Mutagenesis, 2014, 1–8
doi:10.1093/mutage/geu076
Original Article

Original Manuscript

Genetic variants in lncRNA *HOTAIR* are associated with risk of colorectal cancer

Yao Xue[1,2,†], Dongying Gu[3,†], Gaoxiang Ma[1,2,†], Lingjun Zhu[4], Qiuhan Hua[1,2], Haiyan Chu[1], Na Tong[1], Jinfei Chen[3], Zhengdong Zhang[1,2,*,‡] and Meilin Wang[1,2,*,‡]

[1]Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Cancer Center, [2]Department of Genetic Toxicology, the Key Laboratory of Modern Toxicology, Ministry of Education, School of Public Health, [3]Department of Oncology, Nanjing First Hospital, Nanjing Medical University, [4]Department of Oncology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, China

*To whom correspondence should be addressed. Tel: ++86 25 86868423; Fax: +86 25 86862684; Email: drzdzhang@gmail.com
Correspondence may also be addressed to Meilin Wang. Tel: +86 25 86868423; Fax: +86 25 86862684; Email: mwang@njmu.edu.cn

†These authors contributed equally to this work.
‡These authors jointly directed this work.

Abstract

Long non-coding RNA HOX transcript antisenseRNA (*HOTAIR*) has been widely identified to participate in tumour pathogenesis, acting as a promoter in colorectal cancer carcinogenesis. However, the association between genetic variants in *HOTAIR* and cancer risk has not yet been reported. In the present study, we performed a two-stage case–control study to investigate the association between *HOTAIR* tagSNPs and the risk of colorectal cancer. We found that individuals with rs7958904 CC genotype had a significantly decreased risk of colorectal cancer in both Stage 1 and 2, compared with those carrying GG genotype [odds ratio (OR) = 0.70, 95% confidence interval (CI) = 0.51–0.97 in Stage 1; OR = 0.58, 95% CI = 0.37–0.91 in Stage 2; OR = 0.67, 95% CI = 0.51–0.87 in combined stage]. The subsequently stratified analyses showed that the protective effect of rs7958904 was more pronounced in several subgroups. In summary, our study showed that genetic variants in *HOTAIR* were associated with risk of colorectal cancer and rs7958904 may act as a potential biomarker for predicting the risk of colorectal cancer.

Introduction

Colorectal cancer is one of the most common cancers worldwide, with about half a million deaths each year (1). The incidence of colorectal cancer varies significantly among different geographic regions and is highest in North America and Oceania (2). However, with changes in daily habits and recent advances in diagnostic techniques, the incidence of colorectal cancer has increased dramatically in China. According to statistical data reported by WHO, colorectal cancer has the third fastest rising incidence in China, and mortality from colon cancer increased 50% between 1991 to 2010 (3). Currently, colorectal cancer is the third most common cancer and has the fifth highest mortality in China, causing serious problems for public health.

In recent years, a novel kind of non-coding RNA, long non-coding RNA (lncRNAs), has gained attention for its complex and extensive regulatory functions (4). LncRNAs are defined as non-coding RNAs longer than 200 nucleotides (5). Studies have shown that lncRNAs may possess various functions in carcinogenesis, including transcriptional, post-transcriptional and epigenetic regulation of cancer-related genes (6). Among them, some have been identified to be related to specific diseases, e.g. lncRNA MALAT1 has been found to be a negative predictor of outcome in lung cancer (7) and lncRNA DD3 has been demonstrated to be a motivator in the carcinogenesis of prostate cancer (8). Similar dysregulation of lncRNAs has also been observed in colorectal cancer. Kogo *et al.* (9) reported that lncRNA HOX transcript antisenseRNA (*HOTAIR*) is upregulated in colorectal cancer and is associated with poor prognosis of colorectal cancer. In addition, *HOTAIR* has also been identified to participate in some other cancers (10). *HOTAIR* is HOX antisense intergenic RNA, located in the HOXC locus. Given the remarkable dysregulation of *HOTAIR* in tumours, investigations into its biological function have been conducted widely. So far, studies have revealed that the major role of *HOTAIR* involves epigenetic regulation of

transcription in a 40 kb region of HOXD (10,11). The same function of regulation of polycomb-dependent chromatin modification by *HOTAIR* for cancer cells has been found in colorectal cancer (9).

Genetic variants, majorly composed of single nucleotide polymorphisms (SNPs), have long been confirmed to participate in carcinogenesis. These variants may exert various influences on expression or function of a particular gene (12). Recently, studies on the effects of SNPs are no longer confined to cancer-related protein coding genes, but are also extended to functional lncRNAs. SNPs in several lncRNAs previously identified to be involved in cancer have been reported to be associated with cancer risk, e.g. rs6434568 in the prostate cancer-related *PCGEM1* gene (13) and rs7763881 in the hepatocellular cancer-related *HULC* gene (14). These results provide evidence for the important role of lncRNA SNPs in carcinogenesis. Based on the role of *HOTAIR* in the molecular mechanism of colorectal cancer, as well as the influence of genetic variants on gene function, we hypothesised that genetic variants in *HOTAIR* may modify the risk of colorectal cancer.

