An Insertion/deletion Polymorphism within RERT-lncRNA Modulates Hepatocellular Carcinoma Risk

Zhansheng Zhu1†, Xueren Gao†, Yan He††, Hua Zhao2, Qiang Yu3, Deke Jiang4, Pingzhao Zhang4,5, Xiaopin Ma4, Huixing Huang4, Dong Dong1, Jiao Wan1, Zhenyong Gu1, Xinghong Jiang6, Long Yu4*, Yuzhen Gao1*

1 Department of Forensic Medicine, Medical College of Soochow University, Suzhou, 215123, Jiangsu, P.R. China
2 Department of General Surgery, the First Affiliated Hospital of Soochow University, Suzhou, 215006, Jiangsu, P.R. China
3 Department of Gastroenterology, Suzhou Municipal Hospital, Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, 215002, Jiangsu, P.R. China
4 The State Key Laboratory of Genetic Engineering, Fudan University, Shanghai, 200433, P. R. China
5 Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, P. R. China
6 Department of Neurobiology and Psychology, Key Laboratory of Pain Research & Therapy, Medical College of Soochow University, Suzhou, 215123, Jiangsu, P.R. China

† These authors contributed equally to this work.

Running title: Indel polymorphism within RERT-lncRNA modulates HCC risk
Précis: An insertion/deletion polymorphism within a novel long non-coding RNA is associated with EGLN2 expression and HCC susceptibility.
Keywords: hepatocellular carcinoma; rs10680577; indel polymorphism; lncRNA; EGLN2
Grant Support: This study is supported by grants from Natural Science Foundation of China (30800621 and 81171893 to Y. Gao, 81201574 to Y. He).

*Corresponding author:
Yuzhen Gao, Department of Forensic Medicine, Medical College of Soochow University, Suzhou, 215123, Jiangsu, P.R. China. Phone: +86-512-65880109 Fax: +86-512-65125019, E-mail: yuzhengao@suda.edu.cn and Long Yu, The State Key Laboratory of Genetic Engineering, Fudan University, 220 Handan Road, Shanghai 200433, P. R. China. Phone: +86-21-65643954 Fax: +86-21-65643250, E-mail: longyu@fudan.edu.cn

Conflicts of interest: No potential conflicts of interest were disclosed.
Word count for main text (4609), total number of figures (4), and tables (2)
Abstract

The Prolyl hydroxylase 1 (EGLN2) is known to affect tumorigenesis by regulating the degradation of hypoxia-inducible factor (HIF). Polymorphisms in EGLN2 may facilitate cancer cell survival under hypoxic conditions and directly associate with cancer susceptibility. Here, we examined the contribution of a 4-bp insertion/deletion polymorphism (rs10680577) within the distal promoter of EGLN2 to the risk of hepatocellular carcinoma (HCC) in Chinese populations. The contribution of rs10680577 to HCC risk was investigated in 623 HCC cases and 1242 controls and replicated in an independent case-control study consisting of 444 HCC cases and 450 controls. Logistic regression analysis showed that the deletion allele of rs10680577 was significantly associated with increased risk for HCC occurrence in both case-control studies (OR=1.40; 95%C.I.=1.18-1.66, \( P<0.0001 \); OR=1.49; 95%C.I.=1.18-1.88, \( P=0.0007 \)). Such positive association was more pronounced in current smokers (OR=3.49, 95%C.I.=2.24-5.45) than non-smokers (OR=1.24, 95%C.I.=1.03-1.50) (heterogeneity \( P=0.0002 \)). Genotype-phenotype correlation studies demonstrated that the deletion allele was significantly correlated with higher expression of both EGLN2 and RERT-lncRNA (a long non-coding RNA whose sequence overlaps with Ras-related GTP-binding protein 4b (\( RAB4B \)) and EGLN2) \textit{in vivo} and \textit{in vitro}. Furthermore, RERT-lncRNA expression was also significantly correlated with EGLN2 expression \textit{in vivo}, consistent with \textit{in vitro} gain-of-function study that showed overexpressing RERT-lncRNA up-regulated EGLN2. Finally, \textit{in-silico} prediction suggested that the insertion allele could disrupt the structure of
RERT-lncRNA. Taken together, our findings provided strong evidence for the hypothesis that rs10680577 contributes to hepatocarcinogenesis, possibly by affecting RERT-lncRNA structure and subsequently EGLN2 expression, making it a promising biomarker for early diagnosis of HCC.
Introduction

