

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

Zhansheng Zhu<sup>1</sup>, Xueren Gao<sup>1</sup>, Yan He<sup>1</sup>, Hua Zhao<sup>3</sup>, Qiang Yu<sup>4</sup>, Deke Jiang<sup>5</sup>, Pingzhao Zhang<sup>5,6</sup>, Xiaopin Ma<sup>5</sup>, Huixing Huang<sup>5</sup>, Dong Dong<sup>1</sup>, Jiao Wan<sup>1</sup>, Zhenyong Gu<sup>1</sup>, Xinghong Jiang<sup>2</sup>, Long Yu<sup>5</sup>, and Yuzhen Gao<sup>1</sup>

### Abstract

The Prolyl hydroxylase 1 (EGLN2) is known to affect tumorigenesis by regulating the degradation of hypoxia-inducible factor. 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 1,242 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% confidence interval (CI) = 1.18–1.66,  $P < 0.0001$ ; OR = 1.49; 95% CI = 1.18–1.88,  $P = 0.0007$ ]. Such positive association was more pronounced in current smokers (OR = 3.49, 95% CI = 2.24–5.45) than nonsmokers (OR = 1.24, 95% CI = 1.03–1.50; heterogeneity  $P = 0.0002$ ). Genotype-phenotype correlation studies showed that the deletion allele was significantly correlated with higher expression of both *EGLN2* and RERT-lncRNA [a long noncoding RNA whose sequence overlaps with Ras-related GTP-binding protein 4b (*RAB4B*) and *EGLN2*] *in vivo* and *in vitro*. Furthermore, RERT-lncRNA expression was also significantly correlated with *EGLN2* expression *in vivo*, consistent with *in vitro* gain-of-function study that showed overexpressing RERT-lncRNA upregulated *EGLN2*. Finally, *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. *Cancer Res*; 72(23); 1–10. ©2012 AACR.

### 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 more than 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 O<sub>2</sub> homeostasis (9). HIF regulates a program of gene expression that facilitates cell survival under hypoxic conditions through cell-intrinsic

**Authors' Affiliations:** <sup>1</sup>Department of Forensic Medicine, <sup>2</sup>Department of Neurobiology and Psychology, Key Laboratory of Pain Research & Therapy, Medical College of Soochow University; <sup>3</sup>Department of General Surgery, the First Affiliated Hospital of Soochow University; <sup>4</sup>Department of Gastroenterology, Suzhou Municipal Hospital, Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, Jiangsu, China; <sup>5</sup>The State Key Laboratory of Genetic Engineering, and <sup>6</sup>Institutes of Biomedical Sciences, Fudan University, Shanghai, China

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Z. Zhu, X. Gao, and Y. He contributed equally to this work.

**Corresponding Authors:** Yuzhen Gao, Department of Forensic Medicine, Medical College of Soochow University, Suzhou, 215123, Jiangsu, 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, China. Phone: 86-21-65643954; Fax: 86-21-65643250; E-mail: longyu@fudan.edu.cn

doi: 10.1158/0008-5472.CAN-12-0010

©2012 American Association for Cancer Research.

changes in metabolism and cell-extrinsic changes affecting oxygen delivery. Prolyl hydroxylases (PHD1, also known as EGLN2) is 1 of the 3 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 2 independent case–control sets including 1,067 newly diagnosed incident HCC cases and 1,692 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). In brief, the diagnosis of patients was validated by combinations of pathologic examination and positive imaging (i.e., MRI and/or computerized tomography). A total of 1,692 controls, 1,242 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 1 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 1 year before the interview; the nonsmokers 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 1 year. Those who consumed more than 2 alcohol drinks per week for more than 1 year were categorized as heavy drinkers. Nondrinkers 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 pathologic 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 and International Union Against Cancer standard. After surgical resection, fresh tissues were immediately stored at  $-80^{\circ}\text{C}$  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 Supplementary Table S1. Genotyping was conducted 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 2 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  $-1,724$  to  $+46$  relative to the transcriptional start site of *EGLN2* was amplified with primers clone-F and clone-R as listed in Supplementary Table S1 from a homozygous human genomic DNA sample containing a 4-bp "TACT" (four-basepairs DNA sequence that represents the insertion/deletion sequence of rs10680577 polymorphism) 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 Lightning Site-Directed Mutagenesis Kit (Cat # 210518, Stratagene) with mutagenesis primers (mutagenesis-F and mutagenesis-R as listed in Supplementary Table S1). 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 supplemented with 10% FBS and 1% penicillin-streptomycin at  $37^{\circ}\text{C}$  in a humidified 5%  $\text{CO}_2$  incubator. The cell lines were characterized by DNA fingerprinting analysis using short tandem repeat markers. All cell lines were placed under cryostage after removal from the bank and used within 3 months after thawing fresh vials. Cells were seeded at a density

