinical characteristics are shown in Additional file 4: Supplementary Table 3.

Association analysis between 14 SNPs and clinical test indexes of HCC patients

In this study, the levels of α -fetoprotein (AFP), AST, ALT and AST/ALT ratio in HCC patients did not conform to the normal distribution. Mann–Whitney U test was used to explore the correlation between the distribution of 14 SNP genotypes and these indicators. FDR correction was performed for all *P* values at the same time. As shown in Fig. 3B, the AST and ALT levels in patients with rs1134492 CC + TC genotype were significantly higher than those with TT genotype (corrected *P* = 0.042 and 0.014, respectively). Meanwhile, AST/ALT ratios were significantly higher in patients with rs10589312 TCTTGC/TCTTGC + TCTTGC/T genotype than those with TT genotype (corrected *P* = 0.021), and AST/ALT ratios were significantly lower in patients with rs84557 CC + CT genotype than

those with TT genotype (corrected *P* = 0.006). In addition, the results of associations of other SNPs with clinical test indexes are shown in Additional file 1: Supplementary Fig. 3.

Acquisition of mRNAs targeted by miRNAs

In this association study, five SNPs (rs7958904, rs3931282, rs1134492, rs10589312 and rs84557) were found to be associated with the occurrence and prognosis of HCC. These five loci located on lncRNA coding genes would affect the binding of six microRNAs (miR-615-3p, miR-205-5p, miR-34b-5p, miR-183-3p, miR-31-5p and miR-33b-5p) to lncRNAs (Fig. 1E). The target genes of these six miRNAs were obtained through screening mRNAs shared by miR-Walk and miRTarBase databases. For miR-615-3p, highly expressed DE mRNAs in cancers were also compared with candidate target mRNAs to obtain cross-genes between them. Finally, 57 target genes of miR-615-3p, 163 target genes of miR-205-5p, 92 target genes of miR-34b-5p, 108 target genes of miR-183-3p, 171 target genes of miR-31-5p and 87 target genes of miR-33b-5p were extracted (Fig. 4A, Fig. 4D, Fig. 4G, Fig. 4J, Fig. 4M and Fig. 4P, respectively).

The modular analysis of PPI network of miRNA-target mRNAs

Firstly, PPI networks of six miRNAs targets were constructed based on String, and the Cytoscape plug-in MCODE was used for module analysis. Then, the PPI networks of miR-615-3p, miR-205-5p, miR-34b-5p, miR-183-3p, miR-31-5p and miR-33b-5p target genes were constructed into two, five, three, three, six and three functional modules, respectively. (Fig. 4B, Fig. 4E, Fig. 4H, Fig. 4K, Fig. 4N and Fig. 4Q, respectively). Furthermore, KEGG or Go-BP enrichment of the most important modules was performed using Cytoscape plug-ins ClueGO and CluePedia, which was the first cluster in the PPI network module analysis of six miRNAs target genes. The results of KEGG or Go-BP enrichment analysis were shown in Fig. 4C, Fig. 4F, Fig. 4I, Fig. 4L, Fig. 4O and Fig. 4R. Therefore, the targets of the above miRNAs may be involved in the occurrence and development of HCC by participating in pathways or biological processes shown in the figure.

Discussion

In this study, eight lncRNA-mediated ceRNAs associated with HCC were screened by bioinformatics methods. Compared with recent researches on screening HCC related

(See figure on next page.)

Fig. 2 The genotype distribution and association analysis of 15 lncRNA-SNPs with HCC risk. **A** The genotype distribution, HWE test, and association analysis under the codominant model 1, codominant model 2 and dominant model. 1 is the TCTTGC/TCTTGC genotype of rs10589312, 2 is the TCTTGC/T genotype of rs10589312, 1 + 2 is the TCTTGC/TCTTGC + TCTTGC/T genotype of rs10589312, and 3 is the T/T genotype of rs10589312. The values in red italics are statistically significant. **B** Association analysis of SNP rs7958904 with HCC risk after adjusted. OR and *P* values were adjusted for age, gender, smoking and drinking by logistic regression. The values in red italics are statistically significant