Hepatocellular carcinoma (HCC) is the most common primary malignancy of liver in adults and the third leading cause of cancer-related mortality (1). More than 80% HCC cases are found in Asia and Africa, and over 50% of such cases are from mainland China (2). It is well-established that hepatocarcinogenesis is a complicated process and that multiple risk factors are involved in its initiation, promotion and progression (3, 4). Although much effort has been put in to study the carcinogenesis of HCC in recent years, the molecular mechanisms of HCC remain poorly understood (5). As one of the important carcinogens for HCC, hepatitis infection has become a significant public health problem in most Asian populations, and much is known about the mechanisms through which hepatitis influences HCC risk (6). Besides hepatitis infection, compelling evidence suggests the involvement of host genetic factors in HCC carcinogenesis (3). Thus, it is of particular interest in identifying HCC susceptibility genes, which will definitely benefit the prediction of HCC risks, and the exploration of approaches to prevent HCC carcinogenesis.

Hypoxia has now been recognized as one of the best validated target in controlling tumor progression and resistance to therapy (7-8), in which hypoxia-inducible factor (HIF), a transcription factor that regulates oxygen homeostasis, plays key roles in balancing O₂ homeostasis (9). HIF regulates a program of gene expression that facilitates cell survival under hypoxic conditions through cell-intrinsic changes in metabolism and cell-extrinsic changes affecting oxygen delivery. Prolyl hydroxylases (PHD1, also known as EGLN2) is one of the three enzymes capable of hydroxylating
the alpha subunit of HIF and results in polyubiquitinylation and proteasomal degradation of HIF (10). Several lines of evidence have indicated that EGLN2 is involved in development of multiple cancers (11-14). Whether EGLN2 promotes or suppresses tumorigenesis depends on its cell and cancer type-specific expression (15). In terms of HCCs, low EGLN2 expression is associated with poor prognosis (16). In the current case control study, by using a candidate-gene-based approach, we evaluated the association between a 4-bp indel (rs10680577) within the distal promoter of *EGLN2* and HCC susceptibility in Chinese populations. Additional experimental and *in-silico* studies were used to assess the possible functional significance of this polymorphism.

**Materials and methods**

**Study populations**

Our study included two independent case-control sets including 1067 newly diagnosed incident HCC cases and 1692 healthy and genetically unrelated ethnic Han Chinese. The Suzhou case control set contained 623 HCC patients who had been diagnosed, hospitalized and treated in the affiliated hospitals of Soochow University or the Suzhou Municipal Hospital between 2007 and 2010; the Shanghai case control set contained 444 HCC patients who were recruited by Qidong Liver Cancer Research Institute of Jiangsu Province from 2008 to 2010. All patients had never received any medical treatments. The diagnosis of the cases, the inclusion and exclusion criteria for the cases and controls were described in detail previously (17-19). Briefly, the
diagnosis of patients was validated by combinations of pathological examination and positive imaging (i.e. Magnetic resonance imaging and/or computerized tomography).

A total of 1692 controls, 1242 from Suzhou and 450 from Shanghai case control set, were selected with comparable age and sex with HCC cases. These cancer-free individuals were obtained from a community nutritional survey that was conducted in the same regions during the same period as the recruitment of cancer patients. After obtaining informed consents from each individual, we extracted genomic DNA from the peripheral blood of cases and controls. All participants were negative for antibodies to hepatitis C virus, hepatitis D virus or HIV. Each subject was interviewed in-person using a structured questionnaire to obtain information on demographic data and related risk factors, including smoking and drinking status. The “current smokers” were individuals who had kept smoking almost every day for more than one year till the time of interview; and the “former smokers” were those who experienced the same degree of smoking as the “current smokers”, but stopped smoking at least one year prior to the interview; the non-smokers were those either never smoked or seldom did. Subjects were considered as “light drinkers”, if they consumed 1-2 alcohol drinks per week for more than one year. Those who consumed more than 2 alcohol drinks per week for more than one year were categorized as “heavy drinkers”. “Non-drinkers” were those either never drank or seldom did.

Tumor tissues from a total of 72 patients with a diagnosis of HCC were collected according to the availability of frozen-stored tissues of HCC resections from 2004 to 2006. The 72 HCC cases were confirmed by pathological diagnosis and none of these
patients had ever received preoperative chemotherapy or radiotherapy. Tumor stages were determined according to a modified American Joint Committee on Cancer (AJCC) and international union against cancer (UICC) standard. After surgical resection, fresh tissues were immediately stored at -80°C until DNA/RNA isolation and protein extraction. The design of the study was approved by the Ethical Committee of Soochow University.