of  $1 \times 10^5$  cells per well in 24-well plates (BD Biosciences). Note that 16 hours after plating, cells were transfected by Lipofectamin 2000 (Cat #11668-019, Invitrogen) according to manufacturer's protocol. In each well, 500 ng pGL3-basic vector containing either TACT insertion or deletion allele and 50 ng pRL-TK vector (Promega) were cotransfected into cells. The empty pGL3-basic vector was added as negative control. Note that 24 hours after transfection, cells were harvested immediately after addition of 100  $\mu$ L 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 conducted and each experiment was repeated at least 3 times.

### **In silico prediction of rs10680577 on RERT-lncRNA folding structures**

As certain conserved structures more likely serve important biologic 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-lncRNA.

### **Real-time 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 (RT-PCR) was conducted using Roche LightCycler 480 to quantify relative RAB4B, EGLN2, and RERT-lncRNA expression in these samples. Glyceraldehyde-3-phosphate dehydrogenase (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 10-fold dilution with 5 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  $C_T$  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 (lnc-F and lnc-R) spanning exon 7 and exon 8 of the transcript was designed. Because the primers designed for RAB4B and EGLN2 overlapped with RERT-lncRNA (Supplementary Fig. S1), 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 Supplementary Table

S1. The 25  $\mu$ L total volume final reaction mixture consisted of 1  $\mu$ mol/L of each primer, 12.5  $\mu$ L of Master Mix (Applied Biosystems), and 2.5  $\mu$ L of cDNA. The negative control experiments were conducted with distilled H<sub>2</sub>O 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 conducted 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 blotting**

Approximately 50  $\mu$ g of protein extract from HCC tumor tissues and cell lines were separated on a 12% polyacrylamide gel. Proteins were transferred to a polyvinylidene difluoride 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-conjugated secondary antibodies (1:2500, Santa Cruz Biotechnology). Films were exposed in dark room using an enhanced chemiluminescence system (Cell Signaling Technologies). Quantification of protein levels was conducted with ImageJ software.

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

For overexpression studies, the complete 2849-bp cDNA sequence of RERT-lncRNA (NR\_037791) was directly synthesized by Genewiz Company and cloned into XhoI 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 hours in a 6-well plate. The total RNA was then extracted to quantify expression levels of RERT-lncRNA and EGLN2. Six replicates were conducted for each group each time, and each experiment was repeated at least 3 times.

### **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 hepatitis B virus (HBV) infection status. In addition, stratified analysis by these common confounders was conducted using binary logistic regression model. Because of the limited number of del/del genotype, they were integrated with the ins/del group when appropriate. Student *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 nonparametric Mann-Whitney U test. The normalized expression levels of EGLN2 and RERT-lncRNA in cell lines were compared using 1-way ANOVA or student *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 2-sided.

## Results

### The associations of rs10680577 with HCC susceptibility

The baseline characteristics of the subjects in 2 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 Supplementary Fig. S2. 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 codominant 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% confidence interval (CI) = 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 Supplementary Table S2, in Suzhou case-control set, the positive association was more pronounced in current smokers (OR = 2.96, 95% CI = 1.72–5.11) than nonsmokers (OR = 1.29, 95% CI =

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

Because rs10680577 was located at around 1.65-kb upstream of *EGLN2* and 1.75-kb downstream of *RAB4B*, we examined the expression of *EGLN2* and *RAB4B* in HCC tissue samples with different genotypes. As shown in Fig. 1A, results of q-PCR showed that *EGLN2* mRNA level ( $0.292 \pm 0.054$ ) in samples with ins/del and del/del genotype was significantly higher than that in the ins/ins genotype ( $0.065 \pm 0.017$ ). However, in terms of *RAB4B*, no significant difference was observed between 2 genotypic groups. To validate our findings, we further examined the genotype-phenotype correlations in 4 common hepatoma cell lines (HepG2, Hep3B, sk-Hep-1, and SMMC-7721) and L02 hepatic cell line. The *EGLN2* mRNA expression level ( $0.333 \pm 0.041$ ) in Hep3B with del/del genotype was significantly higher than the average level of other 4 cell lines carrying ins/ins genotype ( $0.042 \pm 0.007$ ; Fig. 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 blotting. 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 (Fig. 1C). Quantification of Western blotting 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