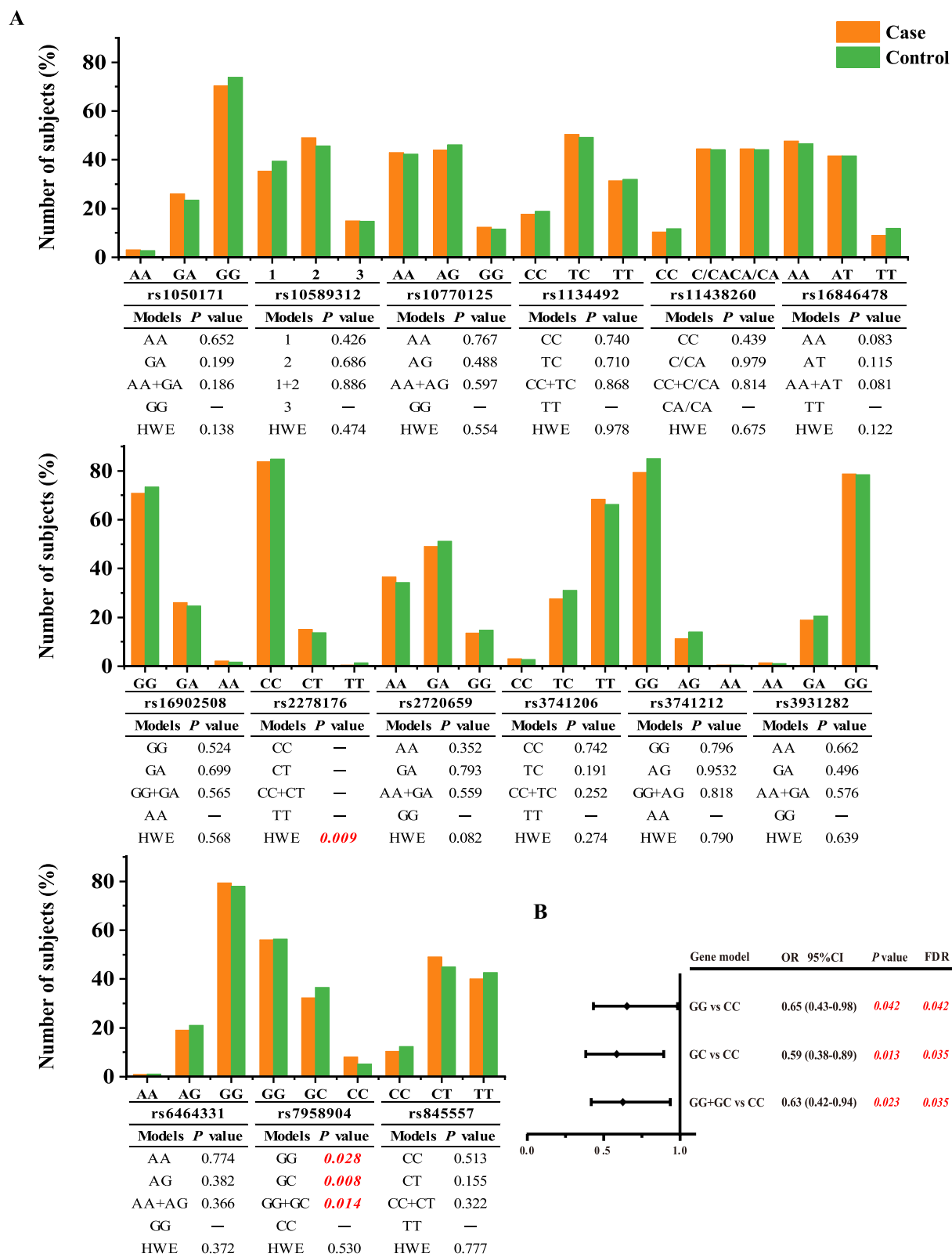
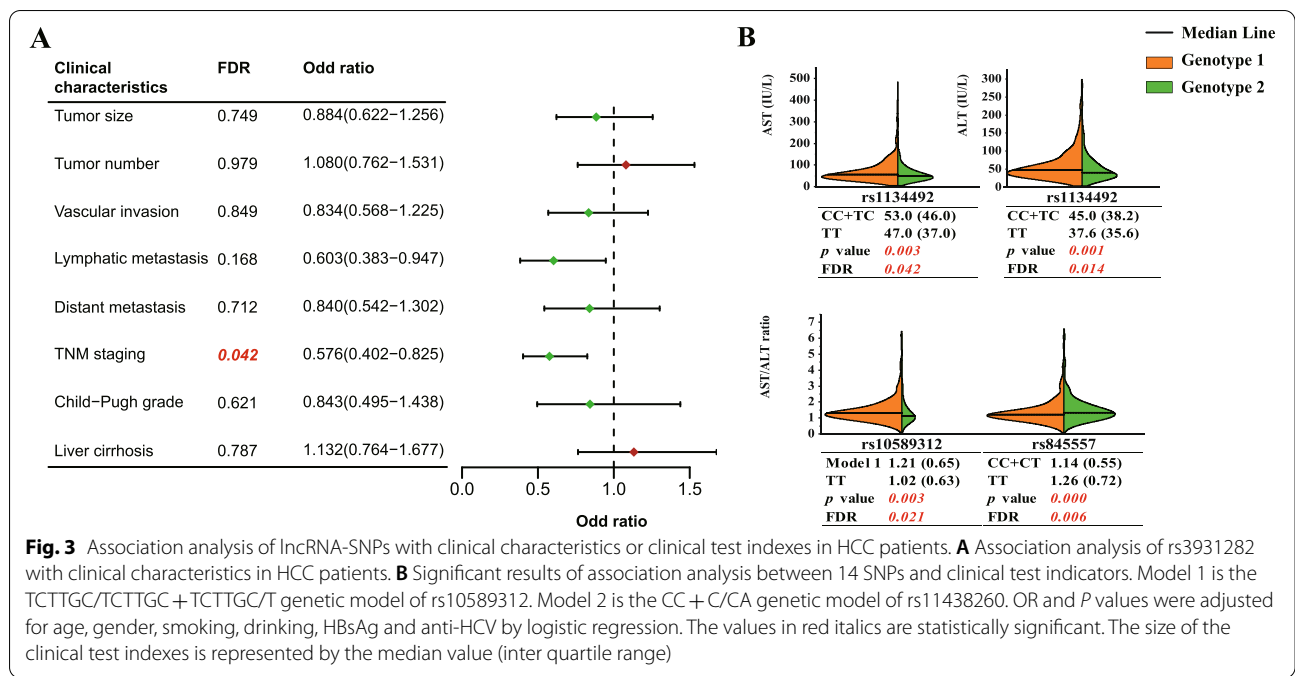


Fig. 2 (See legend on previous page.)



ceRNA networks based on TCGA, such as the study of Ye et al. [22], the screening process of this study has significant characteristics on the basis of the same scientific nature. We not only used the Lnc2Cancer database to extract the lncRNAs supported by the low throughput experimental verification, but also considered the subcellular localization of lncRNAs. The latter is due to the fact that the biological process of lncRNA competing with mRNA to bind miRNA mainly occurs in the cytoplasm [23]. Then eight lncRNAs were annotated and the functional SNPs that affect the binding of miRNAs to lncRNAs were extracted. After evaluating these SNPs, 15 SNPs were finally selected and a case-control study was conducted in a Han Chinese population to explore the potential associations between these candidate SNPs and the risk of HCC.

