Carcinogenesis vol.35 no.9 pp.2062–2067, 2014
doi:10.1093/carcin/bgu103
Advance Access publication April 30, 2014

The identification of an ESCC susceptibility SNP rs920778 that regulates the expression of lncRNA *HOTAIR* via a novel intronic enhancer

Xiaojiao Zhang[†], Liqing Zhou[1,†], Guobin Fu[2,†], Fang Sun[3], Juan Shi, Jinyu Wei, Chao Lu, Changchun Zhou[4], Qipeng Yuan and Ming Yang*

State Key Laboratory of Chemical Resource Engineering, College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China, [1]Department of Radiation Oncology, Huaian No. 2 Hospital, Huaian, Jiangsu Province, China, [2]Department of Oncology, Provincial Hospital affiliated to Shandong University, Jinan, Shandong Province, 250021, China, [3]Biochemistry Department of Bethune Military Medical College, Shijiazhuang, Hebei Province, 223002, China and [4]Clinical Laboratory, Shandong Cancer Hospital, Shandong Academy of Medical Sciences, Jinan, Shandong Province, 250117, China

*To whom correspondence should be addressed. College of Life Science and Technology, Beijing University of Chemical Technology, PO Box 53, Beijing 100029, China. Tel: +86 10 64447747; Fax: +86 10 64437610; Email: yangm@mail.buct.edu.cn.

Long noncoding RNA (lncRNA) HOX transcript antisense RNA (HOTAIR), which could induce genome-wide retargeting of polycomb-repressive complex 2, trimethylates histone H3 lysine-27 (H3K27me3) and deregulation of multiple downstream genes, is involved in development and progression of esophageal squamous cell carcinoma (ESCC). We hypothesized that the functional single nucleotide polymorphisms (SNP) in *HOTAIR* may affect HOTAIR expression and/or its function and, thus, ESCC risk. Therefore, we examined the association between three haplotype-tagging SNPs (htSNP) across the whole *HOTAIR* locus and ESCC risk as well as the functional relevance of an ESCC susceptibility SNP rs920778. Genotypes were determined in three independent case–control sets consisted of 2098 ESCC patients and 2150 controls. The allele-specific regulation on *HOTAIR* expression by the rs920778 SNP was investigated *in vitro* and *in vivo*. We found that the *HOTAIR* rs920778 TT carriers had a 1.37-fold, 1.78-fold and 2.08-fold increased ESCC risk in Jinan, Shijiazhuang and Huaian populations, respectively, compared with the CC carriers (P = 0.003, 7.7×10^{-4} and 5.9×10^{-4}). During inspecting functional relevance of the rs920778 SNP, we identified a novel intronic *HOTAIR* enhancer locating between +1719bp and +2353bp from the transcriptional start site through reporter assays. Moreover, there is an allelic regulation of rs920778 on HOTAIR expression via this enhancer in both ESCC cell lines and normal esophageal tissue specimens, with higher HOTAIR expression among T allele carriers. These results demonstrate that functional genetic variants influencing lncRNA regulation may explain a fraction of ESCC genetic basis.

Introduction

Esophageal squamous cell carcinoma (ESCC) is one of the most common and fatal malignant tumors in the world, with a relatively high incidence in Eastern Asian compared to Western countries (1). Currently, the entire etiology of ESCC still remains to be clarified. However, accumulated epidemiology evidences indicate that tobacco smoking, heavy alcohol drinking, micronutrient deficiency as well as dietary carcinogen exposure might be main environmental risk-factors of this lethal disease (2,3). Interestingly, only a portion of exposed individuals develop ESCC, suggesting that genetic makeup may also contribute to malignant transformation of esophageal epithelial cells (2–5).

Long non-coding RNAs (lncRNAs) are an important class of genes involved in various biological processes. LncRNA HOX transcript antisense RNA (HOTAIR), which is expressed from the homebox C gene (HOXC) locus, is involved in development and progression of multiple cancers including ESCC (6–24). It has been reported that lncRNA HOTAIR could induce genome-wide retargeting of polycomb-repressive complex 2 (PRC2), lead to trimethylates histone H3 lysine-27 (H3K27me3), and promote cancer cell metastasis (6,12,15). In ESCC, upregulation of lncRNA HOTAIR expression was observed in cancerous tissue samples compared with normal esophageal tissues (20–22). Silencing HOTAIR in ESCC cell lines diminished the capability of cancer cells to form foci, migrate and invasion, altered cell cycle progression, and increased apoptosis (21,22). In xenograft mice, the tumors formed by HOTAIR-knockdown ESCC cells were smaller, both in size and weight, than the tumors formed by the control cells (21). All these evidences demonstrate the oncogenic role of HOTAIR during ESCC tumorigenesis.