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

Characteristics	Suzhou		Shanghai		Overall	
	Case (n = 623)	Control (n = 1,242)	Case (n = 444)	Control (n = 450)	Case (n = 1067)	Control (n = 1,692)
Age (mean $\pm$ SD)	53.4 $\pm$ 10.1	52.7 $\pm$ 10.5	53.7 $\pm$ 10.3	53.1 $\pm$ 9.8	53.5 $\pm$ 10.2	52.8 $\pm$ 10.1
Gender						
Male	399 (0.64)	795 (0.64)	272 (0.61)	285 (0.63)	671 (0.63)	1,080 (0.64)
Female	224 (0.36)	447 (0.36)	172 (0.39)	165 (0.37)	396 (0.37)	612 (0.36)
Smoking status						
Nonsmokers	427 (0.69)	893 (0.72)	304 (0.69)	324 (0.72)	731 (0.69)	1,217 (0.72)
Former smokers	107 (0.17)	206 (0.17)	76 (0.17)	76 (0.17)	183 (0.17)	282 (0.17)
Current smokers	89 (0.14)	143 (0.12)	64 (0.14)	50 (0.11)	153 (0.14)	193 (0.11)
Drinking status						
Nondrinkers	361 (0.58)	671 (0.54)	257 (0.58)	243 (0.54)	618 (0.58)	914 (0.54)
Light drinkers	169 (0.27)	372 (0.30)	122 (0.27)	135 (0.30)	291 (0.27)	507 (0.30)
Heavy drinkers	93 (0.15)	199 (0.16)	65 (0.15)	72 (0.16)	158 (0.15)	271 (0.16)
HBsAg						
Positive	423 (0.68)	119 (0.10)	304 (0.68)	41 (0.09)	727 (0.68)	160 (0.10)
Negative	200 (0.32)	1,123 (0.90)	140 (0.32)	409 (0.91)	340 (0.32)	1,532 (0.90)

**Table 2.** Associations between rs10680577 and HCC susceptibility in Suzhou and Shanghai case-control sets recruited during 2007 to 2010

	Genetic model	Genotype	Cases(%)	Control(%)	OR (95% CI) <sup>a</sup>	OR (95% CI) <sup>b</sup>
Suzhou	Codominant model	ins/ins	359 (57.6)	826 (66.5)	1.00 (Reference)	1.00 (Reference)
		ins/del	235 (37.7)	384 (30.9)	1.42 (1.15–1.75)	1.43 (1.16–1.76)
		del/del	29 (4.7)	32 (2.6)	2.09 (1.21–3.61)	2.30 (1.34–3.96)
	Dominant model	ins/ins	359 (57.6)	826 (66.5)	1.00 (Reference)	1.00 (Reference)
		ins/del + del/del	264 (42.4)	416 (33.5)	1.47 (1.20–1.80)	1.49 (1.22–1.83)
	Recessive model	ins/ins + ins/del	594 (95.3)	1,210 (97.4)	1.00 (Reference)	1.00 (Reference)
		del/del	29 (4.7)	32 (2.6)	1.85 (1.07–3.17)	1.84 (1.07–3.16)
	Additive model	ins allele	954 (76.5)	2,036 (82.0)	1.00 (Reference)	1.00 (Reference)
		del allele	293 (23.5)	448 (18.0)	1.40 (1.18–1.66)	1.40 (1.18–1.66)
	Shanghai	Codominant model	ins/ins	248 (55.9)	299 (66.4)	1.00 (Reference)
ins/del			171 (38.5)	138 (30.7)	1.51 (1.13–2.02)	1.53 (1.14–2.00)
del/del			25 (5.6)	13 (2.9)	2.33 (1.11–4.92)	2.61 (1.29–5.28)
Dominant model		ins/ins	248 (55.9)	299 (66.4)	1.00 (Reference)	1.00 (Reference)
		ins/del + del/del	196 (44.1)	151 (33.6)	1.58 (1.19–2.09)	1.58 (1.21–2.08)
Recessive model		ins/ins + ins/del	419 (94.4)	437 (97.1)	1.00 (Reference)	1.00 (Reference)
		del/del	25 (5.6)	13 (2.9)	2.01 (0.97–4.20)	2.00 (1.01–3.95)
Additive model		ins allele	667 (75.1)	736 (81.8)	1.00 (Reference)	1.00 (Reference)
		del allele	221 (24.9)	164 (18.2)	1.49 (1.18–1.88)	1.49 (1.18–1.88)
Overall		Codominant model	ins/ins	607 (56.9)	1,125 (66.5)	1.00 (Reference)
	ins/del		406 (38.0)	522 (30.8)	1.45 (1.23–1.72)	1.45 (1.23–1.70)
	del/del		54 (5.1)	45 (2.7)	2.23 (1.46–3.42)	2.32 (1.54–3.49)
	Dominant model	ins/ins	607 (56.9)	1,125 (66.5)	1.00 (Reference)	1.00 (Reference)
		ins/del + del/del	460 (43.1)	567 (33.5)	1.52 (1.29–1.78)	1.65 (1.40–1.93)
	Recessive model	ins/ins + ins/del	1,013 (94.9)	1,647 (97.3)	1.00 (Reference)	1.00 (Reference)
		del/del	54 (5.1)	45 (2.7)	1.95 (1.28–2.98)	1.95 (1.30–2.92)
	Additive model	ins allele	1,620 (75.9)	2,772 (81.9)	1.00 (Reference)	1.00 (Reference)
		del allele	514 (24.1)	612 (18.1)	1.44 (1.26–1.64)	1.44 (1.26–1.64)

