The Effect of SNPs in lncRNA as CeRNA on The Risk and Prognosis of Hepatocellular Carcinoma

Han Mo
Harbin Medical University
Xi Wang
Harbin Medical University
Guohua Ji
Harbin Medical University
Xiao Liang
Harbin Medical University
Yi Yang
Harbin Medical University
Wenjing Sun
Harbin Medical University
Xueyuan Jia
Harbin Medical University
Lidan Xu
Harbin Medical University
Yuandong Qiao
Harbin Medical University
Henan Zhou
Harbin Medical University
Wenhui Zhao
Harbin Medical University
Songbin Fu
Harbin Medical University
Xuelong Zhang (zhangxuelong@hrbmu.edu.cn)
Harbin Medical University https://orcid.org/0000-0002-1344-8619

Research

Keywords: HOTAIR, PVT1, EGFR-AS1, HCC, risk, prognosis, SNP, ceRNA

Posted Date: June 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-626542/v1

Abstract

Background

Most the HCC susceptible loci identified by GWAS are located in non-coding regions, and the mechanism of action remains unclear. The objective of this study was to explore the relationship among SNPs on lncRNA that affects ceRNA mechanism and the risk and prognosis of HCC.

Methods

In this study, combined with multiple databases, eight lncRNA genes that affect HCC were systematically screened through the mechanism of lncRNA-mediated ceRNA, and 15 SNPs that affect miRNA binding in the eight lncRNA genes were annotated. Genotyping was conducted in 800 HCC cases and 801 healthy controls from a Han population in Northeast China to examine the genetic risk associated with HCC.

Results

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

Conclusions

This first study demonstrated the association of SNPs on lncRNA -mediated ceRNAs with HCC. The results provide new insights of how non-coding loci confer the interindividual differences to susceptibility and progression of HCC.

Background

According to the Global Cancer Statistics 2018, primary liver cancer is the 6th most common cancer and the 4th leading cause of cancer mortality worldwide (new cases, 841 000/y; deaths, 782 000/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 [1], with more than 50% of the world's HCC cases coming from China [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, which may partly explain why only a small number of people develop HCC when exposed to the

same environment [5, 6]. And single nucleotide polymorphisms (SNPs) associated with gene expression, function, phenotype and disease represent the most common genetic variations among individuals [7, 8]. In the past few decades, more and more genome-wide association studies (GWAS) have been applied to identify HCC susceptibility SNPs [9], most of which are located in non-coding regions [10]. 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 [11]. Therefore, for SNPs in non-coding regions, explanation of the mechanism of action in HCC is from the perspective of regulatory.

Long non-coding RNAs (lncRNAs) are defined as a class of transcripts greater than 200 nucleotides, which are not used as protein templates and are usually transcribed by RNA polymerase II [12]. LncRNAs play a key role in various biological processes by participating in transcription, post-transcriptional, post-translational regulation and intercellular signal transduction through a variety of mechanisms, and are involved in complex human diseases, including autoimmune diseases, neurological diseases and various tumors [13]. MicroRNAs (miRNAs) are another small non-coding RNA molecule with a nucleotide length of about 22 that mainly mediate the process of post-transcriptional gene silencing [14]. 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 experiments [15]. For example, lncRNA PTENP1 can function as a decoy for PTEN-targeting miRNAs in tumor suppression [16]. In addition, with the development of high-throughput technologies, an increasing number of tumor-related studies have been published using The Cancer Genome Atlas (TCGA). The HCC related lncRNA-mediated ceRNA networks have also been reported in many recent TCGA-based articles [17].

It is well known that SNPs located in lncRNA-miRNA binding regions may affect their interactions (create or destroy), thereby altering their function and leading to genome-wide butterfly effects [18]. Several evidences indicate that this type of SNPs is involved in the occurrence and development of a variety of tumors. For example, studies showed that rs2147578 CG and GG genotypes were significantly associated with an increased risk of colorectal cancer by affecting the binding of lnc-LAMC2-1:1 to 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 [11]. Encouragingly, a large number of SNPs located in the lncRNA-miRNA binding region 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.

Therefore, one hypothesis is that SNPs on lncRNAs could affect the interaction between lncRNAs and miRNAs and participate in the occurrence and development of HCC. In order to verify the above hypothesis, experiments were systematically conducted to evaluate the functional SNPs affecting the binding of HCC related ceRNAs and miRNAs through the combined analysis of multiple databases including TCGA, RNALocate [21], lncATLAS [22] and lncRNASNP2. A case-control study was then conducted in a northeastern Han Chinese population to unearth the relationship between these candidate

SNPs and HCC risk. Finally, investigations were conducted to study the effect of positive SNPs on the potential regulatory axis of ceRNA associated with the occurrence and development of HCC.

