



## Genetic variants in pseudogene *E2F3P1* confer risk for HBV-related hepatocellular carcinoma in a Chinese population

Li Liu<sup>a,b,Δ</sup>, Yao Liu<sup>c,Δ</sup>, Jibin Liu<sup>b</sup>, Xiangjun Zhai<sup>d</sup>, Juan Wen<sup>c</sup>, Kaipeng Xie<sup>c</sup>, Hongbing Shen<sup>c</sup>,  
Zhibin Hu<sup>c</sup>, Zhining Fan<sup>a,✉</sup>

<sup>a</sup>Institute of Digestive Endoscopy and Medical Center for Digestive Diseases, the Second Affiliated Hospital, Nanjing Medical University, Nanjing, Jiangsu 210011, China;

<sup>b</sup>Department of Hepatobiliary Surgery, Nantong Tumor Hospital, Nantong, Jiangsu 226361, China;

<sup>c</sup>MOE Key Laboratory of Modern Toxicology, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention, and Treatment, and State Key Laboratory of Reproductive Medicine, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 210029, China;

<sup>d</sup>Department of Infection Diseases, Jiangsu Province Center for Disease Prevention and Control, Nanjing, Jiangsu 210009, China.

Received 20 February 2013, Revised 05 March 2013, Accepted 12 April 2013, Epub 25 April 2013

### Abstract

Recent studies showed that pseudogenes can regulate the expression of their coding gene partners by competing for miRNAs. The E2F family plays a crucial role in the control of cell cycle checkpoint. *E2F3P1* is a pseudogene of *E2F3*. Few studies focused on genetic variations on pseudogenes. In this study, we performed a case-control study to assess the association between single nucleotide polymorphisms (SNPs) in *E2F3P1* and hepatocellular carcinoma (HCC) risk in 1050 hepatitis B virus (HBV)-positive HCC cases and 1050 chronic HBV carriers. Logistic regression analysis was applied to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the associations between genotypes and HCC risk. We found that the variant CT/TT genotypes of rs1838149 were associated with a significantly decreased risk of HCC (adjusted OR = 0.66, 95% CIs = 0.51-0.86,  $P = 0.002$ ) compared to those with wildtype CC homozygote. Furthermore, the AA genotype of rs9909601 had an increased HCC risk with an adjusted OR of 1.41 (95% CIs = 1.07-1.86), and the A allele of rs9909601 was significantly associated with HCC risk compared to those with the G allele (adjusted OR = 1.17, 95% CIs = 1.03-1.33,  $P = 0.017$ ). These results indicate that genetic variations in the pseudogene *E2F3P1* may confer HCC risk.

**Keywords:** E2F3P1, single nucleotide polymorphism (SNP), hepatocellular carcinoma (HCC), susceptibility

This work was funded by the National Key Basic Research Program (2013CB911400), the Foundation for the Program for New Century Excellent Talents in University (NCET-10-0178), the Author of National Excellent Doctoral Dissertation (201081), the National Natural Science Foundation of China (30800946 and 81072344), the State Key Infectious Disease Project of China (2012ZX10002010, 2012ZX10002016), the National Major S&T Projects (2011ZX10004-902), the National Science Fund for Creative Research Groups (30921006), and the Priority Academic Program for the Development

of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

<sup>Δ</sup>These authors contributed equally to this study.

✉ Corresponding author: Zhining Fan, MD, Institute of Digestive Endoscopy and Medical Center for Digestive Diseases, the Second Affiliated Hospital, Nanjing Medical University, 121 Jiangjiayuan Rd, Nanjing, Jiangsu 210011, China. Tel: +86-25-585-09900. E-mail: [fanzhining@hotmail.com](mailto:fanzhining@hotmail.com).

The authors reported no conflict of interests.

## INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common cancers worldwide, with over 748,300 new HCC cases diagnosed per year<sup>[1]</sup>, among which about 55% are in China<sup>[2]</sup>. Chronic HBV infection is the most important risk factor for HCC, and the prevalence of HBV surface antigen (HBsAg) exhibited a similar distribution with HCC worldwide<sup>[2]</sup>. In China, about 80% of HCC patients have a history of HBV infection<sup>[3]</sup>. However, only a small fraction of chronic HBV carriers develop HCC. The segregation analysis of familial HCC indicated interaction between HBV infection and a major genetic locus<sup>[4,5]</sup>, suggesting that genetic variations may play roles in the carcinogenesis of HCC.