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

(Fig. 1D). Together, these data showed 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 –1,641 bp 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 2 constructs in any of the 5 cell lines tested ( $P > 0.05$ ; Supplementary Fig. S3).

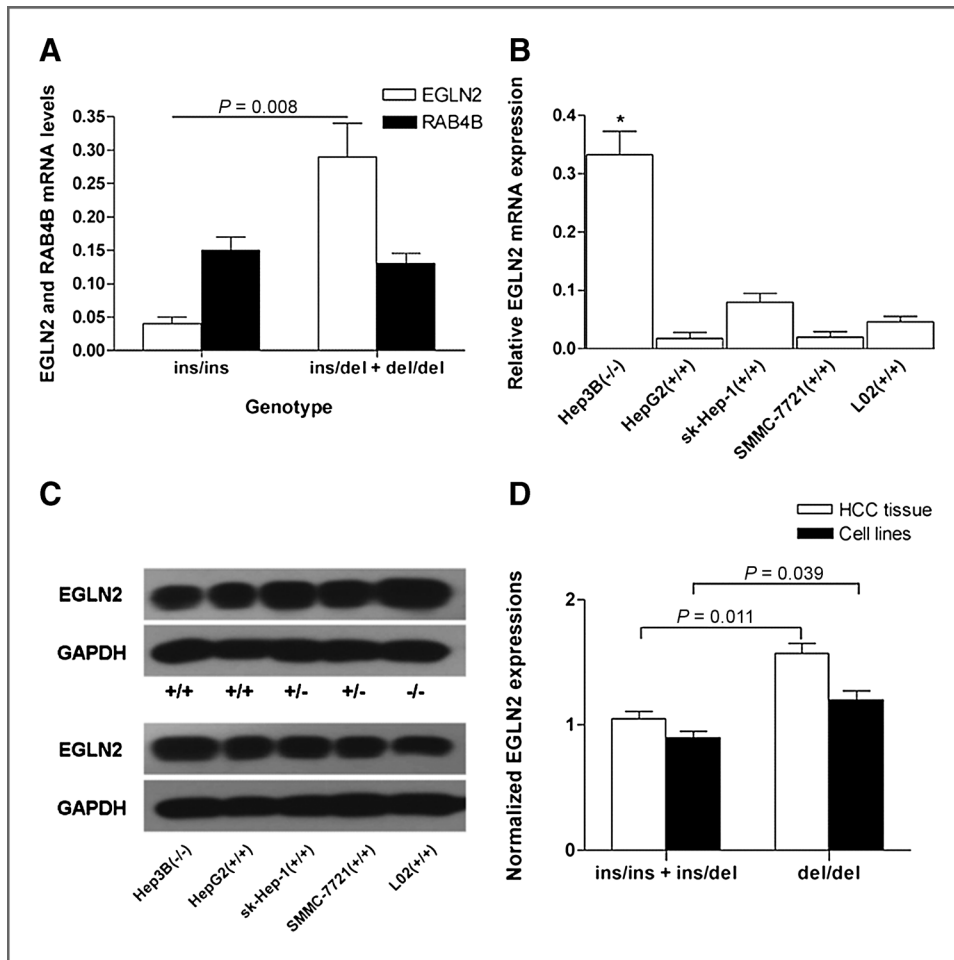
#### In silico analysis of rs10680577 on RERT-lncRNA folding