To date, little or nothing has been known about the single nucleotide polymorphisms (SNP) and their functional significance in the lncRNA *HOTAIR* locus. On the basis of previous findings mentioned above, we hypothesized that the functional genetic variants in the *HOTAIR* gene may affect HOTAIR expression and/or its function, which in turn, may influence consequential risk of developing ESCC. To test this hypothesis, we selected haplotype-tagging SNPs (htSNP) across the whole *HOTAIR* locus and conducted three large independent case–control studies to investigate the association between *HOTAIR* genotypes and ESCC risk. During inspecting functional relevance of the ESCC susceptibility SNP rs920778, we identified a novel intronic enhancer of the *HOTAIR* gene as well as the allelic regulation of rs920778 on HOTAIR expression via this novel enhancer in ESCC cell lines and in patient tissue specimens.

Materials and methods

Study case–control sets

This study consisted of three case–control sets: (a) Jinan set: 1000 patients with ESCC and sex- and age-matched (± 5 years) 1000 controls. (b) Shijiazhuang set: 510 ESCC cases and 550 sex- and age-matched controls. (c) Huaian study: 588 ESCC patients and sex- and age-matched 600 controls. Sixty esophagus normal tissues adjacent to the tumors were obtained from surgically removed specimens of patients in Bethune International Peace Hospital and Huaian No. 2 Hospital. Part of these case–control sets and the tissue samples has been reported previously (26–28). The detailed case–control characteristics are described in the Supplementary Material and Methods, available at *Carcinogenesis* Online. This study was approved by the institutional Review Boards.

SNP selection and genotyping

The *HOTAIR* gene covers a 6232bp region of chromosome 12q13.13 and contains many SNPs. Therefore, we used an htSNP approach to analyze the *HOTAIR* polymorphisms globally (29). SNPs with a minor allele frequency > 5% were selected, and as many SNPs as possible from those located in potentially functional regions were included. The htSNPs were selected from a genotyped SNP database of the Han Chinese population (HCB data) of the HapMap Project (HapMap Rel 27, NCBI B36) covering 10232bp region (6232bp *HOTAIR* locus and 2kb upstream as well as 2kb downstream regions of the *HOTAIR* gene). Three htSNPs (rs920778, rs1899663 and rs4759314) were selected with Haploview version 4.2 software on a block-by-block basis, using a method described previously with the sample size inflation factor, Rh^2, of ≥ 0.8 (Supplementary Table 1, available at *Carcinogenesis* Online) (29).

Abbreviations: CI, confidence interval; ESCC, esophageal squamous cell carcinoma; H3K27me3, trimethylates histone H3 lysine; HOXC, homebox C; HOTAIR, HOX transcript antisense RNA; htSNP, haplotype-tagging SNP; lncRNA, long non-coding RNA; OR, odds ratio; PRC2, polycomb-repressive complex; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphisms.

[†]These authors contributed equally to this work.

HOTAIR htSNPs (rs920778, rs1899663 and rs4759314) were genotyped using PCR-based restriction fragment length polymorphism (RFLP) as described in Supplementary Table 2 and Figure 1, available at *Carcinogenesis* Online. Genotyping was performed without knowledge of case or control status. A 15% random sample was reciprocally tested by different person, and the reproducibility was 99.9%. In addition, a 5% random sample was also examined by Sanger sequencing, and the reproducibility was 100%.

Intronic enhancer reporter constructs

Specific primer pairs (Supplementary Table 3, available at *Carcinogenesis* Online) with XhoI and HindIII restriction sites were used to amplify multiple deletion fragments spanning intron 2 region of *HOTAIR* (from +1463bp to +2353bp, relative to the transcription start site) from human genomic DNA using Pyrobest™ DNA Polymerase (TaKaRa). The PCR products were then digested with XhoI and HindIII (New England Biolabs) and ligated into an appropriately digested pGL3-Basic vector (Promega) containing the firefly luciferase gene as a reporter. The resultant plasmid, designated p-891, was sequenced to confirm containing exclusively G or C allele at rs1899663 or rs920778 SNP position. After the p-891 plasmid was digested with XhoI and EcoRV or EcoRV and HindIII (New England Biolabs), the long restricted DNA products were recovered, blunted with Mung Bean Nuclease (TaKaRa) and ligated. The resultant constructs were named as p-635 and p-256. Similarly, the p-891 plasmid was digested with XhoI and PstI or PstI and HindIII (New England Biolabs). The long restricted DNA products were then recovered, blunted with Mung Bean Nuclease and ligated. The resultant plasmids were designated p-C and p-636. The p-C construct (contains rs920778C) was then site-specifically mutated to create construct p-T, which contains rs920778T, with the mutagenesis primers as shown in Supplementary Table 3, available at *Carcinogenesis* Online. Both p-C and p-T constructs were identical, except for the different allele at rs920778 polymorphic site. Restriction analysis and complete DNA sequencing confirmed the orientation and integrity of these constructs.