Pseudogenes are defined as defunct genomic loci with sequence similarity to functional genes but lacking coding potential due to the presence of disruptive mutations, such as frame shifts and premature stop codons<sup>[6]</sup>. Pseudogenes have long been considered as nonfunctional genomic sequences. However, there is accumulating evidence that many of them might have some biological activities<sup>[7,8]</sup>. Recent studies have suggested that pseudogenes can regulate the expression of their coding gene partners<sup>[9]</sup>. For sequence similarity between pseudogene and its coding gene partner, they may compete for miRNAs and then form a regulatory network across the transcriptome<sup>[8,9]</sup>. Genetic variations in both pseudogenes and miRNAs may disrupt the network. Recently, several studies paid attention to the association between the genetic variations of miRNAs and HCC risk<sup>[10-12]</sup>, but few publications focused on pseudogenes and HCC.

The E2F family plays a crucial role in the control of cell cycle<sup>[13]</sup>. E2F3a is a transcription factor that has been shown to be overexpressed in liver cancer tissues<sup>[14]</sup>. The activity of E2F1-3 is tightly controlled by the retinoblastoma (Rb) family of proteins<sup>[15]</sup>. Recent studies suggested that the expression of E2Fs is regulated by several miRNAs, including the miR-17-92 cluster<sup>[16]</sup>. E2F3P1 is a pseudogene with similar sequence to E2F3. E2F3P1 may regulate E2F3 expression by competing for miRNAs, and the sequence variation in E2F3P1 may influence miRNA binding affinity and may interrupt further cellular activity, including proliferation and apoptosis<sup>[14]</sup>.

## SUBJECTS AND METHODS

### Study subjects

This case control study protocol was approved by the institutional review board of the authors' affiliated

institution. The newly diagnosed HCC patients were consecutively recruited from January 2006 to December 2010 at the authors' affiliated hospital. The controls came from cohorts of Changzhou and Zhangjiagang, which were established in 2004 and 2009, respectively. Informed consent was obtained from each subject at recruitment. A 5-mL sample of whole blood was obtained from each participant as a source of genomic DNA for further genotyping analyses and serological testing for HBV/HCV markers.

### Serological testing

HBsAg, anti-HBs, anti-HBc and anti-HCV were detected by the enzyme-linked immunosorbent assay (Kehua Bio-engineering, Shanghai, China) in the serum following the manufacturer's instructions as described previously<sup>[17]</sup>. All cases were screened for HBV/HCV markers, and 1,050 HCC cases positive for HBV surface antigen (HBsAg) and negative for HCV antibody (anti-HCV) were included in this study. The controls were chronic HBV carriers, positive for both HBsAg and antibody to hepatitis B core antigen (anti-HBc), and negative for HCV antibody (anti-HCV). We randomly selected 1,050 chronic HBV carriers matched to the HCC cases on age and gender.

### SNP selection and genotyping

We surveyed identical sequences between *E2F3* and *E2F3P1*, and found five common SNPs located in *E2F3P1*, with at least 50 bps flanking regions identical to *E2F3*. Next, we searched miRBase (<http://www.mirbase.org/>) and Patrocles (<http://www.patrocles.org/>) to see whether the allelic change of the five SNPs may influence miRNA binding. Eventually, we found that two SNPs (rs9909601 G>A and rs1838149 C>T) might affect miRNA binding.

Genomic DNA was extracted from a leukocyte pellet by traditional proteinase K digestion and followed by phenol-chloroform extraction and ethanol precipitation. The two SNPs were genotyped by using the TaqMan allelic discrimination assay on the 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). The primers and probes for rs9909601 G>A were listed as follows: Primer: sense, 5'-CACGGTACGATAGTCTCTTGATCTG-3', anti-sense, 5'-TCCGCTGCCTTGTTCAAAC-3'; Probe: allele G, FAM-AGCTCCTGAGCCAGTCACCC-TAMRA, allele A, HEX-CAGCTCCTGAAC-CAGTCACCC-TAMRA. The primers and probes for rs1838149 C>T were listed as follows: Primer: sense, 5'-CACCCAACCTGGTTCTCCTTT-3', antisense, 5'-CCATGCTTCCATTCCCAGAA-3'; Probe: allele C, FAM-TCAGCAAGCCCCAGGACTG-TAMRA,