Next, we probed other ways through which rs10680577 could control *EGLN2*. We noted the presence of a 2849-bp RAB4B-

*EGLN2* read-through long noncoding RNA (RERT-lncRNA) overlapping *EGLN2* and the upstream *RAB4B* gene. Long noncoding RNAs (lncRNA) 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-lncRNA, 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 Fig. 2, the TACT insertion appeared to disrupt the unbase paired region surrounded by 2 highly base paired regions (usually indicative of a loop, Fig. 2A and B).

#### 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



**Figure 1.** Expression of EGLN2 in HCC tissues and cell lines with different genotypes. A, EGLN2 and RAB4B mRNA expression (mean  $\pm$  SEM) in HCC tissue samples by rs10680577 genotype. B, EGLN2 mRNA expression (mean  $\pm$  SD) in cell lines with different rs10680577 genotypes. \*,  $P < 0.0001$ . C, Western blot analysis of EGLN2 in different genotypic HCC tissues (top) and cell lines (bottom). D, quantification (mean  $\pm$  SD) of immunoblotted EGLN2 from 3 independent experiments.

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 4 cell lines (Fig. 3A). As shown in Fig. 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 noncarriers in both HCC tissue samples and cell lines (Fig. 3C).

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

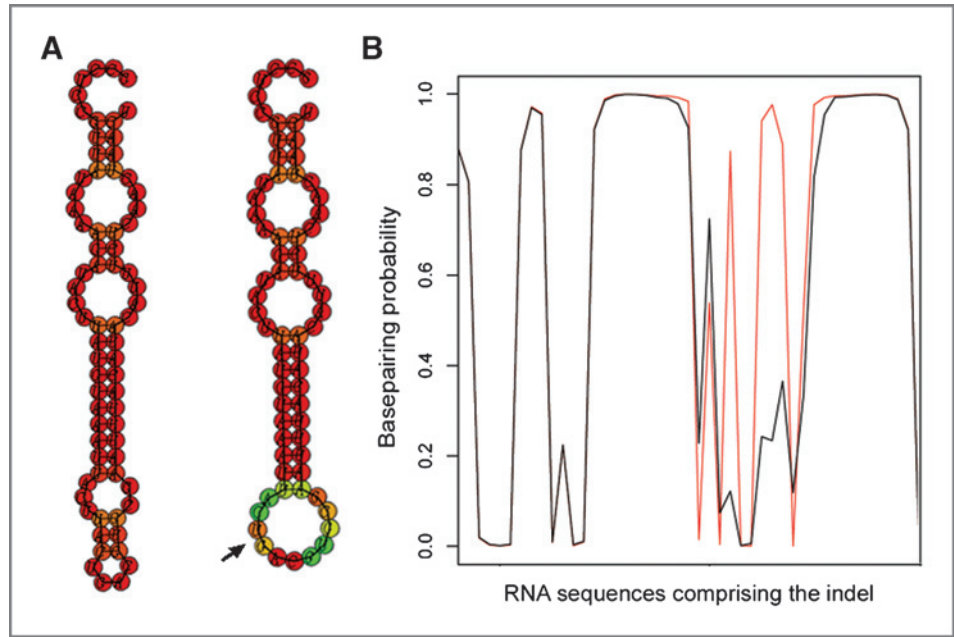
To determine whether RERT-lncRNA can positively regulate EGLN2 expression, we conducted *in vitro* RERT-lncRNA gain-of-function analysis using a transient transfection system. As shown in Fig. 4, compared with cells transfected with empty vectors, cells transfected with constructed RERT-lncRNA vectors displayed 4.73- to 9.82-fold expression of RERT-lncRNA, and 1.61- to 1.75-fold expression of EGLN2. Therefore, overexpression of RERT-lncRNA resulted in a significant upregulation of EGLN2 in 3 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 noncoding 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 biologic 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 single-nucleotide polymorphisms (SNP) 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 tuning