In the association analysis, five positive SNPs that were significantly associated with susceptibility or prognosis of HCC were identified. To our knowledge, this study first reported the relationship between rs3931282, rs1134492, rs10589312 and rs84557 and the prognosis of HCC. In HCC cohort, the TNM stage of AA + GA genotype of *PVT1* rs3931282 was earlier than that of GG genotype, suggesting that the former had a better prognosis. In

addition, our results showed that *PVT1* rs1134492 genotypes were associated with AST and ALT levels, while *PVT1* rs10589312 and *EGFR-AS1* rs84557 were associated with AST/ALT ratios of HCC patients. Clinical biochemical indicators often reflect the liver function of patients. For example, the higher the AST and ALT level and AST/ALT ratio, the more serious the hepatocyte injury, the worse the liver function and the worse the prognosis. These results indicated that some promising SNPs of ceRNAs could be used as important cancer biomarkers and contribute to the individualized prognosis of HCC in a specific population.

This study is the first to show that *HOTAIR* rs7958904 was associated with HCC susceptibility, although it has been reported to be associated with the risk of other cancers. In a case-control study of South Korean population [24], individuals with rs7958904 GG genotype were found to have a lower risk of developing colorectal cancer compared to those with CC + GC and GC genotypes. This trend of increased cancer risk of allele C is consistent with that observed in our study. The present study showed that SNP rs7958904 was strongly associated with the HCC risk. Compared with CC genotypes, individuals

(See figure on next page.)

Fig. 4 Bioinformatics analysis of positive SNPs-related miRNAs. **A-C** Bioinformatics analysis of *HOTAIR*-rs7958904-miR-615-3p. **D-F** Bioinformatics analysis of *PVT1*-rs3931282-miR-205-5p. **G-I** Bioinformatics analysis of *PVT1*-rs1134492-miR-34b-5p. **J-L** Bioinformatics analysis of *PVT1*-rs10589312-miR-183-3p. **M-O** Bioinformatics analysis of *PVT1*-rs10589312-miR-31-5p. **P-R** Bioinformatics analysis of *EGFR-AS1*-rs84557-miR-33b-5p. Each node in enrichment analysis of functional modular genes represents a term, the connection between the node and gene reflects the existence of correlation, and the color reflects the enrichment classification of node and gene