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 differentially expressed lncRNAs (DElncRNAs), mRNAs (DEmRNAs) and miRNAs (DEmiRNAs) between the HCC group and the normal group were identified by edgeR package [23], and multiple test correction was performed using Benjamini and Hochberg false discovery rate (FDR) [24]. The differential expression threshold was FDR < 0.5 and |log2FC| (fold change) > 1.5. In this study, the pheatmap package was used to draw the hierarchical clustering heat map of DElncRNAs, DEmiRNAs and DEmRNAs data in R software.

Screening of SNPs in HCC related ceRNA that affect miRNA binding

As shown in the flow chart (Fig. 1A), firstly, DElncRNAs in HCC were screened through the Lnc2Cancer database [25], and DElncRNAs with consistent expression results verified by low-throughput experiments were selected. Then, the name and subcellular localization of lncRNAs were used as keywords in PubMed literature retrieval, and lncRNAs with evidence of subcellular localization in the cytoplasm were screened using RNALocate and lncATLAS tools. Next, the SNPs affecting the binding of lncRNA to miRNA on lncRNAs were annotated through the lncRNASNP2 database, and the miRNAs affected by these SNPs were DEmiRNAs filtered to extract the relevant SNPs after filtering. Finally, 22 SNPs with MAF (minor allele frequency) > 0.05 in Asian population were extracted from NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/snp/?term=).

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. And none of the patients received any chemotherapy or radiation therapy before sampling. 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.

Sample Collection Information

Each healthy control individual underwent the examination of antigen and antibody, at least one typical morphologic finding from CT or ultrasound. 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). 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 α-fetoprotein (AFP), aspartate aminotransferase (AST), alanine aminotransferase (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 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 Hardy-Weinberg equilibrium (HWE) for controls. ANONA (for continuous variables) and χ^2 test (for categorical variables) were conducted to compare the demographic characteristics between HCC cases and healthy controls. 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 [26] and miRTarBase [27] 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) [28] 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) and Gene Ontology (GO) were mainly used to annotate and analyze biological processes (BP) of genes. In this study, ClueGO (version 2.5.7) [29] and CluePedia (version 1.5.7) [30] 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 ≤ 0.01, GO Tree Interval Min Level = 4, GO Tree Interval Max Level = 7, Kappa Score = 0.6, Layout = yFiles Organic Layout.

Results

Differential expression analysis

The gene expression heatmaps of DElncRNAs, DEmRNAs and DEmiRNAs were shown in Fig. 1B, Fig. 1C and Fig. 1D. 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. 1A) and the Method section of this paper, eight HCC related lncRNA-mediated ceRNAs were selected after screening. The lncRNASNP2 database was used to annotate the above eight lncRNAs, and 2699 SNPs affecting the binding of lncRNA-miRNA were extracted from the database. These 2699 SNPs affected the binding of 2288 miRNAs, and 153 miRNAs were retained through DEmiRNAs filtering, of which 505 SNPs were involved. Finally, 22 SNPs with MAF > 0.05 in Asian

populations were screened from the 1000 Genome. Genomic visualization of the 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 the 22 candidate SNPs was evaluated. First, in the 48-Plex SNPscan™ typing system, the 22 candidate SNPs were classified into four levels: first, second, third and fail (Supplementary Fig. 1A). Subsequent genotyping was then abandoned for fail grade SNPs (rs7336379, rs9986879 and rs10259737). 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 19 other SNPs. The results showed strong LD between some SNPs (Supplementary Fig. 1B). Specifically, rs1050171 on *EGFR-AS1* can represent rs7795743, rs16846478 on GAS5 can represent rs58994962 and rs2235095, and rs3931282 on *PVT1* can represent rs3931281 for subsequent genotyping. Finally, the remaining 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 such as age, gender, smoking and alcohol consumption of 1601 samples enrolled in this study, as well as the clinical information and clinical test indexes of 800 HCC patients are shown in Table 1. No significant differences were observed for age, gender and cigarette smoking between cases and controls (all $P > 0.05$). 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 the 15 SNPs is shown in Fig. 2A, among which only rs2278176 does not conform to HWE and will be eliminated in subsequent studies. The results of χ^2 test showed that the distribution of GG, GC, GG + GC genotype and CC genotype of rs7958904 was significantly different between the case group and the control group (GG vs CC, $P = 0.028$; GC vs CC, $P = 0.008$; GG + GC vs CC, $P = 0.014$). Then binary logistic regression was used to explore the association between rs7958904 and HCC risk. Since age, gender, cigarette smoking and alcohol drinking were identified as risk factors for HCC [31], they were included as covariates to adjust for confounders (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). For other SNPs, there were no significant correlations between different genotypes and HCC risk.

Association analysis between 14 SNPs and clinical characteristics of HCC patients

The association between 14 SNPs and the 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 [32]. 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 ($P = 0.003$). In detail, the frequency of AA + GA genotypes was higher in early stage than that in advanced stage and showed a decreased risk for HCC progression (corrected $P = 0.042$, OR = 0.58, 95% CI = 0.40–0.83). In addition, the results of the association analysis of the remaining clinical characteristics are shown in Supplementary Fig. 2.