In the present study, we evaluated the association between *HOTAIR* tagSNPs and colorectal cancer risk in a two-stage case–control study of Chinese population. As a result, we found that rs7958904, located on exon 6 of *HOTAIR* gene, was significantly associated with decreased colorectal cancer risk.

Materials and methods

Study population

There were 1147 colorectal cancer cases and 1203 cancer-free controls included in Stage 1, and Stage 2 included an additional 587 cases and 652 controls. All subjects were genetically unrelated ethnic Han Chinese. Briefly, patients were recruited in an ongoing study of colorectal cancer, which was conducted in the First Affiliated Hospital and Nanjing First Hospital of Nanjing Medical University from September 2010. Detailed information of the Stage 1 has been described previously (15–17). Subjects in Stage 2 were enrolled in a later time period than those in Stage 1 and can be used as independent samples for replication of results in Stage 1. All cases were newly diagnosed on the time of enrollment without restrictions of age and sex and have been histologically confirmed to be colorectal adenocarcinoma. The controls were randomly recruited from a pool of >25 000 subjects in the same geographical region, who were seeking physical examinations (i.e. they came to the hospital for a routine examination of physical conditions, not for illness or symptoms). All the controls were genetically unrelated with cases. Subjects with histories of cancer were excluded from the control group. In order to diminish the influence of demographic differences between cases and controls on the statistical results, controls were frequency-matched to cases on age (±5 years) and sex. We also obtained clinical information for all the cases. The pathological stage of colorectal cancer was divided into Dukes A, B, C and D. Tumour grade was classified into low, intermediate and high. After informed consent was obtained, a questionnaire about lifestyle factors was completed by all the subjects through face-to-face interviews. Finally, each subject donated 5 ml of blood for genomic DNA extraction. This study was approved by the institutional review boards of Nanjing Medical University.

Samples of colorectal cancer tissues and corresponding non-tumour tissues were obtained from patients who underwent surgical operation at the First Affiliated Hospital and Nanjing First Hospital of Nanjing Medical University. Samples were frozen with liquid nitrogen immediately after surgical resection. Finally, a total of 95 pairs of colorectal cancer and the corresponding normal tissues were collected in the present study.

SNP selection and genotyping

Using genotype data of Han Chinese in Beijing from HapMap database (HapMap Data Rel 24/Phase II, Nov 08, on NCBI B36 assembly, dbSNP b126), we selected three tagSNPs capturing all the common SNPs (minor allele frequency, MAF > 0.05) located in the chromosome locus transcribed into *HOTAIR* and its flanking 2000 bp region. The selection was conducted with the pairwise option of the Haploview 4.0 software (Cambridge, MA, USA) and the threshold for analyses was set as $r^2 > 0.8$.

Overall flow of SNP selection and relative locations of the three selected tagSNPs are summarised in Figure 1.

Genomic DNA was extracted from white blood cell fractions using the Qiagen Blood Kit (Qiagen). We used TaqMan allelic discrimination methods to detect genotypes of the three tagSNPs. Assays were performed with the ABI 7900HT real-time PCR system (Applied Biosystems, Foster City, CA, USA). Sequences of primers and probes of each SNP are listed in Supplementary Table 1, available at *Mutagenesis* Online. Ten percentage of samples were selected for repeated genotyping and the concordance rate was 100.0%. The call rates were >98% for all variants.

Total RNA extraction and real-time reverse transcription-polymerase chain reaction

Total RNA was extracted from colorectal cancer tissues and cells using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Then cDNA was synthesised with M-MLV reverse transcriptase (Invitrogen). Real-time reverse transcription-polymerase chain reaction (RT-PCR) with SYBR Green assay (TaKaRa Biotechnology, Dalian, China) was performed to examine expression level of *HOTAIR*. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as an internal control. The assay was conducted using the ABI 7300 system (Applied Biosystems). The primers for *HOTAIR* and *GAPDH* are available on request. All reactions were done in triplicate and expression level of *HOTAIR* was calculated according to the equation $2^{-\Delta\Delta C_t}$.