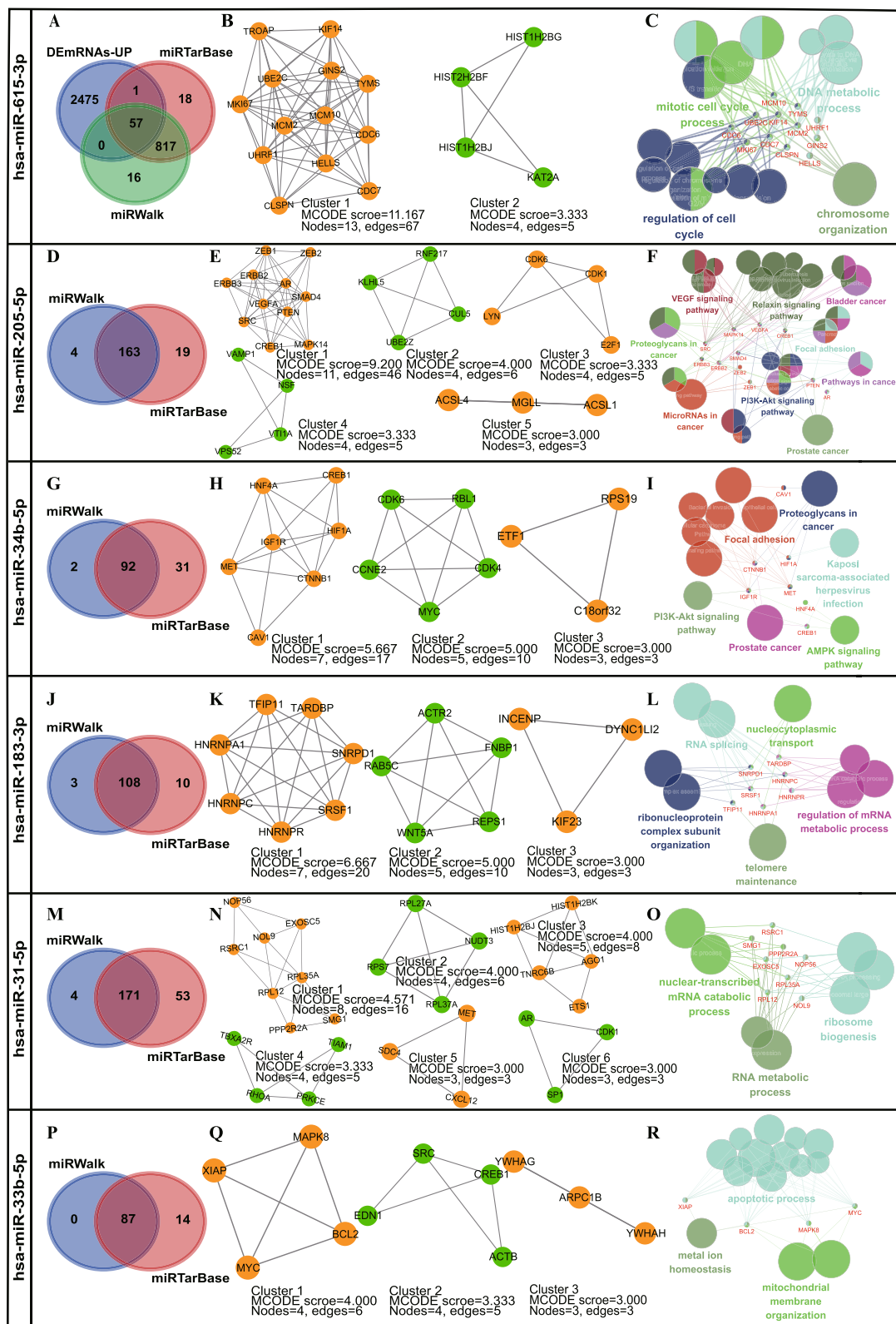
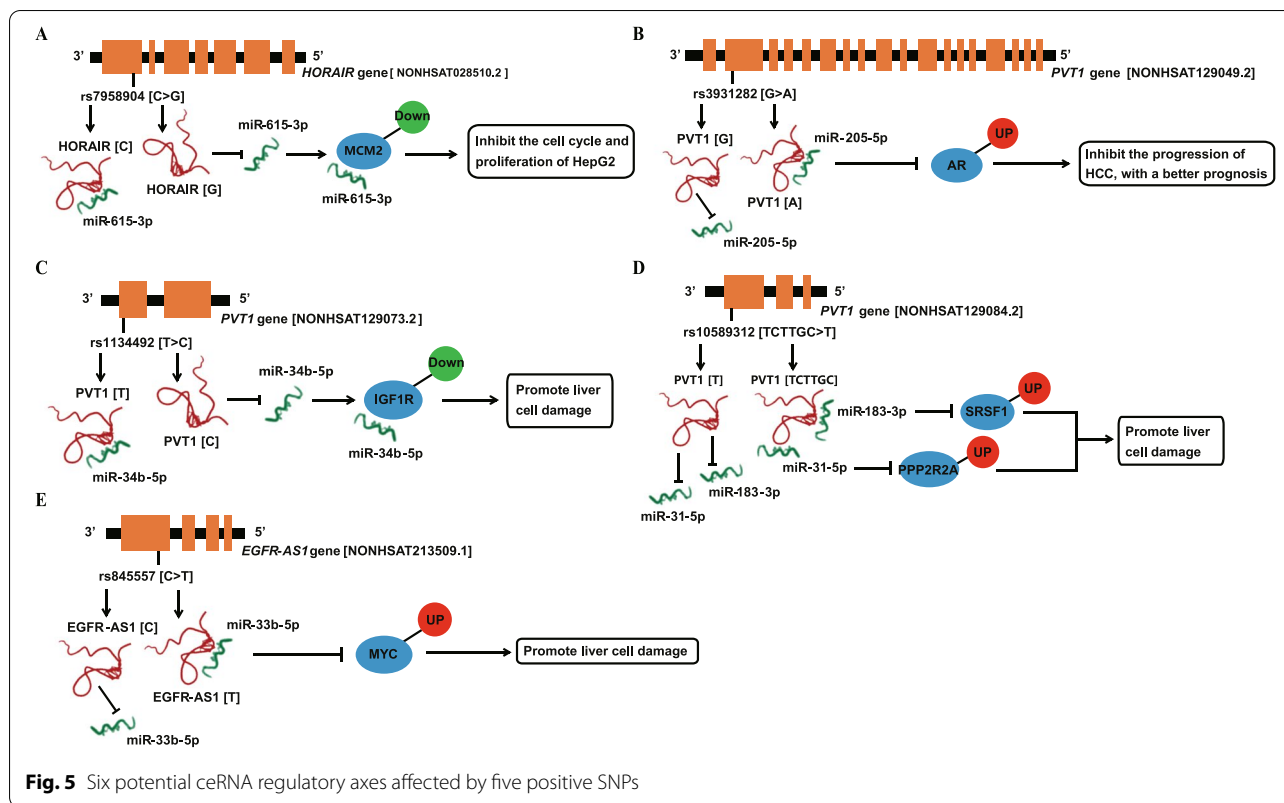


Fig. 4 (See legend on previous page.)



with rs7958904 GG or GC genotypes showed 0.65-fold or 0.59-fold decreased HCC risk. Therefore, rs7958904 may serve as a promising predictor for HCC risk. However, in a study linking rs7958904 with ovarian cancer in a southern Chinese population, the CC genotype showed a protective effect [25]. One of the reasons for this contradiction may be that the regulatory axis of ceRNA affected by rs7958904 is different in different types of cancers. Therefore, in order to further explain how the five positive SNPs are related to the risk and prognosis of HCC through ceRNA mechanism, the potential regulatory axis of HCC related ceRNA affected by positive SNPs was constructed through miRNA target gene prediction and functional module analysis.