Dual luciferase reporter assay

KYSE30 and KYSE150 ESCC cells (6×10^4) were placed in 24-well plates and transfected with both 50 ng of reporter constructs (pGL3-Basic, p-891, p-635, p-256, p-C, p-T and p-636) using Lipofectamine™ 2000 (Invitrogen) when grown to 50% confluence. pRL-SV40 (1 ng) (Luciferase Assay System; Promega) containing renilla reniformis luciferase was cotransfected to standardize transfection efficiency. Dual luciferase activities (both firefly luciferase activity and renilla luciferase activity) were determined at 48 h after transfection using a luciferase assay system (Promega) as described previously (28). For each plasmid construct, three independent transfection experiments were performed, and each was done in triplicate. Fold increase was calculated by defining the activity of empty pGL3-Basic vector as 1. In brief, the relative luciferase value for each pGL3-Basic derived plasmid equals to ratio of its firefly luciferase activity (standardized by the renilla luciferase activity of the same sample) and pGL3-Basic's firefly luciferase activity (standardized by the renilla luciferase activity of the same sample). That is, the relative luciferase value of pGL3-Basic transfected cells was 1, and that value for each pGL3-Basic derived plasmid equals to folds of 1.

Real-time analyses of lncRNA HOTAIR

Total RNA was extracted from normal esophagus tissue samples using TRIzol Reagent (Invitrogen). After that, we treated each RNA sample with RNase-Free DNase to remove genomic DNA (Ambion). SYBR-Green real-time quantitative PCR method was used to examine lncRNA HOTAIR levels in normal esophagus tissues. The expression of individual HOTAIR measurements was calculated relative to expression of *β-actin* using the 2^{-dCt} method as described previously (26,28). The detailed procedures are described in the Supplementary Material and Methods and Table 4, available at *Carcinogenesis* Online.

Statistical analyses

Pearson's χ^2 test was used to examine the differences in demographic variables and genotype distributions of *HOTAIR* htSNPs (rs920778, rs1899663 and rs4759314) between ESCC patients and controls. Associations between *HOTAIR* htSNPs genotypes and risk of the development of ESCC were estimated by odds ratio (OR) and their 95% confidence intervals (CIs) computed using unconditional logistic regression model. All ORs were adjusted for age, sex, drinking and smoking status, where it was appropriate. A P value of less than 0.05 was used as the criterion of statistical significance, and all statistical tests were two-sided. All analyses were performed with SPSS software package (Version 16.0, SPSS, Chicago, IL).

Results

Allelic frequencies and genotype distributions of HOTAIR htSNPs in the training case–control set

There are no statistically significant differences between patients and control subjects for all three case–control sets in terms of median age and sex distributions (all $P > 0.05$), indicating that the frequency matching was adequate (Table I). However, more smokers and alcohol drinker were observed among ESCC cases compared with controls in all case–control sets.

Allele frequencies and genotype distributions of *HOTAIR* htSNPs (rs920778 C>T, rs1899663 G>T and rs4759314 A>G) in patients and controls from the training set are showed in Table II. The allele frequencies for rs920778 T, rs1899663 T and rs4759314 G were 0.278, 0.147 and 0.043 in cases and 0.220, 0.151 and 0.046 in controls in Jinan training case–control set. All observed genotype frequencies in both controls and patients conform to Hardy–Weinberg equilibrium. Distributions of the rs920778, rs1899663 and rs4759314 genotypes were then compared among cases and controls. Frequencies of rs920778 CC, CT and TT genotypes among ESCC cases differed significantly from those among controls ($\chi^2 = 20.23$, $P = 4.04 \times 10^{-5}$,

Table I. Distribution of selected characteristics among ESCC cases and controls

Variable	Jinan case–control set (training set)			Shijiazhuang case–control set (validation set 1)			Huaian case–control set (validation set 2)		
	Cases	Controls	P^a	Cases	Controls	P^a	Cases	Controls	P^a
	No. (%)	No. (%)		No. (%)	No. (%)		No. (%)	No. (%)	
	1000	1000		510	550		588	600	
Age (year)[b]			0.474			0.433			0.725
≤57(≤59)	516 (51.6)	500 (50.0)		271 (53.1)	279 (50.7)		288 (49.0)	300 (50.0)	
>57(>59)	484 (48.4)	500 (50.0)		239 (46.9)	271 (49.3)		300 (51.0)	300 (50.0)	
Sex			0.426			0.218			0.678
Male	776 (77.6)	761 (76.1)		398 (78.0)	446 (81.1)		413 (70.2)	428 (71.3)	
Female	224 (22.4)	239 (23.9)		112 (22.0)	104 (18.9)		175 (29.8)	172 (28.7)	
Smoking status			<0.001			<0.001			<0.001
Yes	752 (75.2)	396 (39.6)		381 (74.7)	263 (47.8)		437 (74.3)	203 (33.8)	
No	248 (24.8)	604 (60.4)		129 (25.3)	287 (52.2)		151 (25.7)	397 (66.2)	
Drinking status			<0.001			<0.001			<0.001
Yes	553 (55.3)	401 (40.1)		288 (56.5)	228 (41.5)		334 (56.8)	242 (40.3)	
No	447 (44.7)	599 (59.9)		222 (43.5)	322 (58.5)		254 (43.2)	358 (59.7)	