**DNA extraction and genotyping**

Genomic DNA of blood samples, HCC tumor tissues and cell lines were extracted using genomic DNA purification kit (Qiagen). DNA fragments containing rs10680577 were amplified with genotyping primers (genotyping-F and genotyping-R) listed in supplemental Table 1. Genotyping was performed without knowledge of case or control status as previously described (17). To validate the genotyping method, we also analyzed 50 randomly selected DNA samples by direct sequencing; the results for these two methods were 100% concordant. Approximately 10% of the case and control samples were randomly selected and tested in duplicate by independent technicians, with 100% concordance of results.

**EGLN2 promoter-reporter constructs**

To investigate the influence of rs10680577 on EGLN2 promoter activity, a genomic region from -1724 to +46 relative to the transcriptional start site of EGLN2 was amplified with primers clone-F and clone-R as listed in supplemental Table 1 from a
homozygous human genomic DNA sample containing a “TACT” insertion. The PCR products were separated in 1% agarose gel and extracted, purified, and cloned using TA cloning Kit (Cat # A1360, Promega). The insert containing “TACT” was confirmed by direct sequencing. Fragments were digested with Hind III and Bgl II, and subcloned into pGL3-basic vector (Cat # E1751, Promega). To obtain the same promoter sequence with a different allele of rs10680577, another allele was generated using QuikChange® Lightening Site-Directed Mutagenesis Kit (Cat # 210518, Stratagene) with mutagenesis primers (mutagenesis-F and mutagenesis-R as listed in supplemental Table 1). The resulting constructs were verified by direct sequencing.

**Cell culture and luciferase reporter assay**

The HepG2, Hep3B, sk-Hep-1, SMMC-7721 hepatoma cell lines and L02 immortalized hepatic cell line were obtained directly from Shanghai Cell Bank of Chinese Academy of Sciences. Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37°C in a humidified 5% CO₂ incubator. The cell lines were characterized by DNA fingerprinting analysis using short tandem repeat (STR) markers. All cell lines were placed under cryostage after removal from the bank and used within three months after thawing fresh vials. Cells were seeded at a density of 1×10^5 cells per well in 24-well plates (BD Biosciences). 16 hours after plating, cells were transfected by Lipofectamin 2000 (Cat #11668-019, Invitrogen) according to manufacturer’s protocol. In each well, 500ng pGL3-basic vector containing either
“TACT” insertion or deletion allele and 50ng pRL-TK vector (Promega) were co-transfected into cells. The empty pGL3-basic vector was added as negative control. 24 hr after transfection, cells were harvested immediately after addition of 100μl passive lysis buffer. Firefly luciferase activity in cell lysate was measured with the Dual Luciferase assay system (Cat #E1910, Promega) in TD-20/20 luminometer (Turner Biosystems) and was normalized with the Renilla luciferase activity. Six replicates per group were performed and each experiment was repeated at least three times.

**In-silico prediction of rs10680577 on RERT- IncRNA folding structures**

As certain conserved structures more likely serve important biological functions, a 35-bp region flanking the polymorphism was analyzed using RNAfold (20) and SNPfold (21) to predict the putative influence of rs10680577 on local folding structures of RERT-IncRNA.

**Real-time RT-PCR analysis**

Total RNA was isolated from tumor tissue specimens and cell lines using RNA isolation kit (Cat #74106, Qiagen). cDNA was generated using random primers and Superscript II reverse transcriptase (Cat #18064-014, Invitrogen). A SYBR ® Green real-time PCR was performed using Roche LightCycler® 480 to quantify relative RAB4B, EGLN2 and RERT-IncRNA expression in these samples. GAPDH was chosen as the internal control. Because the expression levels of EGLN2 and
RERT-lncRNA in Hep3B cells were the highest among all cell lines, the same stock Hep3B cDNA was used as calibrator (basis sample) to generate relative standard curves for both target genes and GAPDH by series of a tenfold dilution with five measuring points in triplicates. Thus, the relative calculated values can be compared across multiple plates. The initial individual concentration of EGLN2, RERT-lncRNA and GAPDH in the stock Hep3B cDNA was assigned with an arbitrary unit according to their raw C_\text{T} values in Hep3B calibrator, respectively. For all experimental samples, target genes and GAPDH quantity were determined by interpolating from their relative standard curves, respectively. Finally, the expression levels of target genes were normalized to internal control. To overcome the overlapping problem between the sequence of RERT-lncRNA transcript and that of RAB4B and EGLN2, a pair of specific primers (Inc-F and Inc-R) spanning “exon” 7 and “exon” 8 of the transcript was designed. Since the primers designed for RAB4B and EGLN2 overlapped with RERT-lncRNA (Supplemental Figure 1), the relative amount of RERT-lncRNA was subtracted from RAB4B and EGLN2 expression level for each sample to further compare their expression levels between different groups. Primer sequences used for RAB4B, EGLN2, RERT-lncRNA and GAPDH were shown in supplemental Table 1. The 25\mu l total volume final reaction mixture consisted of 1 \mu M of each primer, 12.5 \mu l of Master Mix (Applied Biosystems), and 2.5\mu l of cDNA. The negative control experiments were performed with distilled H_2O as template. PCR efficiencies were calculated with relative standard curves and the regression coefficients were above 0.98. In addition, the melting curve analysis was performed for the PCR products to
verify primer specificity. To further validate the specificity of primers for RERT-lncRNA, we analyzed 10 randomly selected RERT-lncRNA PCR products by direct sequencing.