The first regulation axis of HCC related ceRNA was affected by the positive SNP rs7958904. As shown in Fig. 5A, according to the annotation of lncRNASNP2 database, the rs7958904 G allele of *HOTAIR* will disrupt the adsorption of *HOTAIR* to miR-615-3p, thereby increasing the silencing effect of miR-615-3p on its target genes. Among the targets of miR-615-3p, inhibition of MCM2 expression could significantly inhibit HepG2 proliferation and cell cycle through the cyclin D-dependent kinases (CDKs) 2/7 pathway [26]. The regulatory axis expounds the possible reason why individuals with *HOTAIR* rs7958904 GG and GC genotypes show a lower

risk of developing HCC than those with CC genotype. The second regulation axis of HCC related ceRNA was affected by the positive SNP rs3931282. As shown in Fig. 5B, the rs3931282 A allele of *PVT1* will enhance the adsorption of *PVT1* to miR-205-5p, thereby reducing the silencing effect of miR-205-5p on its target. Among the targets of miR-205-5p, the increased expression of AR gene can increase HCC cell adhesion and inhibit HCC cell migration by activating AR-β1-integrin-AKT signal transduction [27, 28]. This is the possible reason why HCC patients with rs3931282 AA+GA genotype have earlier TNM staging and better prognosis than those with GG genotype.

The third, fourth and fifth regulation axes of HCC related ceRNA were affected by the positive SNP rs1134492, rs10589312 and rs84557, respectively. Specifically, in the third regulatory axis, the rs1134492 C allele of *PVT1* will destroy the adsorption of *PVT1* to miR-34b-5p, thus increasing the silencing effect of miR-34b-5p on its target. Among the targets of miR-34b-5p we excavated, the IGF1R inhibition can promote Fas-induced liver injury [29] (Fig. 5C). In the fourth regulatory axis, the rs10589312 TCTTGC allele of *PVT1* will enhance the adsorption of *PVT1* to miR-183-3p and miR-31-5p, and thus weaken the inhibition of miR-183-3p and miR-31-5p on their targets SRSF1

and PPP2R2A. Among them, SRSF1 overexpression can promote liver injury induced by caspase-dependent apoptosis pathway [30, 31]. Overexpression of PPP2R2A can inhibit Akt phosphorylation, thus blocking the PI3K/Akt signaling pathway and causing liver injury [32, 33] (Fig. 5D). In the fifth regulatory axis, the rs84557 T allele of *EGFR-AS1* will enhance the adsorption of *EGFR-AS1* to miR-33b-5p, thereby reducing the inhibition of miR-33b-5p on its target MYC. MYC overexpression can activate the p14ARF/MDM2 pathway, thereby stimulating p53-mediated apoptosis and promoting liver injury [34] (Fig. 5E). The influence of the above SNPs on the possible regulatory axis may be the molecular mechanism behind the positive genetic association results in this study. Specifically, HCC patients with rs1134492 CC + TC genotype had higher AST and ALT levels and worse prognosis compared to those with TT genotype, HCC patients with rs10589312 TCTTGC/TCTTGC + TCTTGC/T genotype had higher AST/ALT ratios than those with TT genotype, and patients with rs84557 TT genotype had higher AST/ALT ratios than those with CC + CT genotype.

In recent years, three lncRNAs located by the five positive association SNPs found in this study, namely *HOTAIR*, *PVT1* and *EGFR-AS1*, have been reported to be involved in the occurrence and development of a variety of cancers through the ceRNA mechanism. For example, it has been found that *HOTAIR* acts as an endogenous "sponge" for miR-148b to regulate the expression of DNMT1/MEG3/p53 pathway in hepatic stellate cells, which is related to the occurrence of HCC [35]. Another study showed that lncRNA *PVT1*, as a ceRNA, could compete with Atg3 to bind to microRNA-365 and promote autophagy in HCC cells [36]. Recently, lncRNA *EGFR-AS1* was found to be associated with migration, invasion and apoptosis of glioma cells by targeting miR-133b/RACK1 [37].

Several limitations of this study should be considered. First, we recommend that the findings of this association study should be expanded to other ethnic groups in different regions in the world, beyond the northeastern Chinese Han population. Second, the five potential regulatory axes of HCC related ceRNA affected by positive SNPs need to be verified by further in vitro and in vivo experiments.

Conclusions

This is the first study to show that SNPs on lncRNA-mediated ceRNAs are associated with the HCC risk and prognosis. *HOTAIR* rs7958904, *PVT1* rs3931282, rs1134492 and rs10589312, and *EGFR-AS1* rs84557 might be involved in the occurrence or prognosis of

HCC. The reason why these positive SNPs were associated with HCC are also explained from the perspective of the regulatory axis of ceRNA.

Methods

Data acquisition and identification of differentially expressed genes

In this study, liver hepatocellular carcinoma (LIHC) RNA-seq data (lncRNA and mRNA, level 3; including 374 HCC samples and 50 normal samples, Illumina HiSeq RNA-Seq platform) and miRNA-seq data (including 375 HCC samples and 50 normal samples, Illumina HiSeq miRNA-Seq platform) were downloaded from TCGA (<https://tcga-data.nci.nih.gov/>) in September 2019. The lncRNA and mRNA gene symbol was annotated using the Ensembl database (<http://www.ensembl.org/>). First, prior to differential expression analysis, all unexpressed RNAs were excluded by removing all lines in the gene expression matrix with a mean value of less than or equal to 1. Then the DElncRNAs, mRNAs DEMRNAs and miRNAs DEmiRNAs between the HCC group and the normal group were identified by edgeR package [38], and multiple test correction was performed using Benjamini and Hochberg false discovery rate (FDR) [39]. The differential expression threshold was $FDR < 0.5$ and $|\log_2FC|$ (fold change) > 1.5 . In this study, the heatmap package was used to draw the hierarchical clustering heatmap of DElncRNAs, DEmiRNAs and DEMRNAs data in R software (version 3.5.3).

Study population

In this study, 800 HCC patients were recruited from the inpatient department in the Harbin Medical University Cancer Hospital between January 2007 and December 2016. At the same time, 801 healthy controls were collected from the Physical Examination Center of the First Affiliated Hospital of Harbin Medical University. The healthy control group was frequency-matched with the HCC case group according to gender and age. In addition, all subjects must be stable residents of the area. "Stable residents" means that all subjects are Han Chinese from northeast China who have lived in Harbin for at least three generations.