**Table 1** Demographic characteristics of the study subjects

Variable	Cases (n = 1,050)		Controls (n = 1,050)		OR (95% CI)	P
	n (%)	n (%)	n (%)	n (%)		
<b>Age (years)</b>						
≤ 53	556 (52.95)		539 (51.33)		1.00 (0.99-1.01)	0.692
> 53	494 (47.05)		511 (48.67)			
<b>Gender</b>						
Male	993 (94.57)		994 (94.67)		1.02 (0.70-1.49)	0.923
Female	57 (5.43)		56 (5.33)			
<b>Smoking status</b>						
Ever	633 (60.29)		606 (57.71)		0.90 (0.76-1.07)	0.231
Never	417 (39.71)		444 (42.29)			
<b>Drinking status</b>						
Ever	625 (59.52)		485 (46.19)		0.58 (0.49-0.69)	< 0.001
Never	425 (40.48)		565 (53.81)			

allele T, HEX-TGTCAGCAAGCCTCAGGACTG-TAMRA. Genotyping was performed blindly without knowing the subjects' case or control status. Two blank (water) controls in each 384-well plate were used for quality control, and more than 10% samples were randomly selected to repeat, yielding a 100% concordance rate.

### Statistical analysis

Differences in demographic characteristics between cases and controls were detected by Student's *t*-test and  $\chi^2$  test for continuous variables and categorical variables, respectively. Logistic regression analyses were performed to compute odds ratios (ORs) and 95% confidence intervals (CIs) from the association between the genotypes and risk of HCC. The crude ORs were calculated from univariate logistic regression, while the adjusted ORs from multivariate logistic regression with the adjustment of age, gender, and smoking and drinking status. Chi-square-based *Q*-test was applied to test the heterogeneity of association

between subgroups. All tests were two-sided. All of the statistical analyses were performed with R software (version 2.15.1; The R Foundation for Statistical Computing, Vienna, Austria).

### RESULTS

The demographic characteristics of 1,050 HBV positive HCC cases and 1,050 chronic HBV carriers are summarized in **Table 1**. No significant differences were found in age and gender distributions between the cases and controls ( $P = 0.692$  and  $0.923$ , respectively). The smoking rates of the two groups were also similar ( $P = 0.231$ ). However, the drinking rates were significantly higher among the cases than those of the controls ( $P < 0.001$ ).

The genotyping call rates for rs9909601 and rs1838149 were 99.57% and 99.43%, respectively. The observed genotype frequencies among the study subjects were in Hardy-Weinberg Equilibrium (HWE) ( $P = 0.284$  and  $0.269$  for rs9909601 and rs1838149, respectively). The genotype distributions of the two

**Table 2** Distribution of alleles and genotypes of two SNPs and their association with hepatocellular carcinoma (HCC) risk

Genotype	Cases (N = 1050)		Controls (N = 1050)		Crude OR (95% CIs)	P	Adjusted OR (95% CIs) <sup>a</sup>	P
	n	%	n	%				
<b>rs9909601</b>								
	n = 1046	n = 1045						
GG	381	36.4	423	40.5	1		1	
AG	505	48.3	499	47.8	1.12 (0.93-1.35)		1.12 (0.93-1.35)	
AA	160	15.3	123	11.8	1.44 (1.10-1.90)		1.41 (1.07-1.86)	
AG/AA	665	63.6	622	59.5	1.19 (1.00-1.42)	0.056	1.18 (0.98-1.41)	0.074
A allele				1.18 (1.04-1.34)	0.010	1.17 (1.03-1.33)	0.017	
<b>rs1838149</b>								
	n = 1043	n = 1045						
CC	936	89.7	893	85.5	1		1	
CT	104	10.0	150	14.4	0.66 (0.51-0.86)		0.65 (0.50-0.85)	
TT	3	0.3	2	0.2	1.43 (0.24-8.58)		1.48 (0.24-9.02)	
CT/TT	107	10.3	152	14.5	0.67 (0.52-0.87)	0.002	0.66 (0.51-0.86)	0.002
T allele					0.69 (0.54-0.90)	0.005	0.68 (0.53-0.88)	0.004

<sup>a</sup>Logistic regression analyses adjusted for age, gender, and smoking status and drinking status. OR: odds ratios; CIs: confidence intervals.