**Western blot**

Approximately 50 µg of protein extract from HCC tumor tissues and cell lines were separated on a 12% polyacrylamide gel. Proteins were transferred to a PVDF membrane (GE Healthcare) and probed with primary antibodies against EGLN2 (Cat # sc-46024, 1:1000, Santa Cruz Biotechnology) and GAPDH (Cat # sc-48167, 1:2000, Santa Cruz Biotechnology). The primary antibodies were detected by horseradish peroxidase (HRP)-conjugated secondary antibodies (1:2500, Santa Cruz Biotechnology). Films were exposed in dark room using an enhanced chemiluminescence system (ECL, Cell Signaling Technologies). Quantification of protein levels was performed with ImageJ software.

**Construction of RERT-lncRNA-expressing plasmids and transient transfection**

For over-expression studies, the complete 2849-bp cDNA sequence of RERT-lncRNA (NR_037791) was directly synthesized by Genewiz Company (Suzhou, China) and cloned into Xhol and HindIII sites of pcDNA3.1(+) expression vector (Invitrogen). The constructed RERT-lncRNA-pcDNA3.1(+) plasmid or empty vector pcDNA3.1(+) were transfected into HepG2, SMMC-7721 and sk-Hep-1 cells cultured for 48 h in a six-well plate. The total RNA was then extracted to quantify expression levels of
Statistical Analysis

The genotype distribution was analyzed by Hardy-Weinberg equilibrium using $\chi^2$ test. Logistic regression was used to analyze the association between rs10680577 and HCC risk, adjusted by sex, age, smoking status, drinking status and HBV infection status. In addition, stratified analysis by these common confounders was performed using binary logistic regression model. Due to the limited number of del/del genotype, they were integrated with the ins/del group when appropriate. Student’s $t$ test was used to examine the differences in luciferase reporter gene expression. The normalized expression values of RAB4B, EGLN2 and RERT-lncRNA in tissue samples were compared using non-parametric Mann-Whitney $U$ test. The normalized expression levels of EGLN2 and RERT-lncRNA in cell lines were compared using one way ANOVA or student’s $t$ test. Correlations were assessed using Pearson correlation coefficient. These statistical analyses were implemented in Statistic Analysis System software (version 8.0, SAS Institute). $P<0.05$ was used as the criterion of statistical significance. All statistical tests were two sided.

Results

The associations of rs10680577 with HCC susceptibility

The baseline characteristics of the subjects in two case control studies were listed in
Table 1. No apparent differences were found between HCC patients and control subjects in terms of sex, age, smoking and drinking status, suggesting that the frequency matching was adequate. Approximately 68% of the cases and 10% of the controls were HBsAg-positive, in accordance with the fact that HBV infection was a major risk factor for HCC. Example output from sequencing and genotyping assays of rs10680577 were shown in Supplemental Figure 2. The observed genotypic frequencies for rs10680577 were in agreement with the Hardy-Weinberg equilibrium in both cases and controls ($P$>0.05). As shown in Table 2, in both case control sets and pooled analysis, compared with the ins/ins genotype, subjects with the heterozygous ins/del or homozygous del/del had a significantly increased risk of HCC under co-dominant model (adjusted OR=1.45 and 2.32; $P$<0.0001, respectively). Similar trends were observed in all other genetic models (Table 2). Each additional copy of the 4-bp deletion allele was associated with a 44% increased risk in pooled analysis (OR=1.44, 95% C.I.=1.26-1.64, $P$<0.0001).