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

In this study, the levels of 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 in patients with TT genotype (corrected $P = 0.042$ and 0.014, respectively). Meanwhile, AST/ALT ratio was significantly higher in patients with rs10589312 TCTTGC/TCTTGC + TCTTGC/T genotype than in patients with TT genotype (corrected $P = 0.021$), and AST/ALT ratio were significantly lower in patients with rs84557 CC + CT genotype than in patients with TT genotype (corrected $P = 0.006$). In addition, the results of the association analysis of the clinical test indexes are shown in 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 by selecting the miRWalk and miRTarBase and screening the mRNAs shared by the two databases. For miR-615-3p, highly expressed DEmRNAs in tumors 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 Analyses 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 the 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 the 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 similar studies on HCC related ceRNA networks constructed based on TCGA database [17], the screening process of this study has significant characteristics: not only the Lnc2Cancer database was used to extract the lncRNAs supported by low-throughput experimental verification, but also the subcellular localization of lncRNAs was considered. The latter is due to the fact that the biological process of lncRNA competing with mRNA to bind miRNA mainly occurs in the cytoplasm [33]. Then, based on the lncRNASNP2 database, the 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 summary, in the association analysis, five positive SNPs that were significantly associated with susceptibility or prognosis of HCC were identified. Furthermore, a PubMed literature search was conducted using five SNPs as key words, with findings that 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 a better prognosis. The association between candidate SNPs and clinical test indexes in HCC cohort was also investigated. The results showed that AST and ALT levels of individuals with *PVT1* rs1134492 CC + TC genotype were higher than those in TT genotype. The AST/ALT ratio of individuals carrying *PVT1* rs10589312 TCTTGC allele was significantly higher than that of TT homozygous genotype. The AST/ALT ratio of individuals carrying *EGFR-AS1* rs84557 CC + TC genotype was significantly lower than that of TT genotypes. 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 tumor biomarkers and could be beneficial 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 many studies have reported that this SNP is associated with the risk of other cancers. For example, in a case-control study of South Korean population [34], individuals with rs7958904 GG genotype were found to have a lower risk of developing colorectal cancer compared to individuals with CC + GC and GC genotypes. This trend of increased risk of tumor in individuals with allele C is consistent with what was observed in our study. The present study showed that SNP rs7958904 was strongly correlated with the HCC risk. Compared with CC genotypes, individuals 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 [35]. One of the reasons for this contradiction may be that the regulatory axis of ceRNA affected by rs7958904 is different in different tumors. 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 variation 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 [36]. 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 variant 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 [37, 38]. 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 variant 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 [39] (Fig. 5C). In the fourth regulatory axis, the rs10589312 TCTTGC allele variant 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 [40, 41], and PPP2R2A overexpression can inhibit Akt phosphorylation, thus blocking the PI3K/Akt signaling

pathway and causing liver injury [42, 43] (Fig. 5D). In the fifth regulatory axis, the rs84557 T allele variant 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 [44] (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 a higher AST/ALT ratio than those with TT genotype, and patients with rs84557 TT genotype had a higher AST/ALT ratio 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 tumors 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 [45]. Another study showed that lncRNA *PVT1*, as a ceRNA, could compete with Atg3 to bind to microRNA-365 and promote autophagy in HCC cells [46]. Recently, lncRNA *EGFR-AS1* was found to be associated with migration, invasion and apoptosis of glioma cells by targeting miR-133b/RACK1 [47].

Conclusions

In summary, to our knowledge, this is the first study to show that SNPs in lncRNA-mediated ceRNAs are associated with the risk and prognosis of HCC. We found that *HOTAIR* rs7958904 can be used as a new predictor of HCC risk, *PVT1* SNPs (rs3931282, rs1134492 and rs10589312) and *EGFR-AS1* rs84557 can be used as new predictors of poor prognosis of HCC, which provides a scientific basis for improving the prevention and treatment of HCC. In addition, we also explained the reason why these positive SNPs were associated with HCC from the perspective of the regulatory axis of ceRNA. However, some limitations should be admitted here. Firstly, the samples in this HCC association study were only from the Han population in Northeast China, so the above association results need to be further verified in different populations. Secondly, 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.

Abbreviations

HBV, hepatitis B virus; HBsAg, hepatitis B surface antigen; HCV, hepatitis C virus; HCC, hepatocellular carcinoma; ALT, alanine aminotransferase; AST, aspirate aminotransferase; AFP, α-fetoprotein; BP, biological process; ceRNA, 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, Odd ratio; PPI, protein-protein interaction; SNP, single nucleotide polymorphism.

Declarations

Ethics approval statement

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

Consent for publication

Not applicable.

Availability of data and materials

The data-sets used and/or analysed during the current study available from the corresponding author on reasonable request.

Competing interests

All authors donot have anyconflicts of interest.

Funding

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

Authors' contributions

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

Acknowledgements

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

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology. 2007;132(7):2557–76.

3. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology. 2012;142(6):1264–73. e1261.
4. Aravalli RN, Steer CJ, Cressman EN. Molecular mechanisms of hepatocellular carcinoma. Hepatology. 2008;48(6):2047–63.
5. Churpek JE, Marquez R, Neistadt B, Claussen K, Lee MK, Churpek MM, Huo D, Weiner H, Bannerjee M, Godley LA, et al. Inherited mutations in cancer susceptibility genes are common among survivors of breast cancer who develop therapy-related leukemia. Cancer. 2016;122(2):304–11.
6. Wang X, Zhang X, Qiu B, Tang Y, Sun H, Ji H, Liu Y, Shi L, Song G, Yang Y. MDM2 SNP309T > G polymorphism increases susceptibility to hepatitis B virus-related hepatocellular carcinoma in a northeast Han Chinese population. Liver Int. 2012;32(7):1172–8.
7. De Wit P, Pespeni MH, Palumbi SR. SNP genotyping and population genomics from expressed sequences - current advances and future possibilities. Mol Ecol. 2015;24(10):2310–23.
8. Stracquadanio G, Vrugt B, Flury R, Schraml P, Wurl P, Muller TH, Knippschild U, Henne-Bruns D, Breitenstein S, Clavien PA, et al. CD44 SNPrs187115: A Novel Biomarker Signature that Predicts Survival in Resectable Pancreatic Ductal Adenocarcinoma. Clin Cancer Res. 2016;22(24):6069–77.
9. 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.
10. Hindorff 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.
11. 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.
12. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145–66.
13. Li J, Xuan Z, Liu C. Long non-coding RNAs and complex human diseases. Int J Mol Sci. 2013;14(9):18790–808.
14. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet. 2004;5(7):522–31.
15. 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.
16. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465(7301):1033–8.
17. 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 lnc-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. lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. Nucleic Acids Res. 2018;46(D1):D276–80.
21. Zhang T, Tan P, Wang L, Jin N, Li Y, Zhang L, Yang H, Hu Z, Zhang L, Hu C, et al. RNALocate: a resource for RNA subcellular localizations. Nucleic Acids Res. 2017;45(D1):D135–8.
22. Mas-Ponte D, Carlevaro-Fita J, Palumbo E, Hermoso Pulido T, Guigo R, Johnson R. LncATLAS database for subcellular localization of long noncoding RNAs. RNA. 2017;23(7):1080–7.
23. 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.
24. 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.
25. Gao Y, Wang P, Wang Y, Ma X, Zhi H, Zhou D, Li X, Fang Y, Shen W, Xu Y, et al. Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. Nucleic Acids Res. 2019;47(D1):D1028–33.
26. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nat Methods. 2015;12(8):697.
27. 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.
28. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2.
29. 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.
30. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics. 2013;29(5):661–3.
31. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. Cancer Control. 2017;24(3):1073274817729245.
32. 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.
33. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2016;17(1):47–62.

34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. Kedzierska H, Piekielko-Witkowska A. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. Cancer Lett. 2017;396:53–65.
42. 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.
43. 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.
44. McMahon SB. MYC and the control of apoptosis. Cold Spring Harb Perspect Med. 2014;4(7):a014407.
45. 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.
46. 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.
47. 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.

Tables

Table 1 Characteristics of the study population

Variable	Cases (N = 800)	Controls (N = 801)	P value
Age (years)	Mean ± S.D.	Mean ± S.D.	
	56.87 ± 10.25	56.97 ± 10.26	0.851[a]
Gender			
Male	652	688	
Female	148	133	0.319[b]
Cigarette smoking			
No	415	430	0.469[b]
Yes	385	371	
Alcohol consumption			
No	463	514	0.010[b]
Yes	337	287	
Family history			
No	757	-	-
Yes	43	-	-
HBsAg			
Negative	171	-	-
Yes	629	-	-
Anti-HCV			
Negative	697	-	-
Yes	103	-	-
Liver cirrhosis			
No	230	-	-
Yes	570	-	-
Child-Pugh grade			
A	696	-	-

[a] Student's t-test. [b] Chi-square test. S.D., standard deviation.

The bold and italicized value indicates statistical significance.

Variable	Cases (N = 800)	Controls (N = 801)	P value
B or C	104	-	-
Tumor size			
<5cm	418	-	-
≥5cm	382	-	-
Tumor number			
Solitary	400	-	-
Multiple	400	-	-
Vascular invasion			
No	541	-	-
Yes	259	-	-
Lymphatic metastasis			
No	614	-	-
Yes	186	-	-
Distant metastasis			
No	631	-	-
Yes	169	-	-
TNM staging			
I + II	275	-	-
III + IV	525	-	-
[a] Student's t-test. [b] Chi-square test. S.D., standard deviation.			
The bold and italicized value indicates statistical significance.			