ESCC, esophageal squamous cell carcinoma.
[a]Two-sided χ^2 test.
[b]Median ages of cases for Jinan case–control set, Shijiazhuang case–control set and Huaian case–control set are 56, 59 and 59 years, respectively.

df = 2), with the frequency of TT homozygote being significantly higher among patients than among controls (8.3% versus 4.1%). However, no statistically significant differences of rs1899663 and rs4759314 genotypes were observed between ESCC patients and healthy controls (both $P > 0.05$). Therefore, no other studies of these two genetic variants were conducted in the next analyses.

Association between HOTAIR rs920778 SNP and ESCC risk

Associations between genotypes of *HOTAIR* rs920778 and risk of ESCC were firstly estimated in Jinan training set using unconditional logistic regression analyses (Table III). The rs920778 T allele was shown to be risk allele. Individuals having the rs920778 TT genotype had an OR of 1.37 (95%CI = 1.11–1.69, $P = 0.003$) for developing ESCC in Jinan population compared with individual having the CC genotype. This association was successfully validated in the other two independent case–control sets (Shijiazhuang set: OR = 1.78, 95%CI = 1.34–2.38, $P = 7.7 \times 10^{-4}$; Huaian set: OR = 2.08, 95%CI = 1.51–2.85, $P = 5.9 \times 10^{-4}$) (Table III). Intriguingly, the rs920778 CT carriers also showed a 1.56- or 1.53-fold increased ESCC risk compared with those carrying the rs920778 CC genotype in two validation sets (95%CI = 1.19–2.04, $P = 0.001$ or, 95%CI = 1.17–2.00, $P = 0.002$). In the pooled analyses, we observed that the rs920778 CT or TT genotype was also significantly associated with ESCC risk (OR = 1.37, 95%CI = 1.19–1.57, $P = 7.7 \times 10^{-5}$ or, OR = 1.61, 95%CI = 1.39–1.87, $P = 2.4 \times 10^{-6}$) (Table III). All ORs were adjusted for sex, age, smoking and alcohol drinking status. There were no gene–environment interactions between *HOTAIR* rs920778 SNP and sex, age, smoking or alcohol drinking status.

Identification of a novel HOTAIR intronic enhancer

Because the ESCC susceptibility SNP rs920778 is located within the *HOTAIR* intron 2 and previous H3K4me1 and H3K4me3 modification data suggest that there might be a potential enhancer in this intron (Supplementary Figure 2, available at *Carcinogenesis* Online), we therefore examined the enhancer activity of this region by a set of luciferase reporter gene constructs in human ESCC KYSE30 and KYSE150 cells (Figure 1). Interestingly, we observed a highest enhancer activity for construct p-635 which contains a 635bp 3′-region of intron 2, compared with all other constructs, such as p-891 which contains the full length intron 2, as well as p-C which contains a 255bp 3′-region of intron 2. These results indicated that there may be a negative regulatory element between +1463bp and +1718bp

Table II. Associations between candidate SNPs in *HOTAIR* and ESCC risk in Jinan case–control set (training set)

	rs920778			rs1899663			rs4759314		
		ESCC	Control		ESCC	Control		ESCC	Control
Genotypes (%)	CC	52.8	60.1	GG	72.5	72.4	AA	91.7	91.0
	CT	38.9	35.8	GT	25.6	25.0	AG	8.1	8.9
	TT	8.3	4.1	TT	1.9	2.6	GG	0.2	0.1
OR[a] (95% CI)	CT	1.21 (0.99–1.47)		GT	1.00 (0.81–1.25)		AG	0.86 (0.61–1.20)	
	TT	1.37 (1.11–1.69)		TT	0.86 (0.62–1.18)		GG	1.00 (0.30–3.34)	
P	CT	0.063		GT	0.992		AG	0.857	
	TT	0.003		TT	0.344		GG	0.996	
P_{trend}[b]		2.6×10^{-5}			0.723			0.642	
Location		Intron 2			Intron 2			Intron 1	
Position[c]		52646499			52647261			52648102	

ESCC, esophageal squamous cell carcinoma; OR, odds ratio; CI, confidence interval.
[a]Data were calculated by unconditional logistic regression, adjusted for age, sex, drinking and smoking.
[b]Tests for trend of odds were two sided and were based on likelihood ratio tests assuming an additive model.
[c]Position in NCBI build 36.