We further determined the contribution of age, sex, smoking status, drinking status and HBV infection to the association between rs10680577 and risk of HCC. As shown in supplemental Table 2, in Suzhou case control set, the positive association was more pronounced in current smokers (OR=2.96, 95% C.I.=1.72-5.11) than non-smokers (OR=1.29, 95% C.I.=1.02-1.64) (heterogeneity $P$<0.0001). Similar trend was also validated in Shanghai case control set (heterogeneity $P$=0.004). However, other parameters did not seem to modify the positive association between the deletion allele of rs10680577 and risk of HCC.
The genotype-phenotype correlation between rs10680577 and EGLN2 expression

Since rs10680577 was located at around 1.65-kb up stream of EGLN2 and 1.75-kb down stream of RAB4B, we examined the expression of EGLN2 and RAB4B in HCC tissue samples with different genotypes. As shown in Figure 1A, results of q-PCR demonstrated that EGLN2 mRNA level (0.292 ± 0.054) in samples with ins/del and del/del genotype was significantly higher than that in the ins/ins genotype (0.065 ± 0.017). However, in terms of RAB4B, no significant difference was observed between two genotypic groups. To validate our findings, we further examined the genotype-phenotype correlations in four common hepatoma cell lines (HepG2, Hep3B, sk-Hep-1 and SMMC-7721) and L02 hepatic cell line. The EGLN2 mRNA expression level (0.333 ± 0.041) in Hep3B with del/del genotype was significantly higher than the average level of other four cell lines carrying ins/ins genotype (0.042 ± 0.007) (Figure 1B). To further investigate the correlation between rs10680577 genotype and EGLN2 protein level, 5 randomly selected HCC tumor tissues with different genotypes and aforementioned 5 cell lines were analyzed by western blot. For both liver tissues and cell lines, we found that EGLN2 protein level of del/del genotype carriers was higher than that with ins/del or ins/ins genotypes (Figure 1C). Quantification of western blot showed that EGLN2 expression of del/del genotype carriers was approximately 1.50 and 1.33 fold higher than that of ins/ins and ins/del genotype carriers in HCC tissues and cell lines, respectively (Figure 1D). Together, these data demonstrated that the deletion allele of rs10680577 was significantly
correlated with higher EGLN2 expression, at both mRNA and protein levels.

The influence of rs10680577 on EGLN2 promoter activity

rs10680577 is located at -1641bp upstream of the transcription start site of EGLN2. Therefore, the genotype-phenotype correlation may be mediated by a differential promoter polymorphism-associated regulatory mechanism. To test this hypothesis, reporter constructs containing either insertion or deletion allele of rs10680577 were generated using site-directed mutagenesis. They were used to examine the effect of this polymorphism on the promoter activity using a luciferase-based transient transfection system. However, in contrast with our predication, we found that there was no significant difference between the two constructs in any of the five cell lines tested ($P>0.05$) (supplemental Figure 3).

In-silico analysis of rs10680577 on RERT-IncRNA folding

Next, we probed other ways through which rs10680577 could control EGLN2. We noticed the presence of a 2849-bp RAB4B-EGLN2 read-through long non-coding RNA (RERT-IncRNA) overlapping EGLN2 and the upstream RAB4B gene. Long non-coding RNAs (lncRNAs) have been recently found to be pervasively transcribed in the genome and have gene-regulatory roles such as chromosome dosage-compensation, imprinting, epigenetic regulation, transcription, translation, or splicing (22). Considering the fact that rs10680577 is located within the intronic region of RERT-IncRNA, it is plausible that rs10680577 may influence
RERT-lncRNA expression by affecting its folding structures, which in turn mediates EGLN2 expression. Using RNAfold and SNPfold algorithms, we predicted the local structure changes of RERT-lncRNA caused by the 4-bp indel polymorphism. As shown in Figure 2, the “TACT” insertion appeared to disrupt the un-basepaired region surrounded by two highly basepaired regions (usually indicative of a loop, Figure 2A and 2B).

The correlation between EGLN2 and RERT-lncRNA expression

Based on the above *in-silico* analysis, we then examined the expression levels of RERT-lncRNA in both HCC tissue samples and cell lines. The sequencing results confirmed that the PCR products amplified using the specific primers were identical to the sequences of RERT-lncRNA. Intriguingly, the expression level of RERT-lncRNA in Hep3B was approximately 3 fold higher than the average expression level of all other four cell lines (Figure 3A). As shown in Figure 3B, the expression levels of EGLN2 and RERT-lncRNA were significantly correlated in HCC tissue samples ($R^2=0.540$, $P<0.001$). Furthermore, deletion allele carriers have significantly higher expression levels of RERT-lncRNA than deletion allele non-carriers in both HCC tissue samples and cell lines (Figure 3C).