Figures

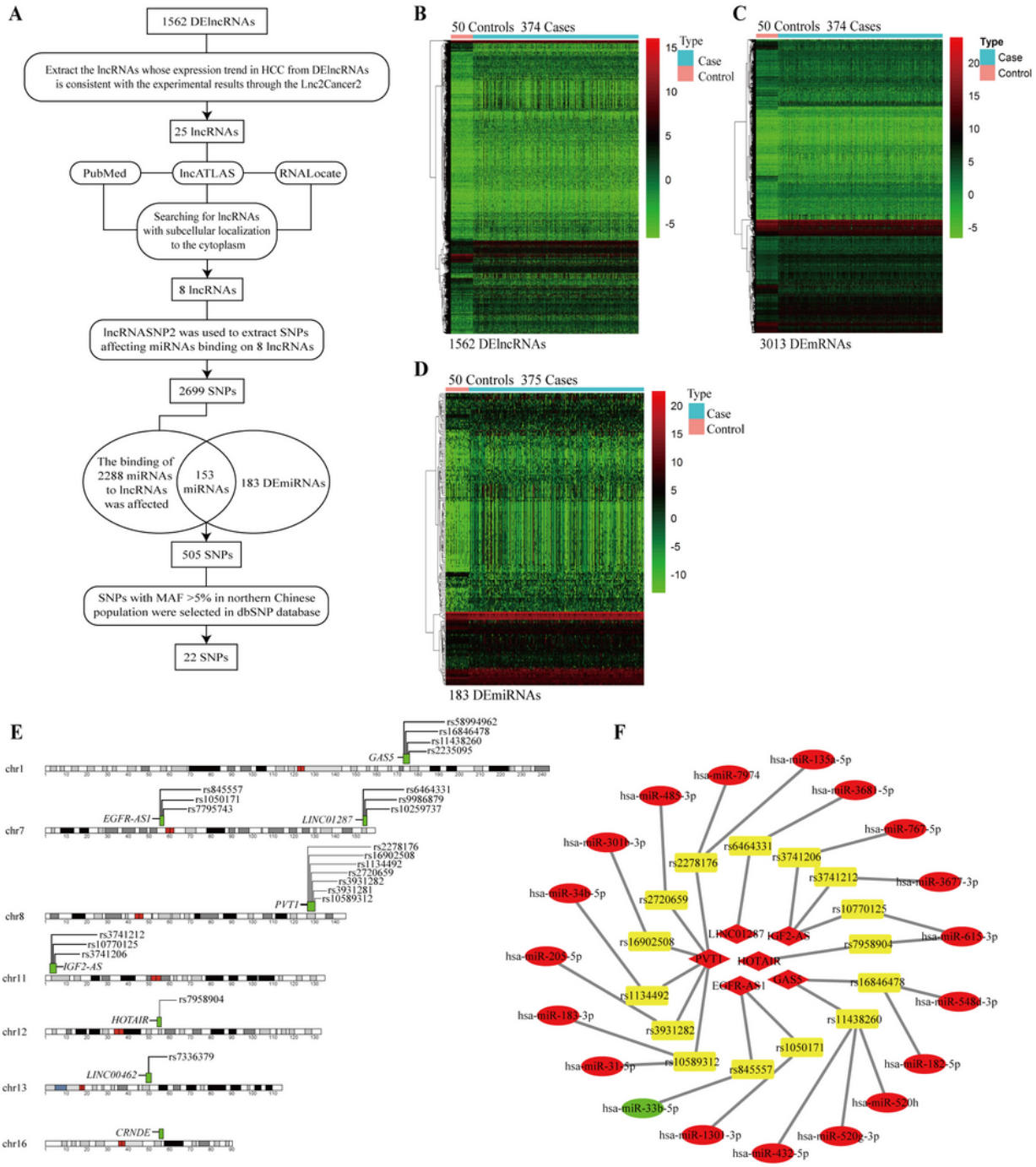

Figure 1

Screening result of lncRNA related SNPs. (A) Flow chart of screening lncRNA related SNPs. (B) Clustering heatmap of DElncRNAs. (C) Clustering heatmap of DEmRNAs. (D) Clustering heatmap of DEmiRNAs. In the heat map, red represents high expression and green represents low expression; pink represents the paracancerous control group and blue represents the HCC case group. (E) Genomic visualization of 8

lncRNAs and 22 candidate SNPs. (F) Related lncRNA-SNP-miRNA network diagram. Yellow rectangles are SNPs, red is high expression, green is low expression, ellipses are miRNAs, and diamonds are lncRNAs.