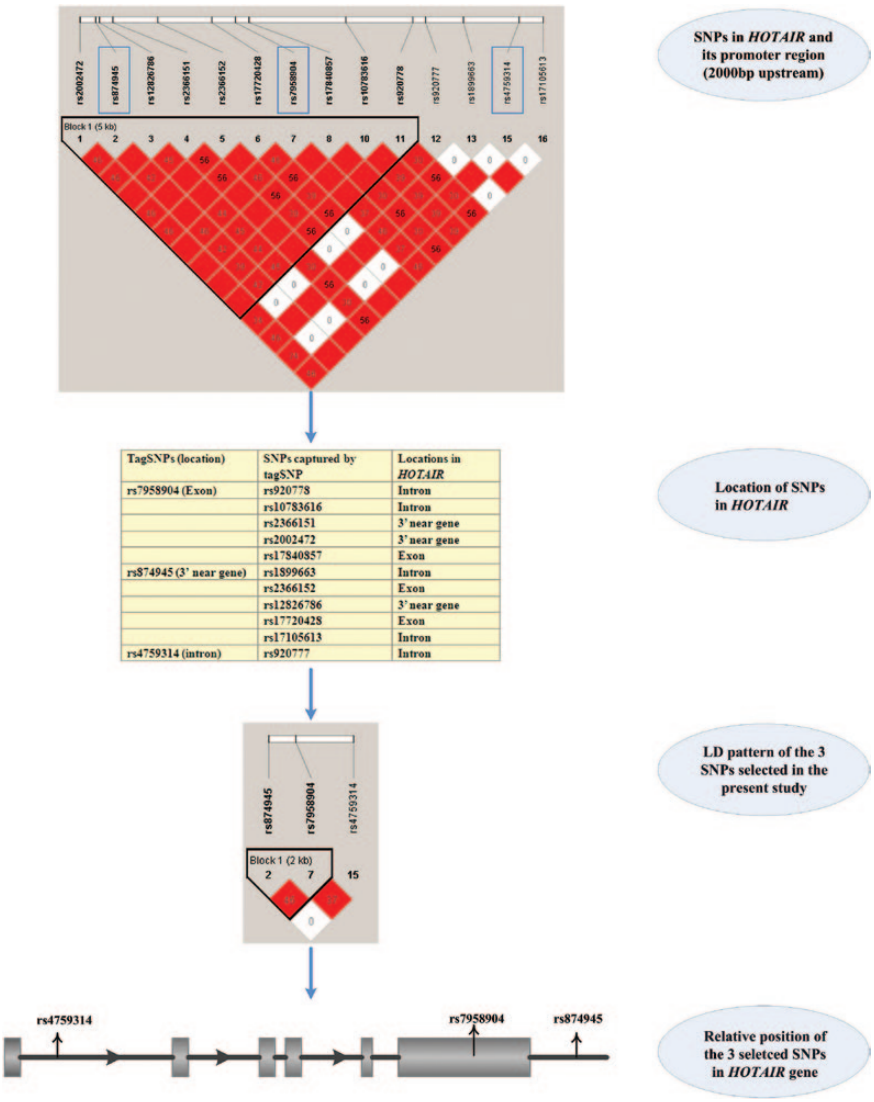
HOTAIR overexpression vector construction and cell transfection

We used *HOTAIR* overexpression vectors, containing different alleles in rs7958904, to detect effects of the associated variant on the expression level of *HOTAIR* and proliferation rate of colorectal cancer cells. The overexpression vectors were prepared by amplifying full lengths of complementary DNA encoding *HOTAIR* and then they were cloned into pcDNA 3.1 vector (Invitrogen). All constructs were sequenced to confirm their authenticity.

We used Lipofectamine 2000 (Invitrogen) to transiently transfect *HOTAIR* overexpression plasmid into colorectal cancer LoVo cells, according to the manufacturer's instructions. All the experiments were conducted by the same technician following uniform protocol, in order to eliminate the influence of inequable transfection efficiencies. After a 48 h transfection, cells were collected and cell lysates were prepared for total RNA extraction.

Cell proliferation assay

Approximately 5.0×10^3 transfected LoVo cells were plated in 96-well plates. After 24, 48 and 72 h of transfection, the proliferation rate of LoVo cells was assessed using the Cell Counting Kit 8 (Dojindo, Kumamoto, Japan). The absorbance value of each well was determined at 450 nm using the Infinite M200 spectrophotometer (Tecan, Germany). Each experiment was repeated three times.



TagSNPs (location)	SNPs captured by tagSNP	Locations in *HOTAIR*
rs7958904 (Exon)	rs920778	Intron
	rs10783616	Intron
	rs2366151	3′near gene
	rs2002472	3′near gene
	rs17840857	Exon
rs874945 (3′ near gene)	rs1899663	Intron
	rs2366152	Exon
	rs12826786	3′near gene
	rs17720428	Exon
	rs17105613	Intron
rs4759314 (intron)	rs920777	Intron

Figure 1. Overall flow of SNP selection and relative location of the three selected tagSNPs in *HOTAIR*. TagSNPs selected in Haploview were marked with blue line box