**RERT-lncRNA increase EGLN2 expression *in vitro***

To determine whether RERT-lncRNA can positively regulate EGLN2 expression, we performed *in vitro* RERT-lncRNA “gain of function” analysis using a transient
transfection system. As shown in Figure 4, compared with cells transfected with empty vectors, cells transfected with constructed RERT-lncRNA vectors displayed 4.73-9.82 fold expression of RERT-lncRNA, and 1.61-1.75 fold expression of EGLN2. Therefore, over-expression of RERT-lncRNA resulted in a significant up-regulation of EGLN2 in three HCC cell lines.

**Discussion**

We presented here the first case control study evaluating the association between the 4-bp indel polymorphism within a novel long non-coding RNA and HCC susceptibility. On the basis of our current findings, we propose a schematic model illustrating the molecular mechanism and functional basis for polymorphism-associated hepatocarcinogenesis conferred by RERT-lncRNA and EGLN2 expression.

LncRNAs are an important class of pervasive genes involved in a variety of biological functions (23). Defects of lncRNAs often contribute to cancer development (24-26). It raises the question of how genetic variations in lncRNAs contribute to cancer predisposition. Several lines of evidence suggest that SNPs residing in the key regulatory regions of an RNA molecule can severely disrupt its function (21). It indicates that polymorphisms could be one of the mechanisms by which disrupted structural motifs of lncRNAs leads to diseases. The use of human genetic studies on lncRNAs would help us to understand the roles of regulatory elements and to interpret the contribution of those genetic variations to the pathogenesis of cancer (27-28).
Based on the results of our current study, we propose that insertion allele of rs10680577 may disrupt the key regulatory region of RERT-lncRNA, resulting in its misexpression, which further alters the fine turning interactions between RERT-lncRNA and EGLN2. Indeed, our in-silico studies have provided a theoretical basis for our hypothesis. More importantly, we observed strong correlations not only between EGLN2 and RERT-lncRNA expression, but also between rs10680577 and RERT-lncRNA expression. Furthermore, in vitro gain-of-function analysis revealed that RERT-lncRNA over-expression could up-regulate EGLN2 expression, which reinforced our hypothesis. LncRNA (i.e HOTAIR) has been suggested to recruit Polycomb Repressive Complex 2 (PRC2) complex to its target genes (29). We thus speculate that RERT-lncRNA may presumably serve as key co-activators and play an activating role by modulating the recruitment of general transcription factors on the promoter of its cognate gene, EGLN2. Indeed, this hypothesis is supported by a series of biochemical assays. For example, the Evf-2 lncRNA can form a complex with the homeodomain-containing protein Dlx2, which acts a transcriptional activator only when Evf-2 lncRNA is also present (30). Alternatively, RERT-lncRNA may also alter chromatin structure through its own transcription since it has been shown that more than 20% of long intergenic ncRNAs associated with chromatin modifying complexes (31). Therefore, the present study provides a novel insight into causative variations residing in lncRNAs.

Our studies indicate that deletion allele is associated (maybe in an indirect manner involving RERT-lncRNA) with higher EGLN2 expression level and would be a risk
factor for HCC carcinogenesis. Higher EGLN2 expression makes cells less sensitive to hypoxia stress, leading to less HIFα stabilization and HIF activation which are detrimental for hepatic cell survival. Consequently, necrosis results in hepatic cell lysis and inflammation which would promote HCC tumorgenesis. This notion has been reinforced by the fact that Hep3B cells carrying del/del genotype do not exhibit orderly G(1)/S arrest in response to severe hypoxia (32). In line with this evidence, it also has been demonstrated that EGLN2+/− mice have a remarkable tolerance to ischemia and a substantially reduced exercise tolerance (33). In addition, loss or silencing of EGLN2 makes hepatocytes less sensitive to acute hypoxia and protects them from ischemia/reperfusion damage (34).

In addition, our studies also indicate that the association between rs10680577 and HCC incidence is more prominent in current smokers. Consistent with the finding that chromosome 19q13 region covering EGLN2 has previously found to be associated with chronic obstructive pulmonary disease, and one SNP (rs7937) exceeds genome-wide significance (35). Moreover, a SNP (rs3733829) residing in the first intron of EGLN2 has been proved to be associated with smoking behaviors (36). It has been validated that rs10680577 has perfect linkage disequilibrium (LD) with one SNP (rs2644898) in the CHB HapMap panel (r² =1) (37). Using haploview analysis, we discover that rs2644898, rs7937 and rs3733829 are in high LD (r² >0.9) (Supplemental Figure 4). Therefore, it is very likely that the block containing these genetic variations is a susceptibility region for smoking behaviors and/or smoking related diseases. However, the real causative loci still remain to be discovered with
further functional studies.