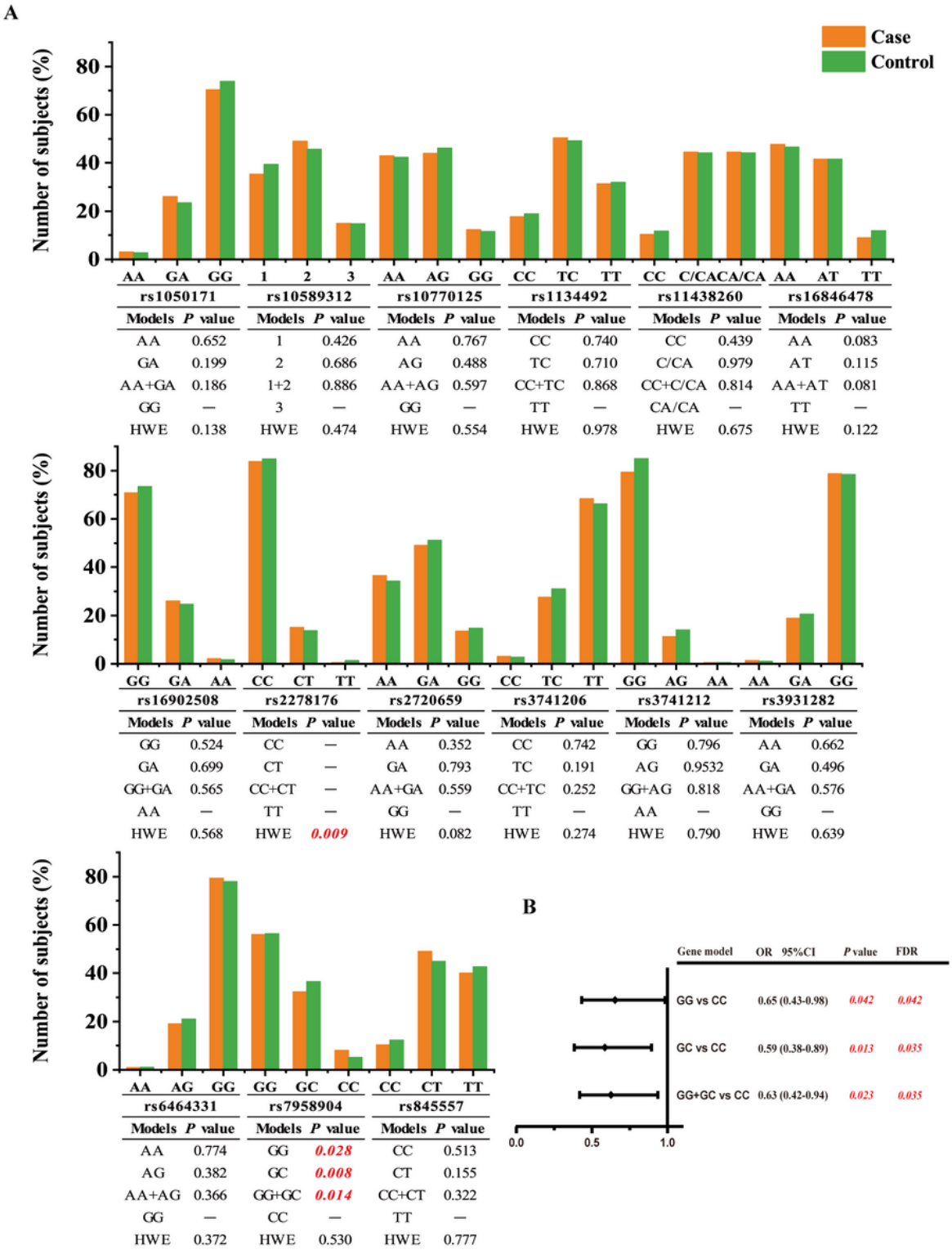

Figure 2

The genotype distribution map of 15 lncRNA-SNPs and association analysis of HCC risk. (A)The genotype distribution and HWE test of 15 lncRNA-SNPs. 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 and HCC risk.OR andP values were adjusted for age, gender, smoking and drinkingby logistic regression. The values in red italics are statistically significant.