Statistical analysis

The associations between *HOTAIR* tagSNPs and risk of colorectal cancer were evaluated by odds ratios (ORs) and their 95% confidence intervals (CIs) from unconditional univariate and multivariate logistic regression analyses. Chi-square (χ^2) test was used to estimate the distribution differences of demographic characteristics as well as genotypes and alleles of the *HOTAIR* variants between cases and controls. All ORs were adjusted for age, gender, smoking and drinking status. Hardy–Weinberg equilibrium (HWE) was tested using a goodness-of-fit χ^2-test. The stratification analyses were performed according to age, gender, smoking status, drinking status, family history of cancer and clinical characteristic. Haploview 4.2 software was used to calculate the D′ and r^2 value for linkage disequilibrium (LD) between the three selected tagSNPs (18). Expression levels of *HOTAIR* in colorectal cancer tissues were analysed using the paired *t*-test. Student's *t*-test was performed to detect CCK-8 results between different plasmids. The full sequences of *HOTAIR* containing G or C alleles in rs7958904 were used to predict the folding structures of *HOTAIR* in RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (19). Putative influences of this variant on secondary structures of *HOTAIR* can be found in the result provided by website. $P < 0.05$ was considered statistically significant. All the statistical analyses were two-sided and were performed with the SAS 9.1 software (SAS Institute, Cary, NC, USA).

Results

Characteristic of the study population

The distribution of selected characteristics between colorectal cancer patients and control subjects is shown in Table I. In total, there were 1734 cases and 1855 controls included in this two-stage study. No significant differences were observed between cases and controls in age ($P = 0.910$) and gender ($P = 0.089$), indicating satisfactory matching by age and gender. We found that there were significant differences in smoking status between cases and controls ($P = 0.004$), while a difference in drinking status was not evident ($P = 0.110$). In the combined set, the low grade proportion of colorectal cancer patients was 8.5%, while the intermediate and high grade proportions were 77.7% and 13.8%, respectively. Approximately 7.8% of patients were in Dukes Stage A, 41.8% in Stage B, 38.1% in Stage C and 12.3% in Stage D.

Table I. Frequency distributions of selected variables between the colorectal cancer cases and controls

Variables	Stage 1		Stage 2		Combined[a]		P^b
	Cases, *N* (%)	Controls, *N* (%)	Cases, *N* (%)	Controls, *N* (%)	Cases, *N* (%)	Controls, *N* (%)	
Age (mean ± SD)	60.0 ± 12.6	59.9 ± 14.3	59.4 ± 11.7	59.6 ± 16.9	59.8 ± 12.3	59.8 ± 15.3	0.910
Sex							
Male	702 (61.2)	698 (58.0)	355 (60.5)	381 (58.4)	1057 (61.0)	1079 (58.2)	0.089
Female	445 (38.8)	505 (42.0)	232 (39.5)	271 (41.6)	677 (39.0)	776 (41.8)	
Smoking							
Never	736 (64.2)	811 (67.4)	378 (64.4)	465 (71.3)	1114 (64.2)	1276 (68.8)	0.004
Ever	411 (35.8)	392 (32.6)	209 (35.6)	187 (28.7)	620 (35.8)	579 (31.2)	
Drinking							
Never	823 (71.8)	898 (74.7)	401 (68.3)	456 (70.0)	1224 (70.6)	1354 (73.0)	0.110
Ever	324 (28.2)	305 (25.3)	186 (31.7)	196 (30.0)	510 (29.4)	501 (27.0)	
FH							
No	904 (78.8)	1076 (89.4)	484 (82.5)	611 (93.7)	1388 (80.1)	1687 (90.9)	<0.001
Yes	243 (21.2)	127 (10.6)	103 (17.5)	41 (6.3)	346 (19.9)	168 (9.1)	
Tumour site							
Colon	559 (48.7)		286 (48.7)		845 (48.7)		
Rectal	588 (51.3)		301 (51.3)		889 (51.3)		
Tumour Grade							
Low	85 (7.4)		63 (10.7)		148 (8.5)		
Intermediate	880 (76.7)		467 (79.6)		1347 (77.7)		
High	182 (15.9)		57 (9.7)		239 (13.8)		
Duke stage							
A	97 (8.4)		38 (6.4)		135 (7.8)		
B	494 (43.1)		231 (39.4)		725 (41.8)		
C	422 (36.8)		238 (40.6)		660 (38.1)		
D	134 (11.7)		80 (13.6)		214 (12.3)		

SD, standard deviation, FH, family history of cancer.

[a]Combine Stage 1 and Stage 2 as the combined stage.

[b]Two-sided χ^2 test for the frequency distributions of selected variables between the cases and controls.