Genome-wide association studies (GWAS) have greatly contributed to the identification of common genetic variants related to HCC. Recent GWAS studies focusing on HCC have reported several susceptibility regions such as 3p22.3, 6p21.33 and 14q32.11 (38-39). Recently, using the Affymetrix Genome-Wide Human SNP Array 5.0, Zhang and colleagues have conducted the first HCC GWAS study within chronic HBV carriers of Chinese ancestry and identified a new HBV-mediated HCC susceptibility region (40). Intriguingly, with the same ethnic group, SNPs (eg. rs2644898) that have high LD with rs10680577 has not been captured in their study. This could be explained by the fact that according to data file of Affymetrix SNP Array 5.0 used in the study by Zhang et al (40), rs2644898 was not included in the GWAS chips.

Finally, some limitations in the study should be addressed. Although we observed a strong correlation between rs10680577 and expression of RERT-lncRNA, even between expression of RERT-lncRNA and EGLN2, how genetic variability at this locus can influence RERT-lncRNA and EGLN2 expressions still need to be fully elucidated both at genetic and functional levels.

In summary, we have provided initial evidence that rs10680577 may play a functional role in regulating the expression of RERT-lncRNA and subsequently affecting the production of EGLN2 and development of HCC. Furthermore, we provide the first experimental evidence of the correlation between rs10680577 and RERT-lncRNA expression, highlighting that the genetic variations within lncRNAs may act as key
functional elements in modulating structure and expression of lncRNAs. Finally, we propose a hypothesis that EGLN2 might play a role in HCC tumorigenesis, depending on individuals’ distinct sensitivity to hypoxia caused by different genetic background. Therefore, EGLN2 may be a promising marker for personalized diagnosis and therapy of HCC.
Acknowledgments

We would like to thank Dr. Matthew Halvorsen and Dr. Alain Laederach (University of North Carolina, Chapel Hill, NC, USA) for their helpful assistance in analyzing RNA structure changes using SNPfold. We also thank Dr. Jiejie Xu (Fudan University, Shanghai, China), Dr. Zhiyong Liu (Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA, USA) and Dr. Huimin Zhang (Soochow University, Suzhou, China) for their help in preparing the manuscript. We gratefully acknowledge the participation of patients with HCC and cancer-free individuals.
References


7. Semenza GL. Regulation of oxygen homeostasis by hypoxia-inducible factor 1. Physiology (Bethesda) 2009;24:97-106.


Table 1. Demographic characteristics of HCC cases and controls recruited from Suzhou and Shanghai during 2007-2010