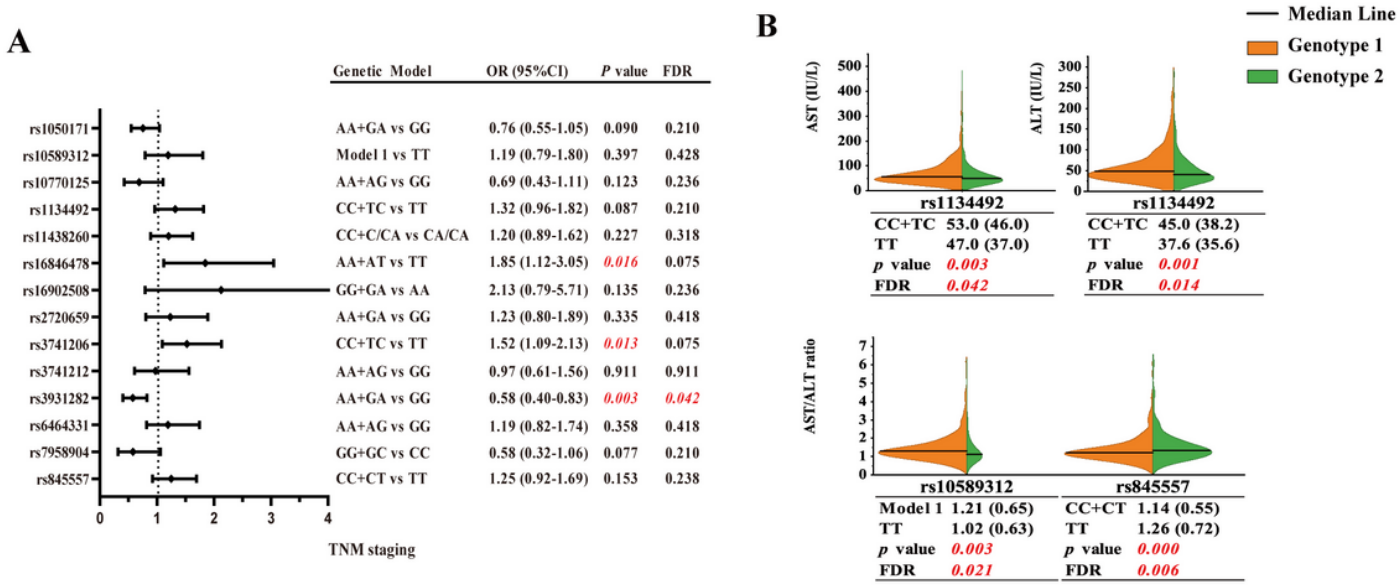

Figure 3

Association analysis of 14 SNPs and clinical characteristics and clinical test indexes in HCC patients.(A) Associationanalysis of 14 SNPs and TNM staging in HCC patients.(B) Significant results of correlation analysis with clinical test indicators. Model 1is the TCTTGC/TCTTGC+TCTTGC/T genetic model of rs10589312.Model 2 is the CC+C/CA genetic model of rs11438260. OR and P values were adjusted for age, gender, smoking, drinking, HBsAg and anti-HCVbylogistic regression. The values in red italics arestatistically significant.The size of theclinical test indexesis represented by the median value (inter quartile range).