Table III. Genotype frequencies of the *HOTAIR* rs920778 C>T polymorphism among ESCC cases and controls and their association with ESCC risk

Studies	Genotypes	Cases No. (%)	Controls No. (%)	OR[a] (95% CI)	*P*[a]
Jinan set		*n* = 1000	*n* = 1000		
	CC	528 (52.8)	601 (60.1)	1.00 (Reference)	
	CT	389 (38.9)	358 (35.8)	1.21 (0.99–1.47)	0.063
	TT	83 (8.3)	41 (4.1)	1.37 (1.11–1.69)	0.003
P_{trend}[b]				2.6×10^{-5}	
		n = 510	*n* = 550		
Shijiazhuang set	CC	256 (50.2)	344 (62.5)	1.00 (Reference)	
	CT	207 (40.6)	186 (33.8)	1.56 (1.19–2.04)	0.001
	TT	47 (9.2)	20 (3.6)	1.78 (1.34–2.38)	7.7×10^{-4}
P_{trend}[b]				2.0×10^{-6}	
		n = 588	*n* = 600		
Huaian set	CC	307 (52.2)	378 (63.0)	1.00 (Reference)	
	CT	203 (39.1)	205 (34.2)	1.53 (1.17–2.00)	0.002
	TT	51 (8.7)	17 (2.8)	2.08 (1.51–2.85)	5.9×10^{-4}
P_{trend}[b]				4.4×10^{-5}	
		n = 2098	*n* = 2150		
Pooled	CC	1091 (52.0)	1323 (61.5)	1.00 (Reference)	
	CT	826 (39.4)	749 (34.9)	1.37 (1.19–1.57)	7.7×10^{-5}
	TT	181 (8.6)	78 (3.6)	1.61 (1.39–1.87)	2.4×10^{-6}
P_{trend}[b]				8.2×10^{-15}	

ESCC, esophageal squamous cell carcinoma; OR, odds ratio; CI, confidence interval.
[a]Data were calculated by logistic regression with adjustment for age, sex, smoking and drinking status.
[b]Test for trend of odds was two-sided and based on likelihood ratio test assuming an additive model.