**Table 3** Stratified analyses of rs1838149 and hepatocellular carcinoma (HCC) risk

Characteristic	Cases		Controls		OR (95% CIs) <sup>a</sup>	P <sup>b</sup>
	CC/CT+TT	CC/CT+TT	CC/CT+TT	CC/CT+TT		
<b>Age (years)</b>						
≤ 53	497 (90.4)/53 (9.6)	452 (84.5)/83 (15.5)	0.54 (0.37-0.79)	0.142		
> 53	439 (89.0)/54 (11.0)	441 (86.5)/69 (13.5)	0.81 (0.55-1.19)			
<b>Gender</b>						
Male	883 (89.5)/104 (10.5)	850 (85.8)/141 (14.2)	0.71 (0.54-0.93)	0.037		
Female	53 (94.6)/3 (5.4)	43 (79.6)/11 (20.4)	0.14 (0.03-0.6)			
<b>Smoking status</b>						
Yes	564 (90.0)/63 (10.0)	528 (87.6)/75 (12.4)	0.73 (0.51-1.06)	0.309		
No	372 (89.4)/43 (10.6)	365 (82.6)/77 (17.4)	0.55 (0.37-0.83)			
<b>Drinking status</b>						
Yes	561 (90.3)/60 (9.7)	408 (84.5)/75 (15.5)	0.60 (0.41-0.86)	0.478		
No	375 (88.9)/47 (11.1)	485 (86.3)/77 (13.7)	0.73 (0.49-1.08)			

<sup>a</sup>Adjusted for age, gender, and smoking and drinking status except for the stratifying factor. <sup>b</sup>P for heterogeneity. OR: odds ratios; CIs: confidence intervals.

SNPs in the cases and controls are shown in Table 2. Compared to individuals carrying the wildtype GG genotype of rs9909601, those with AA genotype had an increased HCC risk with an adjusted OR of 1.41 (95% CIs = 1.07-1.86). In the dominant genetic model, the AG/AA genotype of rs9909601 was associated with HCC risk in borderline significance (adjusted OR = 1.18, 95% CIs = 0.98-1.14,  $P = 0.074$ ). In the additive genetic model, the A allele of rs9909601 was significantly associated with HCC risk (adjusted OR = 1.17, 95% CIs = 1.03-1.33,  $P = 0.017$ ). The minor allele frequency (MAF) of rs1838149 was smaller than that of rs9909601. The variant CT/TT genotypes of rs1838149 significantly decreased HCC risk by 33% (crude OR = 0.67, 95% CIs = 0.52-0.87,  $P = 0.002$ ) and 34% (adjusted OR = 0.66, 95% CIs = 0.51-0.86,  $P = 0.002$ ) before and after adjustment for age, gender, and smoking and drinking status, respectively. Similar results were found in the additive model as the T allele of rs1838149 was significantly associated with HCC risk (adjusted OR = 0.68, 95% CIs = 0.53-0.88,  $P = 0.074$ ).

The association between rs1838149 and the susceptibility to HCC was also evaluated by stratifying by age, gender, and smoking and drinking status (Table 3). However, no significant heterogeneity was detected between the subgroups except for gender. Females exhibited more significant protective effect than males.

## DISCUSSION

In this study, we investigated the association between two SNPs, rs9909601 and rs1838149, and the susceptibility to HCC in a Chinese population. We found that the C to T base change of rs1838149 demonstrated protective effect on HCC and the A allele

of rs9909601 was associated with increased HCC risk compared with the G allele. In this study, the powers of association between rs9909601 and rs1838148 and HCC risk were 45% and 67%, respectively.

According to web-based prediction tools (miRBase and Patrocles), the variant T allele of rs1838149 had a higher affinity to miR-1200. Similarly, the wildtype G allele of rs9909601 is more prone to bind to miR-24 and miR-149 than the variant A allele. In this study, both alleles of the two SNPs in *E2F3P1* predicted greater binding affinity of miRNAs demonstrated protective effect on HCC. Among these miRNAs, both miR-24 and miR-149 may be oncogenic. MiR-24 was known to be directly involved in breast cancer cell invasiveness<sup>[16]</sup> and was found upregulated both in breast cancer patients' sera and cancer tissues<sup>[17]</sup>. MiR-149, a p53-responsive miRNA, leads to resistance to apoptosis in melanoma cells<sup>[18]</sup> and might be involved in the invasion and metastasis of nasopharyngeal carcinoma through the regulation of epithelial-mesenchymal transition<sup>[19]</sup>. So far, few studies focused on miR-1200. In this study, the protective alleles of the two SNPs of *E2F3P1* were predicted to bind to more miRNAs, as it is biologically plausible that the two SNPs were associated with HCC risk.

*E2F3P1* and *E2F3* are located in chromosome 17 and 6, respectively. Though they are not on the same chromosome, they may regulate each other through competing for miRNAs, and their expression was positively correlated<sup>[6]</sup>. *E2F3* is an oncogene with strong proliferative potential and regularly dysregulated or overexpressed in cancer<sup>[20]</sup>. The alleles with higher affinity with miRNAs suggest that fewer amounts of *E2F3P1* RNA exist, as well as *E2F3*, because miRNAs negatively regulated target RNA expression either by

inhibiting its translation or inducing its degradation<sup>[21]</sup>. Thus, it is biologically plausible that the two SNPs were associated with HCC. However, further functional analysis is warranted to verify this hypothesis.