The diagnosis of HCC was based on histological, combined with at least one positive HCC image on computed tomography (CT) or magnetic resonance imaging (MRI), sometimes combined with serum AFP analysis (> 400 ng/ml). And none of the patients received any chemotherapy or radiation therapy before sampling. Moreover, patients with positive laboratory tests for HIV, or suspected autoimmune diseases with antinuclear antibody titer greater than 1:160 were excluded from this study. Each healthy

control individual underwent the examination of antigen and antibody, at least one typical morphologic finding from CT or ultrasound. They were with no primary liver cancer, hepatitis, liver cirrhosis, or hepatic distomiasis.

Sample collection information

Demographic data on participants were collected through in-person interviews, including age, gender, smoking and drinking status. Clinical characteristics of HCC patients were collected from patient medical records, including HBsAg status, anti-HCV, liver cirrhosis, Child–Pugh grade, tumor size, tumor number, vascular invasion, lymphatic metastasis, distant metastasis and TNM stage. Clinical test indexes such as AFP, AST, ALT and AST/ALT ratio in HCC patients were all obtained from the hospital patient information management system. Approximately 2–3 ml of venous blood samples were collected from each subject. Peripheral blood DNA was extracted by the QIAamp DNA Blood Kit (Valencia, CA, USA) and stored at -80 °C.

Candidate SNP assessment and genotyping

Prior to genotyping, 22 candidate SNPs were evaluated to identify those that could not be successfully genotyped in the genotyping system and those with high linkage. For the former, the subsequent genotyping was discarded. For the latter, one SNP was selected as a representative from the highly linked SNPs. Finally, 15 SNPs were genotyped using a custom designed SNPscan™ kit based on the high-throughput SNP genotyping method using double ligation and multiplex fluorescent PCR. In a 200 µL PCR tube, LDR was performed in a final volume of 10 µL containing 3 µL of PCR products, 1 µL 10 × Taq DNA ligase buffer, 0.01 µL of each probe, and 5 U of Taq DNA ligase (NEB, USA). The reaction mixtures were heated for 2 min at 95 °C, followed by 30 thermal cycles at 94 °C for 30 s (denaturation) and 56 °C for 3 min (annealing and ligation). Prior to polyacrylamide gel electrophoresis, the 1 µL LDR products were mixed with 8 µL HiDi (highly deionized-formamide) and denatured for 3 min at 95 °C. The LDR products were separated by electrophoresis using the ABI 3730XL Gene sequencer. All PCR primers and gene probes used in PCR and LDR were designed using Primer 5.0 and Oligo 6.0 software. In terms of quality control, genotyping was performed twice in 5% of case and control random samples (a total of 80 samples) to verify the accuracy of genotyping, and the repeatability of genotyping results was 100%.

Statistical analysis

The Pearson's χ^2 test was used to evaluate HWE for controls. Student's *t* test (for continuous variables) and χ^2 test (for categorical variables) were conducted to compare

the demographic characteristics between HCC cases and healthy controls. Among them, P-P plots and Levene's test were used to assess the normal distribution and homogeneity of variance of continuous variables, respectively. The adjusted odds ratios (ORs) with their 95% confidence intervals (CIs) for the association between genotype frequencies and the risk of HCC plus clinicopathological characteristics (in HCC patients) were evaluated by multiple logistic regression models after controlling for other covariates. Meanwhile, the analysis of polymorphisms and biochemical indicators in HCC cases was performed by Mann–Whitney U test. Statistical analysis was conducted using SPSS version 25.0 software (SPSS, Chicago, IL, U.S.A.). All tests were two-sided and *P*-values less than 0.05 were considered statistically significant. Moreover, the method of FDR was used for multiple test correction.

Acquisition of miRNA target genes and module analysis

In order to identify the effect of statistically significant positive SNPs in genetic association analysis on the potential ceRNA regulatory axis of HCC, target mRNAs of related miRNAs were extracted from miRWalk [40] and miRTarBase [41] databases. In order to improve the reliability of the results, only miRNA–mRNA relationship pairs supported by both databases were extracted. Then the online STRING database (<https://string-db.org/>) was used to construct the protein–protein interaction (PPI) networks for these targets. Interactions with a binding score > 0.4 were considered statistically significant.

The molecular interaction network was visualized using Cytoscape software (version 3.5.1), an open-source bioinformatics software platform. Cytoscape's Molecular Complex Detection (MCODE) plug-in (version 1.4.2) [42] was used to identify functional modules through cluster analysis to obtain the PPI network through topological network analysis. Thus, a protein complex with biological significance or functional module was obtained. The parameters are set as follows: Include Loops = false, Degree Cutoff = 2, Node Score Cutoff = 0.2, Haircut = true, Fluff = false, K Core = 2, Max. Depth from Seed = 100.

KEGG and GO enrichment analyses of module genes

Kyoto Encyclopedia of Gene and Genome (KEGG) [43] and Gene Ontology (GO) [44] were mainly used to annotate and analyze biological processes (BP) of genes. In this study, ClueGO (version 2.5.7) [45] and CluePedia (version 1.5.7) [46] were selected to perform KEGG or GO–BP enrichment analysis on the most important modules in the module analysis. The parameters were set as follows: Show only Pathways with $pV \leq 0.01$, GO Tree Interval Min Level = 4, GO Tree Interval Max Level = 7, Kappa Score = 0.6, Layout = yFiles Organic Layout.