Effects of tagSNPs in *HOTAIR* and colorectal cancer risk

The genotype distributions in the three SNPs were consistent with those expected from HWE (P = 0.260 for rs4759314, P = 0.147 for rs7958904 and P = 0.749 for rs874945, respectively). The genotype distribution of all the tagSNPs and their associations with colorectal cancer risk in our Stage 1 are shown in Table 2. We used ORs and 95% CIs to assess the association of *HOTAIR* SNPs with colorectal cancer risk, in codominant, dominant and additive models. Results of single loci analyses revealed that the genotype frequencies of rs7958904 were significantly different between cases and controls (P = 0.009) and carriers of CC genotype of this loci had a distinct decreased risk of colorectal cancer (adjusted OR = 0.70, 95% CI = 0.51–0.97, P = 0.034), compared to those carrying GG genotype. When we combined GC and CC genotype to construct a dominant model, a significant decreased risk was also found in GC/CC genotypes, with adjusted OR = 0.82, 95% CI = 0.69–0.96 and P = 0.016. Moreover, the frequency of the rs7958904 C allele was lower among the cases than among the controls (P = 0.009). For rs4759314, the subjects carrying the GG genotype had a significantly lower incidence of colorectal cancer than those carrying AA genotype (adjusted OR = 0.10, 95% CI = 0.01–0.79). No notable associations between rs874945 and colorectal cancer risk in the Stage 1 were observed. Therefore, we validated the effects of rs4759314 and rs7958904 in the Stage 2, to further confirm the association observed in the first stage.

As shown in Table 3, the association of rs4759314 disappeared in the Stage 2 (P = 0.355 for GG vs. AA). However, the protective effects of rs7958904 still existed in the Stage 2. When the wild homozygote GG was taken as the reference, the rs7958904 CC genotype and the combined GC/CC genotypes were associated with a significantly decreased colorectal cancer risk (OR = 0.58, 95% CI = 0.37–0.91 for CC; OR = 0.79, 95% CI = 0.63–0.99 for GC/CC). Additionally, in the analyses of additive model and allele frequency, pronounced differences were also found, which was consistent with the results of the Stage 1 (detailed data are shown in Table 3). Subsequently, we analysed the association between rs7958904 and colorectal cancer risk in the combined stage and found significant associations with decreased colorectal cancer risk in the codominant, dominant and additive model, and the distribution of allele frequency was also different between cases and controls (Table 3).

Stratified analysis of rs7958904 in *HOTAIR*

We further evaluated effects of rs7958904 stratified by selected characteristics, e.g. age, gender and smoking status. As shown in Supplementary Table 2, available at *Mutagenesis* Online, the protective effect of rs7958904 was more remarkable in subgroups of older subjects (age > 60) (P = 0.001), females (P = 0.002), never smokers (P = 0.002), never drinkers (P = 0.003) and subjects without a family history of cancer ($P = 1.00 \times 10^{-4}$). It should be interpreted that the cutoff value of age was selected according to the mean age of cases and controls, i.e. 60. We assumed that the two groups with age ≤60 and >60 can be representative of younger and older subjects in our present study, respectively.

Subsequently, we conducted stratified analyses according to clinical features of colorectal cancer, i.e. tumour site, tumour grade and Dukes stage. We found the difference in rs7958904 genotype frequency between cases and controls was more evident in patients with an intermediate grade (adjusted OR = 0.79, 95% CI = 0.69–0.91) (data shown in Supplementary Table 3, available at *Mutagenesis* Online).

Table 2. Association of the selected SNPs with colorectal cancer risk in Stage 1

Genotype	Cases		Controls		Crude OR (95% CI)	Adjusted OR (95% CI)[a]			
	N	%	*N*	%			P^a	$P_{Additive}{}^a$	$P_{Allele}{}^b$
rs4759314									
AA	1011	88.1	1037	86.2	1.00	1.00		0.075	0.073
AG	135	11.8	157	13.1	0.88 (0.69–1.13)	0.89 (0.69–1.14)	0.365		
GG	1	0.1	9	0.7	0.11 (0.01–0.90)	0.10 (0.01–0.79)	0.029		
AG/GG	136	11.9	166	13.8	0.84 (0.66–1.07)	0.84 (0.66–1.08)	0.176		
rs7958904									
GG	672	58.7	646	53.8	1.00	1.00		0.009	0.009
GC	399	34.9	456	38.0	0.84 (0.71–0.99)	0.84 (0.70–1.00)	0.051		
CC	74	6.4	99	8.2	0.72 (0.52–0.99)	0.70 (0.51–0.97)	0.034		
GC/CC	473	41.3	555	46.2	0.82 (0.70–0.97)	0.82 (0.69–0.96)	0.016		
rs874945									
GG	751	65.5	817	68.0	1.00	1.00		0.235	0.225
AG	356	31.0	346	28.8	1.21 (0.94–1.34)	1.14 (0.95–1.36)	0.167		
AA	40	3.5	39	3.2	1.12 (0.71–1.76)	1.06 (0.67–1.69)	0.795		
AG/AA	396	34.5	385	32.0	1.12 (0.94–1.33)	1.13 (0.95–1.34)	0.179		

[a]Adjusted by age, sex, smoking and drinking status in logistic regression analysis.
[b]Two-sided χ^2 test for allele frequencies between the cases and controls.