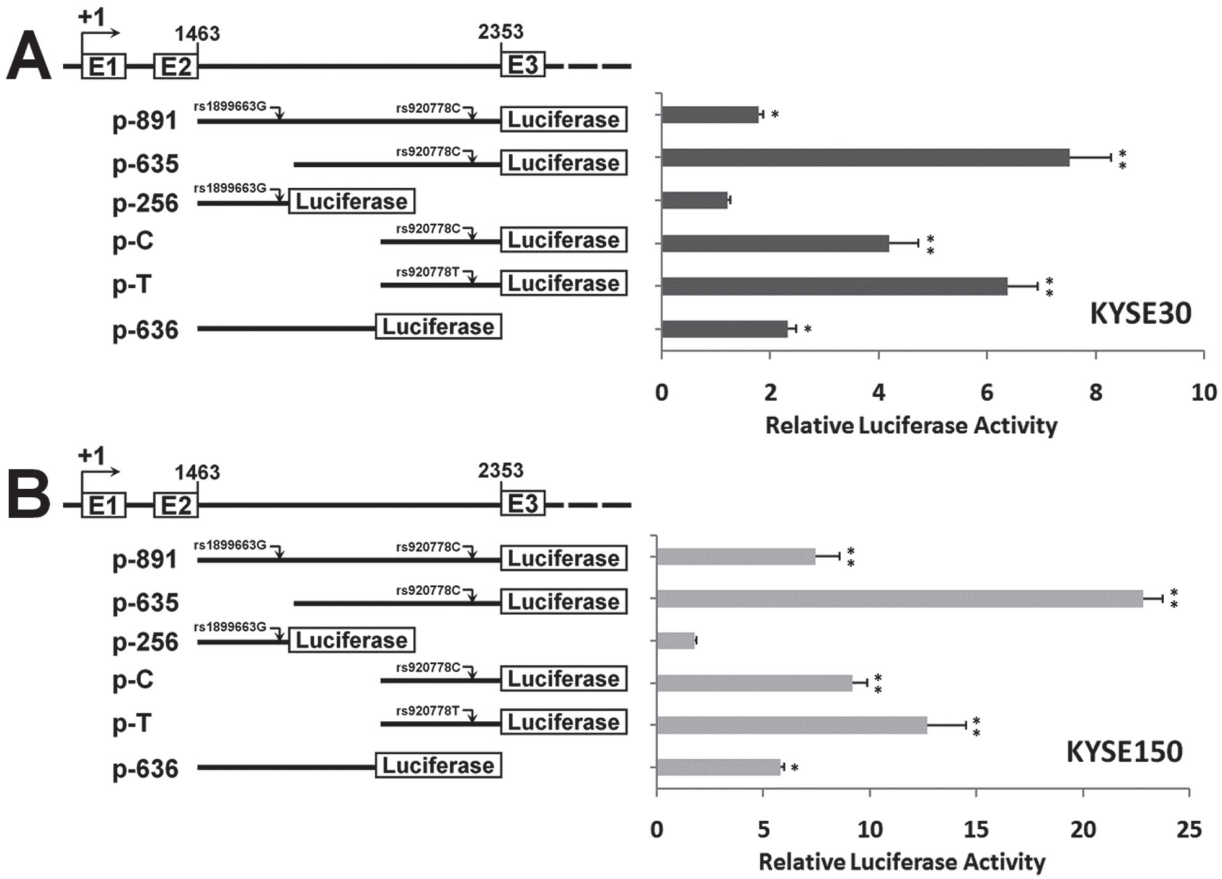

Fig. 1. Transient luciferase reporter gene expression assays with constructs containing different length or different rs920778 allele of *HOTAIR* intron 2 region in KYSE-30 cells (**A**) or KESY-150 cells (**B**). pRL-SV40 were cotransfected with these constructs to standardize transfection efficiency. Fold increase was measured by defining the activity of cells co-transfected with both pGL3-basic as 1. All experiments were performed in triplicates in three independent transfection experiments and each value represents mean ± SD. Compared with pGL3-Basic transfected cells, $*P < 0.05$; $**P < 0.01$.

from the transcriptional start site and the core region of this intronic enhancer might exist between +1719bp and +2353bp from the transcriptional start site.

Functional relevance of rs920778 on HOTAIR expression

We next examined whether the ESCC susceptibility SNP rs920778 has an allele-specific effect on the intronic enhancer activity and, thus, *HOTAIR* expression in both ESCC cells (*in vitro*) and esophagus tissues (*in vivo*). Either KYSE30 cells or KYSE150 cells transfected with *HOTAIR* rs920778C allelic reporter construct (p-C) showed significantly lower luciferase activities compared to cells expression rs920778T allelic reporter construct (p-T) (both $P < 0.05$) (Figure 1A and 1B). As shown in Figure 2, we found that subjects with the rs920778 TT genotype had significantly lower *HOTAIR* RNA levels (mean ± SD) than those with the CC genotypes in normal esophagus tissues (2.178 ± 1.071 [$n = 6$] versus 0.993 ± 1.021 [$n = 30$], $P < 0.01$). Similar results were observed when the *HOTAIR* RNA levels were compared between rs920778 TT and CT genotypes in normal esophagus tissues (2.178 ± 1.071 [$n = 6$] versus 1.022 ± 1.061 [$n = 24$], $P < 0.01$). However, there were no statistically significant differences of HOTAIR expression between TT, CT and CC genotypes in ESCC tissue samples (2.672 ± 1.231 [$n = 6$], 2.429 ± 1.137 [$n = 24$] and 2.388 ± 1.172 [$n = 30$]).

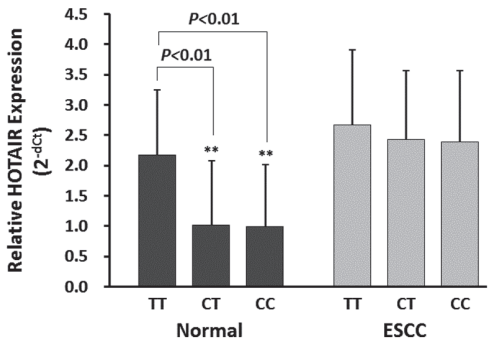

Fig. 2. *HOTAIR* RNA expression (mean ± SD) in normal and cancerous esophagus tissues grouped by *HOTAIR* rs920778 C>T genotypes. The expression of individual HOTAIR measurements was calculated relative to expression of *β-actin* using the 2^{-dCt} method. In normal tissues, there were significantly lower *HOTAIR* RNA levels among subjects with the rs920778 TT genotype compared with the CC genotypes (2.178 ± 1.071 [$n = 6$] versus 0.993 ± 1.021 [$n = 30$], $P < 0.01$). Subjects with rs920778 TT genotype also showed higher *HOTAIR* RNA expression than those carrying the rs920778 CT genotype (2.178 ± 1.071 [$n = 6$] versus 1.022 ± 1.061 [$n = 24$], $P < 0.01$). However, no statistically significant differences of HOTAIR expression were found between TT, CT and CC genotypes in ESCC tissue samples (2.672 ± 1.231 [$n = 6$], 2.429 ± 1.137 [$n = 24$] and 2.388 ± 1.172 [$n = 30$]). $**P < 0.01$.

Discussion

In this study, we examined the association between lncRNA *HOTAIR* htSNPs and risk of developing ESCC through a two stage case–control approach. To the best of our knowledge, this is the first study to examine the role of *HOTAIR* genetic polymorphism in ESCC carcinogenesis. Among three *HOTAIR* htSNPs (rs920778, rs1899663 and rs4759314), only rs920778 was significantly associated with increased ESCC risk in Chinese. Reporter gene assays indicated that

there might be a novel *HOTAIR* intronic enhancer locating between +1719bp and +2353bp from the transcriptional start site. Interestingly, the ESCC susceptibility SNP rs920778 in this enhancer has a genotype-specific effect on lncRNA *HOTAIR* expression. Our observations also support the hypothesis that functional genetic variants influencing lncRNA regulation may explain a part of ESCC genetic basis.