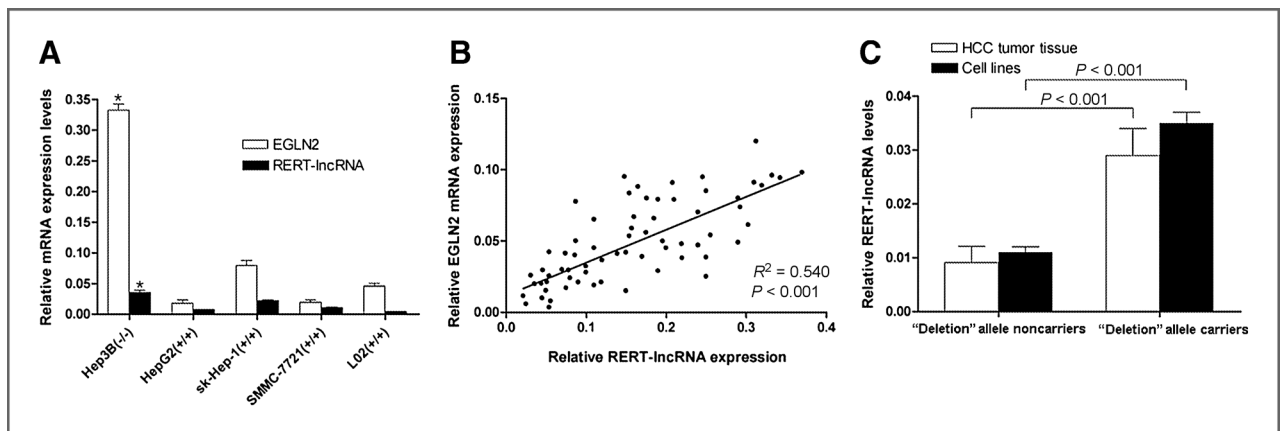
**Figure 2.** Influence of rs10680577 on RERT-lncRNA local folding structures. The local structure changes were illustrated by RNAfold (A) and SNPfold (B), respectively. Arrow (A) indicates the position of TACT deletion. Black line (B) represents deletion allele sequences, whereas red line represents insertion allele sequences.



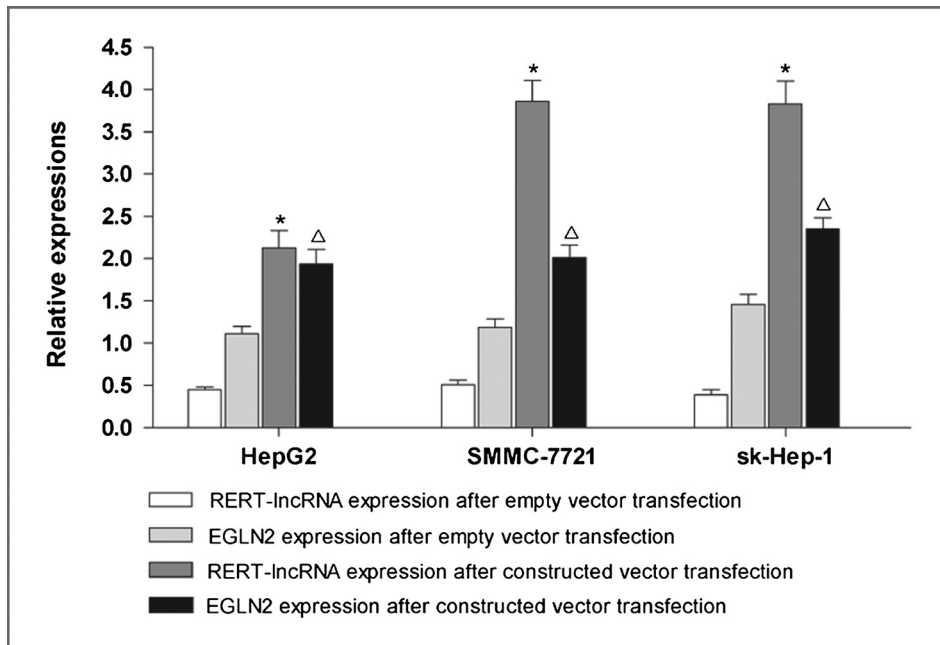
interactions between RERT-lncRNA and EGLN2. Indeed, our *in silico* studies have provided a theoretical basis for our hypothesis. More importantly, we observed strong correlations not only between EGLN2 and RERT-lncRNA expression, but also between rs10680577 and RERT-lncRNA expression. Furthermore, *in vitro* gain-of-function analysis revealed that RERT-lncRNA overexpression could upregulate EGLN2 expression, which reinforced our hypothesis. LncRNA (i.e., HOTAIR) has been suggested to recruit Polycomb Repressive Complex 2 complex to its target genes (29). We thus speculate that RERT-lncRNA may presumably serve as key coactivators 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 because 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 HIF $\alpha$  stabilization and HIF activation, which are detrimental for hepatic cell survival. Consequently, necrosis results in hepatic cell lysis