Table 3. Association of rs4759314 and rs7958904 with colorectal cancer risk in Stage 2 and combined stage

Genotype	Cases		Controls		Crude OR (95% CI)	Adjusted OR (95% CI)			
	N	%	*N*	%			P^a	$P_{Additive}{}^a$	$P_{Allele}{}^b$
Stage 2									
rs4759314									
AA	517	88.2	571	87.6	1.00	1.00		0.551	0.889
AG	65	11.1	79	12.1	0.91 (0.64–1.29)	0.93 (0.66–1.32)	0.690		
GG	4	0.7	2	0.3	2.20 (0.40–12.09)	2.24 (0.41–12.41)	0.355		
AG/GG	69	11.8	81	12.4	0.94 (0.67–1.33)	0.97 (0.68–1.36)	0.842		
rs7958904									
GG	347	59.2	346	53.1	1.00	1.00		0.032	0.009
GC	206	35.2	248	38.1	0.83 (0.65–1.05)	0.84 (0.66–1.07)	0.153		
CC	33	5.6	57	8.8	0.58 (0.37–0.91)	0.58 (0.37–0.91)	0.018		
GC/CC	239	40.8	305	46.9	0.78 (0.62–0.98)	0.79 (0.63–0.99)	0.042		
Combined stage									
rs7958904									
GG	1019	58.9	992	53.6	1.00	1.00		0.002	2.56×10^{-4}
GC	605	35.0	704	38.0	0.84 (0.73–0.96)	0.83 (0.72–0.96)	0.011		
CC	107	6.1	156	8.4	0.67 (0.51–0.87)	0.67 (0.51–0.87)	0.003		
GC/CC	712	41.1	860	46.4	0.81 (0.71–0.92)	0.80 (0.70–0.92)	0.001		

[a]Adjusted by age, sex, smoking and drinking status in logistic regression analysis.
[b]Two-sided χ^2 test for allele frequencies between the cases and controls.

In silico prediction of rs7958904

Furthermore, we also performed *in silico* analyses to predict the influence of G/C variant in rs7958904 on the secondary structure of *HOTAIR*. As shown in Figure 2, the secondary structure was remarkably changed with rs7958904 G/C variant.

HOTAIR expression level in colorectal cancer tissues

To confirm the abnormal expression of *HOTAIR* in Chinese colorectal cancer patients, we evaluated the *HOTAIR* level in 95 paired colorectal cancers and their corresponding normal tissues. As a result, *HOTAIR* was upregulated in cancer tissues and the difference was statistically significant ($P = 0.015$, Figure 3A), suggesting that *HOTAIR* may act as a motivator in carcinogenesis of colorectal cancer.

Effects of *HOTAIR* rs7958904 on cell proliferation

We assessed the proliferation rate of LoVo cells with *HOTAIR*-G and *HOTAIR*-C plasmids, to determine the functional role of rs7958904 in the carcinogenesis of colorectal cancer. After 24, 48 and 72 h of transfection, results of the CCk-8 assay revealed an evident decrease in cell growth of *HOTAIR*-C cells, compared with that of *HOTAIR*-G cells. The decrease was statistically different in all the subgroups ($P < 0.001$ for 24 and 48 h, $P = 0.013$ for 72 h), indicating a growth-inhibiting role of rs7958904 C allele on colorectal cancer cells (Figure 3B and Supplementary Table 4, available at *Mutagenesis* Online).

Discussion

HOTAIR is a HOX transcript antisense RNA and has been widely regarded as a functional lncRNA participating in multiple cancers.