Gupta and colleagues firstly revealed the relationship between lncRNA HOTAIR and H3K27me3 as well metastasis of breast cancer (6). HOTAIR expression was low in normal breast epithelia but high in primary breast cancer as well as metastatic lesions (6). Additionally, breast cancer patients with high HOTAIR expression had a poorer prognosis for overall survival and for metastasis-free survival than did those with low HOTAIR expression (6–8). Similar phenomena have been observed in other cancer types, such as hepatocellular carcinoma (9–11), colorectal cancer (12), sarcoma (13), gastrointestinal stromal tumors (14), pancreatic cancer (15), laryngeal squamous cell carcinoma (16), nasopharyngeal carcinoma (17), lung cancer (18,19), ESCC (20–22), gastric cancer (23,24) and melanoma (25). All these evidences indicate that lncRNA HOTAIR can act as an oncogene and increased HOTAIR expression might result in malignant transformation of normal cells. Interestingly, we found that *HOTAIR* rs920778T allele is associated with increased ESCC risk compared with C allele. Since rs920778T allele is correlated to significantly increased HOTAIR RNA expression in normal esophageal tissues, the association would be expected and biological rational.

The *HOTAIR* intronic enhancer identified in the current study is not a 'conventional' enhancer. However, this kind of enhancers is not discovered for the first time and does exist in many human genes. For example, there is a functional intronic promoter locating in MDM2, a well-known oncogene (30–32). Interestingly, a regulatory SNP (rs2279744) in this intronic promoter could regulate expression of MDM2, attenuate the P53 tumor suppressor pathway and contribute to genetic susceptibility of multiple cancers (32–34). In our case, both H3K4me1 and H3K4me3 modification data as well as our luciferase reporter gene results indicate the existence of this intronic enhancer. Therefore, we believe that the HOTAIR intronic enhancer may work in a similar way as the intronic MDM2 promoter does.

Several limitations may exist in the current case–control study. For example, because it was a hospital-based study and the cases were from the hospital, there might be inherent selection bias. Thus, the findings of our case–control study warrant to be validated in a population-based prospective study in the future. Since the TT genotype of the functional rs920778 SNP is relatively rare (about 3–4% among common populations), the potential clinical translation of this genetic variant might be compromised.

In conclusion, we identified a novel *HOTAIR* intronic enhancer and a functional ESCC susceptibility SNP rs920778 in Chinese populations. The rs920778C>T SNP located in the intronic enhancer region alters the activity of this enhancer, resulting in increased HOTAIR RNA expression, which might be the underlying mechanism in conferring ESCC susceptibility. These data can lead to better understanding, risk assessment, early detection and targeted treatment of ESCC.

Supplementary material

Supplementary Materials and Methods, Tables 1–4 and Figures 1 and 2 can be found at http://carcin.oxfordjournals.org/

Funding

National Natural Science Foundation of China (31271382, 91229126 and 81201586); the open project of State Key Laboratory of Molecular Oncology (SKL-KF-2013-03); Beijing Natural Science Foundation (5122020); Beijing Higher Education Young Elite Teacher Project (YETP0521); Beijing City Talent Training Project (2012D009016000002); Program for Changjiang Scholars and Innovative Research Team in University (IRT13045) and Shandong Provincial Natural Science Foundation (Y2008C36).

Conflict of Interest Statement: None declared.