**Figure 3.** Correlations between EGLN2, RERT-lncRNA, and rs10680577. A, EGLN2 and RERT-lncRNA mRNA expression levels in 5 cell lines. \*,  $P < 0.001$ , compared with the average level of other 4 cell lines. B, correlation between relative mRNA expression of EGLN2 and RERT-lncRNA in 72 HCC tumor tissues. C, the differential expression of RERT-lncRNA in deletion allele carriers and noncarriers. Significant differences were observed in both HCC tissue samples ( $n = 72$ ) and 5 cell lines.



**Figure 4.** RERT-lncRNA overexpression leads to EGLN2 upregulation. 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. 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;  $\Delta$ ,  $P < 0.01$ , level of EGLN2 in cells transfected with constructed RERT-lncRNA vectors versus cells that were transfected with empty vectors.

and inflammation, which would promote HCC tumorigenesis. This notion has been reinforced by the fact that Hep3B cells carrying del/del genotype do not exhibit orderly G<sub>1</sub>/S arrest in response to severe hypoxia (32). In line with this evidence, it also has been shown that EGLN2<sup>-/-</sup> 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 ( $r^2 = 1$ ; ref. 37). Using haplotype analysis, we discover that rs2644898, rs7937, and rs3733829 are in high LD ( $r^2 > 0.9$ ; Supplementary Fig. S4). 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 (e.g., 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 and colleagues (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 for HCC.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Authors' Contributions

**Conception and design:** Z. Zhu, X. Gao, Y. He, Y. Gao

**Development of methodology:** Z. Zhu, X. Gao, Y. He, Y. Gao

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** Z. Zhu, X. Gao, Y. He, H. Zhao, Q. Yu, D. Jiang, P. Zhang, X. Ma, H. Huang, D. Dong, J. Wan, L. Yu, Y. Gao



**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** Z. Zhu, X. Gao, Y. He, Q. Yu, D. Jiang, Y. Gao  
**Writing, review, and/or revision of the manuscript:** Z. Zhu, X. Gao, Y. He, Y. Gao

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** H. Zhao, Z. Gu, X. Jiang, L. Yu, Y. Gao  
**Study supervision:** Y. Gao

## Acknowledgments

The authors would like to thank Dr. Matthew Halvorsen and Dr. Alain Laederach (University of North Carolina, Chapel Hill, NC) for their helpful assistance in analyzing RNA structure changes using SNFold. The authors also thank Dr. Jiejie Xu (Fudan University, Shanghai, China), Dr. Zhiyong Liu (Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA), and

Dr. Huimin Zhang (Soochow University, Suzhou, China) for their help in preparing the manuscript. The authors gratefully acknowledge the participation of patients with HCC and cancer-free individuals.

## Grant Support

This study is supported by grants from Natural Science Foundation of China (30800621 and 81171893 to Y. Gao, 81201574 to Y. He).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 3, 2012; revised September 10, 2012; accepted September 21, 2012; published OnlineFirst October 1, 2012.