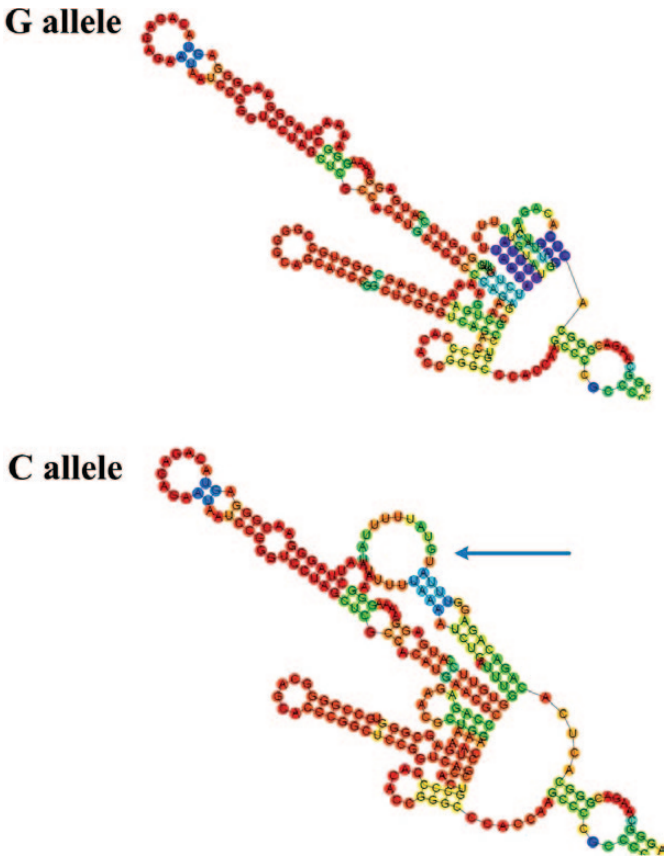


Figure 2. *In silico* analyses predicting secondary structure of *HOTAIR* with RNAfold. The arrow indicates alteration in structures caused by rs7958904 G/C variant.

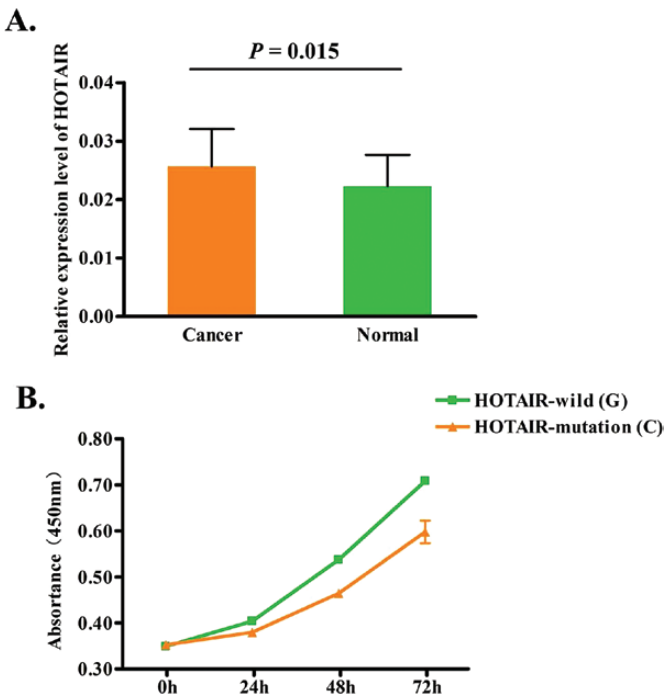


Figure 3. (**A**) Expression level of *HOTAIR* in colorectal cancer tissues was significantly elevated than in adjacent normal tissues. (**B**) Effects of *HOTAIR*-G or *HOTAIR*-C overexpression on colorectal cancer cell proliferation, detected by CCK-8 assay. Proliferation of cells transfected with *HOTAIR*-C plasmid was significantly decreased compared with those transfected with *HOTAIR*-G plasmid. Detailed data are shown in Supplementary Table 4, available at *Mutagenesis* Online.

It is located within the Homeobox C (HOXC) gene cluster on chromosome 12. Studies have demonstrated that *HOTAIR* may interact with the polycomb repressive complex 2 (PRC2) (11) as well as the lysine specific demethylase 1 (LSD1) complexes (20). Taken together, it can act as a scaffold to assemble PRC2 and LSD1 complexes, and cause epigenetic silencing of various cancer-related genes, particularly the HOXD gene, by coupling histone H3K27 methylation and H3K4 demethylation (10,21). Drawn by the remarkable effects of *HOTAIR* on epigenetic regulation in genome-wide level, researchers have focused on the dysregulation of *HOTAIR* in various types of cancer. A Japanese study revealed that *HOTAIR* was overexpressed in colorectal cancer tissues and also that colorectal cancer patients with higher *HOTAIR* level had a worse prognosis than those with lower *HOTAIR* level (9). Thus, it is not surprising that *HOTAIR* is involved in carcinogenesis of colorectal cancer, possibly through regulation of genes related with colorectal cancer. In addition, the promoting role of *HOTAIR* in carcinogenesis has been identified in other cancers. Gupta *et al.* (11) demonstrated firstly that *HOTAIR* was significantly elevated in primary breast cancer and high level of *HOTAIR* was a predictor of metastasis of breast cancer. *In vitro* experiments showed that *HOTAIR* may promote invasion and metastasis of breast cancer cells (11). Subsequently *HOTAIR* was found to participate in a series of human cancers, including hepatocellular carcinoma (22), pancreatic cancer (23) and non-small cell lung cancer (24).

Emerging evidence has shown that genetic variants in lncRNAs may modulate individual susceptibility to cancer (14,25). SNPs in lncRNAs exert various effects on their expressions and functions (26). To date, effects of *HOTAIR* SNPs in human cancer have only been reported by Zhang *et al.* (27). Based on the observation that *HOTAIR* was significantly upregulated in colorectal cancer tissues, we speculated that genetic variant in *HOTAIR* may participate in colorectal cancer by affecting gene expression or functions.