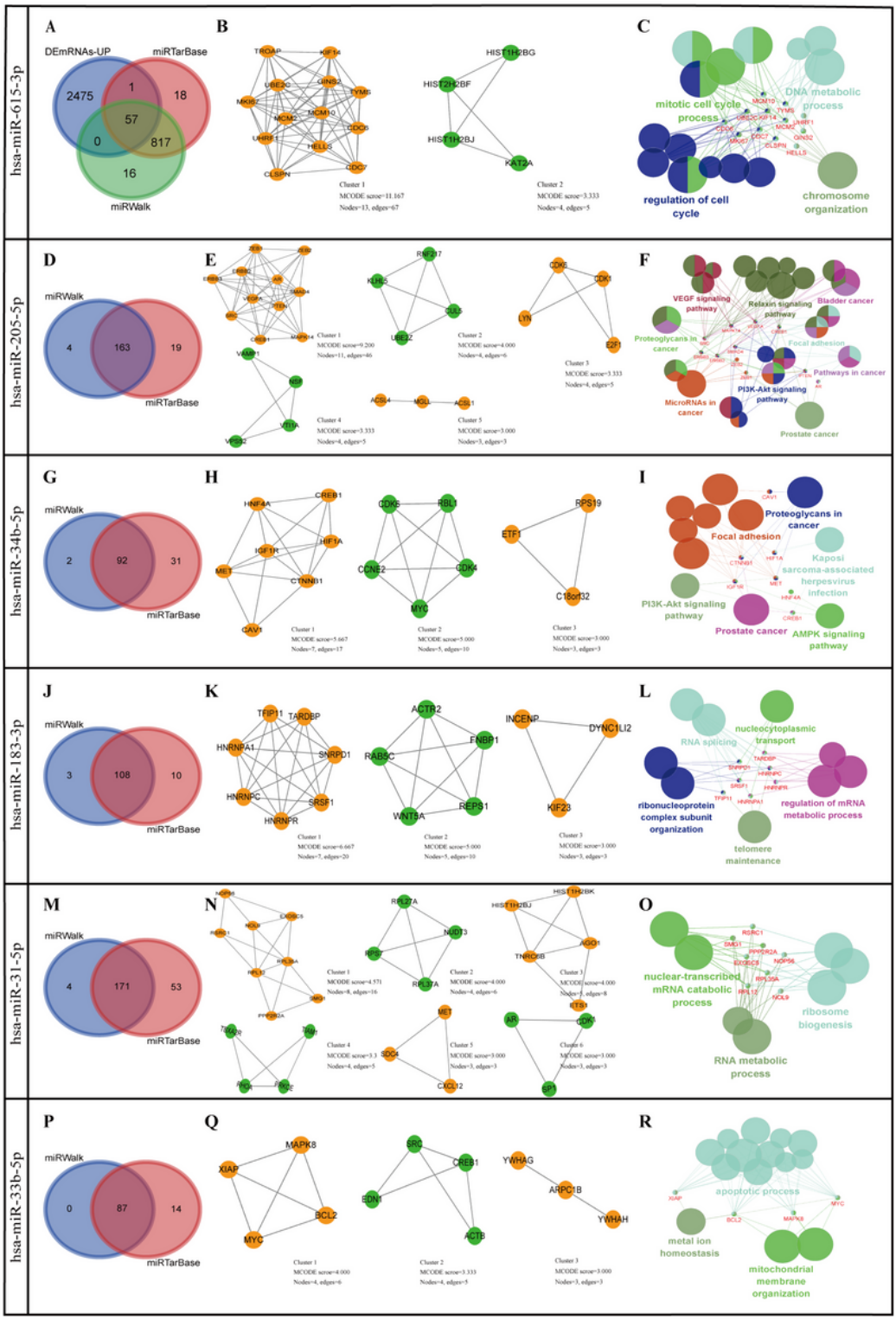

Figure 4

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.

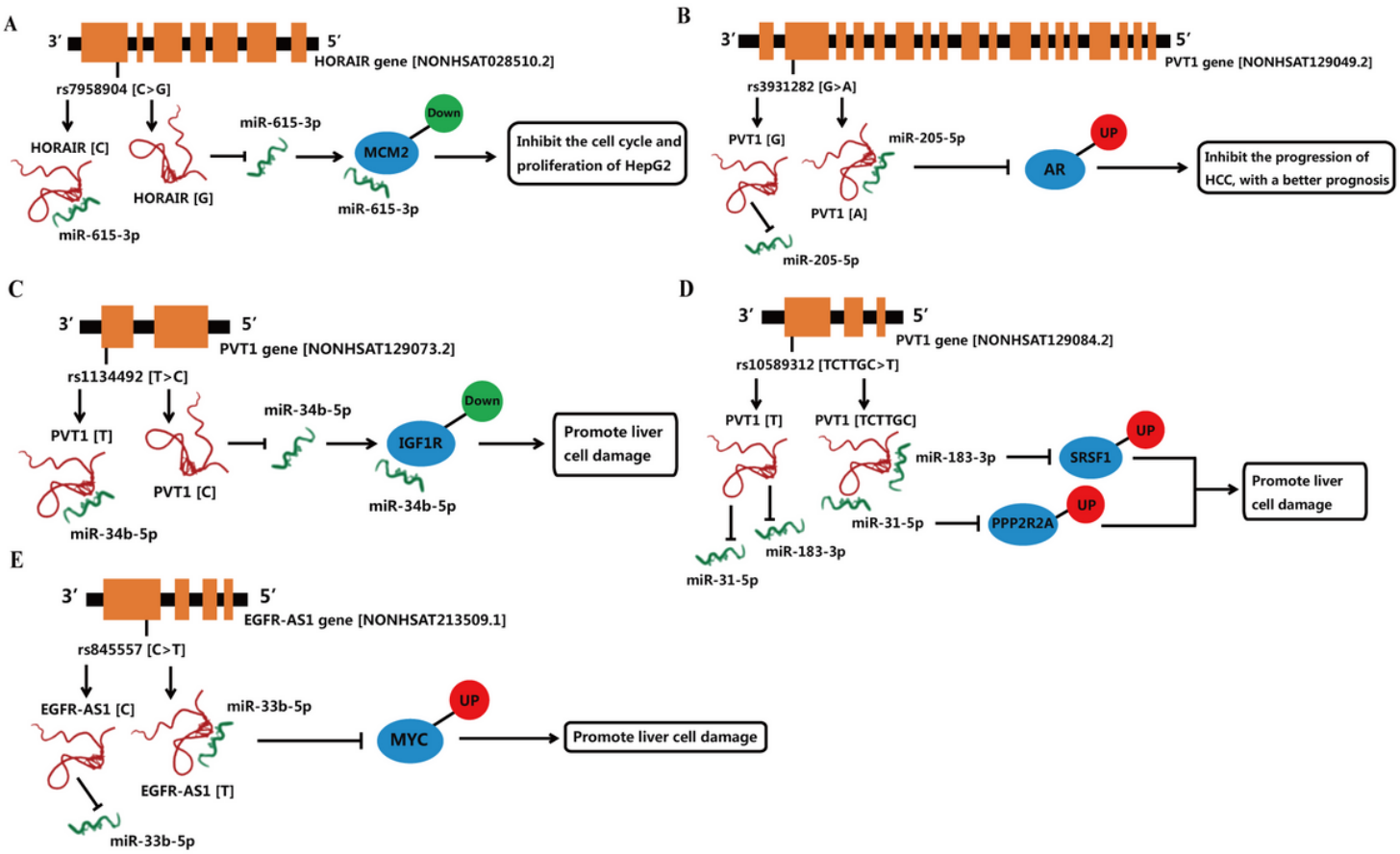

Figure 5

Six potential ceRNA regulatory axes affected by five positive SNPs.

Supplementary Files

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

- SupplementaryFigure1.tif
- SupplementaryFigure2.tif
- SupplementaryFigure3.tif