Abbreviations

HBV: Hepatitis B virus; HBsAg: Hepatitis B surface antigen; HCV: Hepatitis C virus; HCC: Hepatocellular carcinoma; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; AFP: α -Fetoprotein; BP: Biological process; eRNA: Competing endogenous RNA; FDR: False discovery rate; GWAS: Genome-Wide Association Study; HWE: Hardy-Weinberg equilibrium; lncRNA: Long non-coding RNA; MAF: Minor allele frequency; miRNA: MicroRNA; OR: Odds ratio; PPI: Protein-protein interaction; SNP: Single nucleotide polymorphism.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-09010-9>.

Additional file 1: Supplementary Figure 1. Genotyping feasibility assessment of candidate SNPs. Priority "Fail" was not genotyped, and one of the linkage SNPs (red) was selected for genotyping. **Supplementary Figure 2.** P-P plot shows that the age is normally distributed. The data points all fall on the oblique line of 45°, showing strong normality. **Supplementary Figure 3.** Association analysis of 14 SNPs with clinical test indexes in HCC patients. (A) Association analysis of 14 SNPs with AFP in HCC patients. (B) Association analysis of 14 SNPs with AST in HCC patients. (C) Association analysis of 14 SNPs with ALT in HCC patients. (D) Association analysis of 14 SNPs with AST/ALT ratios in HCC patients. Model 1 is the TCTTGC/TCTTGC+TCTTGC/T genetic model of rs10589312. Model 2 is the CC+C/CA genetic model of rs11438260. The values in red italics are statistically significant. The size of the indicator is represented by the median value (inter quartile range).

Additional file 2: Supplementary Table 1. Independent sample test of age.

Additional file 3:

Additional file 4:

Acknowledgements

We gratefully appreciate the numerous sample donors for making this work possible.

Authors' contributions

Xuelong Zhang, Songbin Fu and Han Mo conceived and designed the study; Han Mo, Xi Wang, Guohua Ji, Xiao Liang, Yi Yang, Yuandong Qiao, Xueyuan Jia and Henan Zhou performed the experiments; Xuelong Zhang, Han Mo, Lidan Xu and Xueyuan Jia analyzed the data; and Xuelong Zhang, Wenjing Sun, Wenhui Zhao and Han Mo prepared the manuscript. All authors revised and approved the final draft.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81373220 and No.81302062).

Availability of data and materials

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the Ethics Committee of the Harbin Medical University and all experimental procedures complied with the Declaration of Helsinki. All participants gave written informed consent to take part in the present study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratory of Medical Genetics, Harbin Medical University, Harbin 150081, China. ²Key Laboratory of Preservation of Human Genetic Resources and Disease Control in China (Harbin Medical University), Ministry of Education, Harbin 150081, China. ³Department of Gastroenterology, the First Affiliated Hospital, Harbin Medical University, Harbin 150001, China. ⁴Department of Clinical Laboratory, Harbin Medical University Cancer Hospital, Harbin 150081, China. ⁵Department of Internal Medicine, Harbin Medical University Cancer Hospital, Harbin 150081, China.

Received: 16 July 2022 Accepted: 14 November 2022

Published online: 24 November 2022

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209–49.
- EI-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*. 2007;132(7):2557–76.
- EI-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology*. 2012;142(6):1264–1273 e1261.
- Aravalli RN, Steer CJ, Cressman EN. Molecular mechanisms of hepatocellular carcinoma. *Hepatology*. 2008;48(6):2047–63.
- Yang J, Trepo E, Nahon P, Cao Q, Moreno C, Letouze E, Imbeaud S, Gustot T, Deviere J, Debette S, et al. PNPLA3 and TM6SF2 variants as risk factors of hepatocellular carcinoma across various etiologies and severity of underlying liver diseases. *Int J Cancer*. 2019;144(3):533–44.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106(23):9362–7.
- Doane AS, Elemento O. Alterations in transcriptional networks in cancer: the role of noncoding somatic driver mutations. *Curr Opin Genet Dev*. 2022;75: 101919.
- Downes DJ, Hughes JR. Natural and experimental rewiring of gene regulatory regions. *Annu Rev Genomics Hum Genet*. 2022;23:73–97.
- Zheng J, Huang X, Tan W, Yu D, Du Z, Chang J, Wei L, Han Y, Wang C, Che X, et al. Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat Genet*. 2016;48(7):747–57.
- Wang H, Meng Q, Qian J, Li M, Gu C, Yang Y. Review: RNA-based diagnostic markers discovery and therapeutic targets development in cancer. *Pharmacol Ther*. 2022;234: 108123.
- Huang Z, Zhou JK, Peng Y, He W, Huang C. The role of long noncoding RNAs in hepatocellular carcinoma. *Mol Cancer*. 2020;19(1):77.
- Mirzaei S, Gholami MH, Hushmandi K, Hashemi F, Zabolian A, Canadas I, Zarrabi A, Nabavi N, Aref AR, Crea F, et al. The long and short non-coding RNAs modulating EZH2 signaling in cancer. *J Hematol Oncol*. 2022;15(1):18.
- He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*. 2004;5(7):522–31.
- Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358–69.
- Shen J, Liang C, Su X, Wang Q, Ke Y, Fang J, Zhang D, Duan S. Dysfunction and ceRNA network of the tumor suppressor miR-637 in cancer development and prognosis. *Biomark Res*. 2022;10(1):72.
- Tang X, Ren H, Guo M, Qian J, Yang Y, Gu C. Review on circular RNAs and new insights into their roles in cancer. *Comput Struct Biotechnol J*. 2021;19:910–28.
- Long J, Bai Y, Yang X, Lin J, Yang X, Wang D, He L, Zheng Y, Zhao H. Construction and comprehensive analysis of a ceRNA network to reveal potential prognostic biomarkers for hepatocellular carcinoma. *Cancer Cell Int*. 2019;19:90.

18. Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I, Muller RU, Straka E, Su L, Burki EA, et al. A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res.* 2008;68(20):8535–40.
19. Gong J, Tian J, Lou J, Ke J, Li L, Li J, Yang Y, Gong Y, Zhu Y, Zhang Y, et al. A functional polymorphism in Inc-LAMC2-1:1 confers risk of colorectal cancer by affecting miRNA binding. *Carcinogenesis.* 2016;37(5):443–51.
20. Miao YR, Liu W, Zhang Q, Guo AY. IncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic Acids Res.* 2018;46(D1):D276–80.
21. Vauthey JN, Lauwers GY, Esnaola NF, Do KA, Belghiti J, Mirza N, Curley SA, Ellis LM, Regimbeau JM, Rashid A, et al. Simplified staging for hepatocellular carcinoma. *J Clin Oncol.* 2002;20(6):1527–36.
22. Ye D, Liu Y, Chen Y, Li G, Sun B, Peng J, Xu Q. Identification of lncRNA biomarkers in hepatocellular carcinoma by comprehensive analysis of the lncRNA-mediated ceRNA network. *Front Genet.* 2022;13: 832952.
23. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* 2016;17(1):47–62.
24. Kim JO, Jun HH, Kim EJ, Lee JY, Park HS, Ryu CS, Kim S, Oh D, Kim JW, Kim NK. Genetic Variants of HOTAIR Associated With Colorectal Cancer Susceptibility and Mortality. *Front Oncol.* 2020;10:72.
25. Wu H, Shang X, Shi Y, Yang Z, Zhao J, Yang M, Li Y, Xu S. Genetic variants of lncRNA HOTAIR and risk of epithelial ovarian cancer among Chinese women. *Oncotarget.* 2016;7(27):41047–52.
26. Yang J, Xie Q, Zhou H, Chang L, Wei W, Wang Y, Li H, Deng Z, Xiao Y, Wu J, et al. Proteomic Analysis and NIR-II Imaging of MCM2 Protein in Hepatocellular Carcinoma. *J Proteome Res.* 2018;17(7):2428–39.
27. Zhang H, Li XX, Yang Y, Zhang Y, Wang HY, Zheng XFS. Significance and mechanism of androgen receptor overexpression and androgen receptor/mechanistic target of rapamycin cross-talk in hepatocellular carcinoma. *Hepatology.* 2018;67(6):2271–86.
28. Hu C, Fang D, Xu H, Wang Q, Xia H. The androgen receptor expression and association with patient's survival in different cancers. *Genomics.* 2020;112(2):1926–40.
29. Qadir XV, Chen W, Han C, Song K, Zhang J, Wu T. miR-223 Deficiency Protects against Fas-Induced Hepatocyte Apoptosis and Liver Injury through Targeting Insulin-Like Growth Factor 1 Receptor. *Am J Pathol.* 2015;185(12):3141–51.
30. Jin Q, Yan T, Ge X, Sun C, Shi X, Zhai Q. Cytoplasm-localized SIRT1 enhances apoptosis. *J Cell Physiol.* 2007;213(1):88–97.
31. Kedzierska H, Piekliko-Witkowska A. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. *Cancer Lett.* 2017;396:53–65.
32. Yu L, Zheng J, Li J, Wang Y, Lu X, Fan X. Integrating serum exosomal microRNA and liver microRNA profiles disclose the function role of autophagy and mechanisms of Fructus Meliae Toosendan-induced hepatotoxicity in mice. *Biomed Pharmacother.* 2020;123: 109709.
33. Xu J, Lu Y, Liu Q, Xia A, Zhao J, Xu X, Sun Q, Qi F, Sun B. Long noncoding RNA GMAN promotes hepatocellular carcinoma progression by interacting with eIF4B. *Cancer Lett.* 2020;473:1–12.
34. McMahon SB. MYC and the control of apoptosis. *Cold Spring Harb Perspect Med.* 2014;4(7): a014407.
35. Bian EB, Wang YY, Yang Y, Wu BM, Xu T, Meng XM, Huang C, Zhang L, Lv XW, Xiong ZG, et al. Hotair facilitates hepatic stellate cells activation and fibrogenesis in the liver. *Biochim Biophys Acta Mol Basis Dis.* 2017;1863(3):674–86.
36. Yang L, Peng X, Jin H, Liu J. Long non-coding RNA PVT1 promotes autophagy as ceRNA to target ATG3 by sponging microRNA-365 in hepatocellular carcinoma. *Gene.* 2019;697:94–102.
37. Dong ZQ, Guo ZY, Xie J. The lncRNA EGFR-AS1 is linked to migration, invasion and apoptosis in glioma cells by targeting miR-133b/RACK1. *Biomed Pharmacother.* 2019;118: 109292.
38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40.
39. Madar V, Batista S. FastLSU: a more practical approach for the Benjamini-Hochberg FDR controlling procedure for huge-scale testing problems. *Bioinformatics.* 2016;32(11):1716–23.
40. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods.* 2015;12(8):697.
41. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, Tang Y, Chen YG, Jin CN, Yu Y, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* 2020;48(D1):D148–54.
42. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4:2.
43. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium Nat Genet. 2000;25(1):25–9.
45. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009;25(8):1091–3.
46. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics.* 2013;29(5):661–3.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