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Suzhou</th>
<th>Shanghai</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case (n=623)</td>
<td>Control (n=1242)</td>
<td>Case (n=444)</td>
</tr>
<tr>
<td>Age (mean±S.D.)</td>
<td>53.4±10.1</td>
<td>52.7±10.5</td>
<td>53.7±10.3</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>399(0.64)</td>
<td>795(0.64)</td>
<td>272(0.61)</td>
</tr>
<tr>
<td>Female</td>
<td>224(0.36)</td>
<td>447(0.36)</td>
<td>172(0.39)</td>
</tr>
<tr>
<td>Smoking Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-smokers</td>
<td>427(0.69)</td>
<td>893(0.72)</td>
<td>304(0.69)</td>
</tr>
<tr>
<td>Former smokers</td>
<td>107(0.17)</td>
<td>206(0.17)</td>
<td>76(0.17)</td>
</tr>
<tr>
<td>Current smokers</td>
<td>89(0.14)</td>
<td>143(0.12)</td>
<td>64(0.14)</td>
</tr>
<tr>
<td>Drinking status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-drinkers</td>
<td>361(0.58)</td>
<td>671(0.54)</td>
<td>257(0.58)</td>
</tr>
<tr>
<td>Light drinkers</td>
<td>169(0.27)</td>
<td>372(0.30)</td>
<td>122(0.27)</td>
</tr>
<tr>
<td>Heavy drinkers</td>
<td>93(0.15)</td>
<td>199(0.16)</td>
<td>65(0.15)</td>
</tr>
<tr>
<td>HBsAg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>423(0.68)</td>
<td>119(0.10)</td>
<td>304(0.68)</td>
</tr>
<tr>
<td>Negative</td>
<td>200(0.32)</td>
<td>1123(0.90)</td>
<td>140(0.32)</td>
</tr>
<tr>
<td>Genetic Model</td>
<td>Genotype</td>
<td>Cases (%)</td>
<td>Control (%)</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Suzhou</td>
<td>ins/ins</td>
<td>359</td>
<td>826</td>
</tr>
<tr>
<td></td>
<td>ins/del</td>
<td>235</td>
<td>384</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>Dominant model</td>
<td>ins/ins</td>
<td>359</td>
<td>826</td>
</tr>
<tr>
<td></td>
<td>ins/del + del/del</td>
<td>264</td>
<td>416</td>
</tr>
<tr>
<td>Recessive model</td>
<td>ins/ins + ins/del</td>
<td>594</td>
<td>1210</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>Additive model</td>
<td>ins allele</td>
<td>954</td>
<td>2036</td>
</tr>
<tr>
<td></td>
<td>del allele</td>
<td>293</td>
<td>448</td>
</tr>
<tr>
<td>Shanghai</td>
<td>ins/ins</td>
<td>248</td>
<td>299</td>
</tr>
<tr>
<td></td>
<td>ins/del</td>
<td>171</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>Dominant model</td>
<td>ins/ins</td>
<td>248</td>
<td>299</td>
</tr>
<tr>
<td></td>
<td>ins/del + del/del</td>
<td>196</td>
<td>151</td>
</tr>
<tr>
<td>Recessive model</td>
<td>ins/ins + ins/del</td>
<td>419</td>
<td>437</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>Additive model</td>
<td>ins allele</td>
<td>667</td>
<td>736</td>
</tr>
<tr>
<td></td>
<td>del allele</td>
<td>221</td>
<td>164</td>
</tr>
<tr>
<td>Overall</td>
<td>ins/ins</td>
<td>607</td>
<td>1125</td>
</tr>
<tr>
<td></td>
<td>ins/del</td>
<td>406</td>
<td>522</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>54</td>
<td>45</td>
</tr>
<tr>
<td>Dominant model</td>
<td>ins/ins</td>
<td>607</td>
<td>1125</td>
</tr>
<tr>
<td></td>
<td>ins/del + del/del</td>
<td>460</td>
<td>567</td>
</tr>
<tr>
<td>Recessive model</td>
<td>ins/ins + ins/del</td>
<td>1013</td>
<td>1647</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>54</td>
<td>45</td>
</tr>
<tr>
<td>Additive model</td>
<td>ins allele</td>
<td>1620</td>
<td>2772</td>
</tr>
<tr>
<td></td>
<td>del allele</td>
<td>514</td>
<td>612</td>
</tr>
</tbody>
</table>

<sup>a</sup> Adjusted for age.  <sup>b</sup> Adjusted for sex, age, smoking status, drinking status and HBV infection.
Figure Legends

Figure 1. Expression of EGLN2 in HCC tissues and cell lines with different genotypes.

(A) EGLN2 and RAB4B mRNA expression (mean ± SEM.) in HCC tissue samples, by rs10680577 genotype. (B) EGLN2 mRNA expression (mean ± SD) in cell lines with different rs10680577 genotypes. * indicates $P<0.0001$. (C) Western blot analysis of EGLN2 in different genotypic HCC tissues (upper panel) and cell lines (lower panel). (D) Quantification (mean ± SD) of immunoblotted EGLN2 from three independent experiments.

Figure 2. Influence of rs10680577 on RERT-IncRNA local folding structures.

The local structure changes were illustrated by RNAfold (A) and SNPfold (B), respectively. Arrow (in A) indicates the position of “TACT” deletion. Black line (in B) represents deletion allele sequences whereas red line represents insertion allele sequences.

Figure 3. Correlations between EGLN2, RERT-IncRNA and rs10680577.

(A) EGLN2 and RERT-IncRNA mRNA expression levels in five cell lines. * indicate $P<0.001$, compared with the average level of other four cell lines. (B) Correlation between relative mRNA expression of EGLN2 and RERT-IncRNA in 72 HCC tumor tissues. (C) The differential expression of RERT-IncRNA in “deletion” allele carriers and non-carriers. Significant differences were observed in both HCC tissue samples
(n=72) and five cell lines.

**Figure 4. RERT-lncRNA over-expression leads to EGLN2 up-regulation.**

The constructed RERT-lncRNA-pcDNA3.1 (+) vector as well as pcDNA3.1 (+) empty vectors were transiently transfected into HepG2, SMMC-7721 and sk-Hep-1 cells. Real-time RT-PCR was used to quantify RERT-lncRNA and EGLN2 expression levels after transfection. * P<0.0001, levels of RERT-lncRNA in cells transfected with constructed RERT-lncRNA vectors *versus* cells transfected with empty vectors; △ P<0.01, level of EGLN2 in cells transfected with constructed RERT-lncRNA vectors *versus* cells that were transfected with empty vectors.