## References

- Yang JD, Roberts LR. Hepatocellular carcinoma: a global view. *Nat Rev Gastroenterol Hepatol* 2010;7:448–58.
- McClune AC, Tong MJ. Chronic hepatitis B and hepatocellular carcinoma. *Clin Liver Dis* 2010;14:461–76.
- Chen CJ, Chen DS. Interaction of hepatitis B virus, chemical carcinogen, and genetic susceptibility: multistage hepatocarcinogenesis with multifactorial etiology. *Hepatology* 2002;36:1046–9.
- Whittaker S, Marais R, Zhu AX. The role of signaling pathways in the development and treatment of hepatocellular carcinoma. *Oncogene* 2010;29:4989–5005.
- Aravalli RN, Steer CJ, Cressman EN. Molecular mechanisms of hepatocellular carcinoma. *Hepatology* 2008;48:2047–63.
- El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 2012;142:1264–73.
- Semenza GL. Regulation of oxygen homeostasis by hypoxia-inducible factor 1. *Physiology (Bethesda)* 2009;24:97–106.
- Brahimi-Horn MC, Chiche J, Pouyssegur J. Hypoxia and cancer. *J Mol Med (Berl)* 2007;85:1301–7.
- Wilson WR, Hay MP. Targeting hypoxia in cancer therapy. *Nat Rev Cancer* 2011;11:393–410.
- Kaelin WJ, Ratcliffe PJ. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol Cell* 2008;30:393–402.
- Hsu YC, Yuan S, Chen HY, Yu SL, Liu CH, Hsu PY, et al. A four-gene signature from NCI-60 cell line for survival prediction in non-small cell lung cancer. *Clin Cancer Res* 2009;15:7309–15.
- Andersen S, Donnem T, Stenvold H, Al-Saad S, Al-Shibli K, Busund LT, et al. Overexpression of the HIF hydroxylases PHD1, PHD2, PHD3 and FIH are individually and collectively unfavorable prognosticators for NSCLC survival. *PLoS One* 2011;6:e23847.
- Erez N, Milyavsky M, Eilam R, Shats I, Goldfinger N, Rotter V. Expression of prolyl-hydroxylase-1 (PHD1/EGLN2) suppresses hypoxia inducible factor-1 $\alpha$  activation and inhibits tumor growth. *Cancer Res* 2003;63:8777–83.
- Fox SB, Generali D, Berruti A, Brizzi MP, Campo L, Bonardi S, et al. The prolyl hydroxylase enzymes are positively associated with hypoxia-inducible factor-1 $\alpha$  and vascular endothelial growth factor in human breast cancer and alter in response to primary systemic treatment with epirubicin and tamoxifen. *Breast Cancer Res* 2011;13:R16.
- Jokilehto T, Jaakkola PM. The role of HIF prolyl hydroxylases in tumour growth. *J Cell Mol Med* 2010;14:758–70.
- Calvisi DF, Ladu S, Gorden A, Farina M, Lee JS, Conner EA, et al. Mechanistic and prognostic significance of aberrant methylation in the molecular pathogenesis of human hepatocellular carcinoma. *J Clin Invest* 2007;117:2713–22.
- Gao Y, He Y, Ding J, Wu K, Hu B, Liu Y, et al. An insertion/deletion polymorphism at miRNA-122-binding site in the interleukin-1 $\alpha$  3' untranslated region confers risk for hepatocellular carcinoma. *Carcinogenesis* 2009;30:2064–9.
- Chen S, He Y, Ding J, Jiang Y, Jia S, Xia W, et al. An insertion/deletion polymorphism in the 3' untranslated region of beta-transducin repeat-containing protein (betaTrCP) is associated with susceptibility for hepatocellular carcinoma in Chinese. *Biochem Biophys Res Commun* 2010;391:552–6.
- Jiang Y, Chen S, Jia S, Zhu Z, Gao X, Dong D, et al. Association of HLA-G 3' UTR 14-bp insertion/deletion polymorphism with hepatocellular carcinoma susceptibility in a Chinese population. *DNA Cell Biol* 2011;30:1027–32.
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA website. *Nucleic Acids Res* 2008;36:W70–4.
- Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet* 2010;6:e1001074.
- Hung T, Chang HY. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol* 2010;7:582–5.
- Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011;43:904–14.
- Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 2011;10:38.
- Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res* 2011;71:3–7.
- Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* 2011;71:6320–6.
- Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol* 2011;21:354–61.
- Mattick JS. The genetic signatures of noncoding RNAs. *PLoS Genet* 2009;5:e1000459.
- Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010;329:689–93.
- Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Efv-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev* 2006;20:1470–84.
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 2009;106:11667–72.
- Box AH, Demetrick DJ. Cell cycle kinase inhibitor expression and hypoxia-induced cell cycle arrest in human cancer cell lines. *Carcinogenesis* 2004;25:2325–35.
- Aragónés J, Schneider M, Van Geyte K, Fraisl P, Dresselaers T, Mazzone M, et al. Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. *Nat Genet* 2008;40:170–80.
- Schneider M, Van Geyte K, Fraisl P, Kiss J, Aragónés J, Mazzone M, et al. Loss or silencing of the PHD1 prolyl hydroxylase protects livers of mice against ischemia/reperfusion injury. *Gastroenterology* 2010;138:1143–54.
- Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* 2012;21:947–57.
- Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;42:441–7.

37. Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, et al. Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res* 2011;21:830–9.
38. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* 2011;43:455–8.
39. Clifford RJ, Zhang J, Meerzaman DM, Lyu MS, Hu Y, Cultraro CM, et al. Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma. *Hepatology* 2010;52:2034–43.
40. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, Jia W, et al. Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. *Nat Genet* 2010;42:755–8.