References

1. Jemal,A. *et al.* (2009) Cancer statistics, 2009. *CA Cancer J. Clin.*, **59**, 225–249.
2. Gao,Y.T. *et al.* (1994) Risk factors for esophageal cancer in Shanghai, China. I. Role of cigarette smoking and alcohol drinking. *Int. J. Cancer*, **58**, 192–196.
3. Hu,J. *et al.* (1994) Risk factors for oesophageal cancer in northeast China. *Int. J. Cancer*, **57**, 38–46.
4. Venkitaraman,A.R. (2009) Linking the cellular functions of BRCA genes to cancer pathogenesis and treatment. *Annu. Rev. Pathol.*, **4**, 461–487.
5. Wu,C. *et al.* (2012) Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat. Genet.*, **44**, 1090–1097.
6. Gupta,R.A. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**, 1071–1076.
7. Lu,L. *et al.* (2012) Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer. *Breast Cancer Res. Treat.*, **136**, 875–883.
8. Chisholm,K.M. *et al.* (2012) Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. *PLoS One*, **7**, e47998.
9. Yang,Z. *et al.* (2011) Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Ann. Surg. Oncol.*, **18**, 1243–1250.
10. Geng,Y.J. *et al.* (2011) Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J. Int. Med. Res.*, **39**, 2119–2128.
11. Ishibashi,M. *et al.* (2013) Clinical significance of the expression of long non-coding RNA HOTAIR in primary hepatocellular carcinoma. *Oncol. Rep.*, **29**, 946–950.
12. Kogo,R. *et al.* (2011) Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.*, **71**, 6320–6326.
13. Milhem,M.M. *et al.* (2011) Correlation of MTDH/AEG-1 and HOTAIR Expression with Metastasis and Response to Treatment in Sarcoma Patients. *J. Cancer Sci. Ther.*, **S5**, 4.
14. Niinuma,T. *et al.* (2012) Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors. *Cancer Res.*, **72**, 1126–1136.
15. Kim,K. *et al.* (2013) HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. *Oncogene*, **32**, 1616–1625.
16. Li,D. *et al.* (2013) Long intergenic noncoding RNA HOTAIR is overexpressed and regulates PTEN methylation in laryngeal squamous cell carcinoma. *Am. J. Pathol.*, **182**, 64–70.
17. Nie,Y. *et al.* (2013) Long non-coding RNA HOTAIR is an independent prognostic marker for nasopharyngeal carcinoma progression and survival. *Cancer Sci.*, **104**, 458–464.
18. Zhuang,Y. *et al.* (2013) Induction of long intergenic non-coding RNA HOTAIR in lung cancer cells by type I collagen. *J. Hematol. Oncol.*, **6**, 35.
19. Nakagawa,T. *et al.* (2013) Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochem. Biophys. Res. Commun.*, **436**, 319–324.
20. Lv,X.B. *et al.* (2013) Long noncoding RNA HOTAIR is a prognostic marker for esophageal squamous cell carcinoma progression and survival. *PLoS One*, **8**, e63516.
21. Li,X. *et al.* (2013) Long non-coding RNA HOTAIR, a driver of malignancy, predicts negative prognosis and exhibits oncogenic activity in oesophageal squamous cell carcinoma. *Br. J. Cancer*, **109**, 2266–2278.
22. Chen,F.J. *et al.* (2013) Upregulation of the long non-coding RNA HOTAIR promotes esophageal squamous cell carcinoma metastasis and poor prognosis. *Mol. Carcinog.*, **52**, 908–915.
23. Xu,Z.Y. *et al.* (2013) Knockdown of long non-coding RNA HOTAIR suppresses tumor invasion and reverses epithelial-mesenchymal transition in gastric cancer. *Int. J. Biol. Sci.*, **9**, 587–597.
24. Hajjari,M. *et al.* (2013) Up-regulation of HOTAIR long non-coding RNA in human gastric adenocarcinoma tissues. *Med. Oncol.*, **30**, 670.
25. Tang,L. *et al.* (2013) Long noncoding RNA HOTAIR is associated with motility, invasion, and metastatic potential of metastatic melanoma. *Biomed. Res. Int.*, **2013**, 251098.
26. Liu,L. *et al.* (2012) Functional FEN1 genetic variants contribute to risk of hepatocellular carcinoma, esophageal cancer, gastric cancer and colorectal cancer. *Carcinogenesis*, **33**, 119–123.

27. Shi,J. *et al.* (2013) Leukocyte telomere length-related genetic variants in 1p34.2 and 14q21 loci contribute to the risk of esophageal squamous cell carcinoma. *Int. J. Cancer*, **132**, 2799–2807.
28. Zhang,X. *et al.* (2013) A functional BRCA1 coding sequence genetic variant contributes to risk of esophageal squamous cell carcinoma. *Carcinogenesis* **34**, 2309–2313
29. Yang,M. *et al.* (2011) SLCO2B1 and SLCO1B3 may determine time to progression for patients receiving androgen deprivation therapy for prostate cancer. *J. Clin. Oncol*., **29**, 2565–2573.
30. Zauberman,A. *et al.* (1995) A functional p53-responsive intronic promoter is contained within the human mdm2 gene. *Nucleic Acids Res*., **23**, 2584–2592.
31. Ries,S. *et al.* (2000) Opposing effects of Ras on p53: transcriptional activation of mdm2 and induction of p19ARF. *Cell*, **103**, 321–330.
32. Bond,G. L. *et al.* (2004) A single nucleotide polymorphism in the MDM2 promoter attenuates the TP53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell*, **119**, 591–602.
33. Hu,W. *et al.* (2007) A single nucleotide polymorphism in the MDM2 gene disrupts the oscillation of p53 and MDM2 levels in cells. *Cancer Res*., **67**, 2757–2765.
34. Yang,M. *et al.* (2007) Interaction of P53 Arg72Pro and MDM2 T309G polymorphisms and their associations with risk of gastric cardia cancer. *Carcinogenesis*, **28**, 1996–2001.

Received October 31, 2013; revised April 4, 2014; accepted April 26, 2014