## References

- [1] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; 61: 69-90.
- [2] Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005; 55: 74-108.
- [3] Tanaka M, Katayama F, Kato H, Tanaka H, Wang J, Qiao YL, et al. Hepatitis B and C virus infection and hepatocellular carcinoma in China: a review of epidemiology and control measures. *J Epidemiol* 2011; 21: 401-16.
- [4] Shen FM, Lee MK, Gong HM, Cai XQ, King MC. Complex segregation analysis of primary hepatocellular carcinoma in Chinese families: interaction of inherited susceptibility and hepatitis B viral infection. *Am J Hum Genet* 1991; 49: 88-93.
- [5] Yu MW, Chang HC, Liaw YF, Lin SM, Lee SD, Liu CJ, et al. Familial risk of hepatocellular carcinoma among chronic hepatitis B carriers and their relatives. *J Natl Cancer Inst* 2000; 92: 1159-64.
- [6] Echols N, Harrison P, Balasubramanian S, Luscombe NM, Bertone P, Zhang Z, et al. Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Res* 2002; 30: 2515-23.
- [7] Balakirev ES, Ayala FJ. Pseudogenes: are they "junk" or functional DNA? *Annu Rev Genet* 2003; 37: 123-51.
- [8] Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011; 146: 353-8.
- [9] 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: 1033-8.
- [10] Liu Y, Zhang Y, Wen J, Liu L, Zhai X, Liu J, et al. A genetic variant in the promoter region of miR-106b-25 cluster and risk of HBV infection and hepatocellular carcinoma. *PLoS One* 2012; 7: e32230.
- [11] Li XD, Li ZG, Song XX, Liu CF. A variant in microRNA-196a2 is associated with susceptibility to hepatocellular carcinoma in Chinese patients with cirrhosis. *Pathology* 2010; 42: 669-73.
- [12] Xu Y, Liu L, Liu J, Zhang Y, Zhu J, Chen J, et al. A potentially functional polymorphism in the promoter region of miR-34b/c is associated with an increased risk for primary hepatocellular carcinoma. *Int J Cancer* 2011; 128: 412-7.
- [13] Wong JV, Dong P, Nevins JR, Mathey-Prevot B, You L. Network calisthenics: control of E2F dynamics in cell cycle entry. *Cell Cycle* 2011; 10: 3086-94.
- [14] Lees JA, Saito M, Vidal M, Valentine M, Look T, Harlow E, et al. The retinoblastoma protein binds to a family of E2F transcription factors. *Mol Cell Biol* 1993; 13: 7813-25.
- [15] Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, Major F, et al. An E2F/miR-20a autoregulatory feedback loop. *J Biol Chem* 2007; 282: 2135-43.
- [16] Li W, Ni GX, Zhang P, Zhang ZX, Wu Q. Characterization of E2F3a function in HepG2 liver cancer cells. *J Cell Biochem* 2010; 111: 1244-51.
- [17] Hu L, Zhai X, Liu J, Chu M, Pan S, Jiang J, et al. Genetic variants in human leukocyte antigen/DP-DQ influence both hepatitis B virus clearance and hepatocellular carcinoma development. *Hepatology* 2012; 55: 1426-31.
- [18] Papadimitriou E, Vasilaki E, Vorvis C, Iliopoulos D, Moustakas A, Kardassis D, et al. Differential regulation of the two RhoA-specific GEF isoforms Net1/Net1A by TGF-beta and miR-24: role in epithelial-to-mesenchymal transition. *Oncogene* 2012; 31: 2862-75.
- [19] Wu Q, Wang C, Lu Z, Guo L, Ge Q. Analysis of serum genome-wide microRNAs for breast cancer detection. *Clin Chim Acta* 2012; 413: 1058-65.
- [20] Jin L, Hu WL, Jiang CC, Wang JX, Han CC, Chu P, et al. MicroRNA-149\*, a p53-responsive microRNA, functions as an oncogenic regulator in human melanoma. *Proc Natl Acad Sci U S A* 2011; 108: 15840-5.
- [21] Luo Z, Zhang L, Li Z, Jiang C, Dai Y, Liu X, et al. miR-149 promotes epithelial-mesenchymal transition and invasion in nasopharyngeal carcinoma cells. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* 2011; 36: 604-9.
- [22] Miles WO, Tschop K, Herr A, Ji JY, Dyson NJ. Pumilio facilitates miRNA regulation of the E2F3 oncogene. *Genes Dev* 2012; 26: 356-68.
- [23] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; 120: 15-20.