We selected tagSNPs in *HOTAIR* gene region, and assessed the association between these genetic variants and susceptibility in a two-stage case–control study. Nowadays, the two-stage case–control study is a widely accepted design propocol with significantly reduced false positive rates compared with the traditional one-stage studies. In general, any promising associations observed in Stage 1 are re-evaluated in Stage 2. Associations replicated in Stage 2 are less likely due to chance. The importance of replication using independent samples to reduce type 1 errors in association studies has been demonstrated by other studies. In our study, results of both Stage 1 and independent Stage 2 showed an evident association between decreased risk of colorectal cancer and genetic variants in rs7958904. This observation provided solid confirmation for the protective role of rs7958904 in carcinogenesis of colorectal cancer. However, the effects of rs4759314 observed in first stage disappeared in the second stage. Given the small sample of rs4759314 GG carriers in Stage 1 (1 case vs. 9 controls), we speculated that the association of rs4759314 observed in Stage 1 may be a false positive result.

Subsequently, we performed stratified analyses of rs7958904 according to the epidemiological variables. We found that the decreased colorectal cancer risk associated with rs7958904 in dominant model was more pronounced in subgroups of older subjects, females, non-smokers, non-drinkers and subjects without a family history of cancer. These results indicated that effects of *HOTAIR* variant on colorectal cancer risk may be modulated by specific demographic factors as well as environmental exposures. A possible explanation is that some environmental hazards (e.g. tobacco smoking and alcohol drinking) have colorectal cancer promoting effects, which diminished the protective effects of variants in rs7958904. This is also evidence supporting that carcinogenesis is a complex process involving genetic and environmental factors. In the analyses stratified by clinical characteristics, we found the decreased risk was more significant in subjects with intermediate tumour grade. It is rational that different molecular mechanisms underlying different colorectal cancer grades (28) may interfere with the effects of *HOTAIR* SNPs and lead to our present result. Moreover, the relatively larger sample size in the group of intermediate grade may also lead to the result.

The location of rs7958904 was on the exon of *HOTAIR* gene. Therefore, we performed *in silico* analyses to predict the influence of the G/C variant in rs7958904 on the secondary structure of *HOTAIR*. As a result, the secondary structure was remarkably changed with rs7958904 G/C variant, indicating that the SNP may participate in colorectal cancer through alteration of *HOTAIR* structure. Further studies into the detailed biological mechanisms of *HOTAIR* SNPs are warranted.

In addition, we demonstrated that *HOTAIR* was notably upregulated in colorectal cancer tissues than in adjacent normal tissues, providing further evidence for the important biological role of *HOTAIR* in of colorectal carcinogenesis in the Chinese. Furthermore, our results suggest that G/C variation in rs7958904 may impair proliferation of colorectal cancer cells. We speculate that this may be the molecular mechanism underlying the association between rs7958904 and colorectal cancer risk that we observed in the present study.

There were some strengths of this study that should be noted. First, although the role of *HOTAIR* in carcinogenesis of multiple cancers has been reported, this is the first study investigating genetic variation of *HOTAIR* in relation to cancer susceptibility. Our observation may provide a novel insight into the role of *HOTAIR* in human cancers, especially colorectal cancer. The relatively large sample size and the two-stage study design, ensuring sufficient statistical power to detect subtle differences, is another strength of our study. The whole study provided strong evidence for the roles of *HOTAIR* SNPs in colorectal cancer. Third, our control group was selected from people who came to hospital for a routine examination, not for illness, which significantly diminished the effect of selection bias. Furthermore, the controls and the cases were matched on sex and age, and the genotype distributions in our control group were similar to those reported in public database. Therefore, we believed that selection bias was not substantial and not likely to influence the analyses of our study. There is also a limitation of our study. Many studies have focused on the environmental exposure of colorectal patients. In the present study, we only obtained information on smoking and drinking. Further studies should focus on more environmental risk factors, such as dietary habits of the subjects, to make the result more informative.

In conclusion, we identified a novel SNP located in *HOTAIR* gene that was significantly associated with decreased risk of colorectal cancer in our two-stage case–control study. Larger prospective studies are warranted to confirm our results.

Supplementary data

Supplementary Tables 1–4 is available at *Mutagenesis* Online.

Funding

This study was partly supported by National Natural Science Foundation of China (81373091, 81230068, 81201570); Key

Program for Basic Research of Jiangsu Provincial Department of Education (12KJA330002); Jiangsu Provincial Science and Technology Innovation Team; Priority Academic Program Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

Acknowledgements

Yao Xue, Lingjun Zhu and Meilin Wang were involved in conception and design; Gaoxiang Ma and Qiuhan Hua were involved in the acquisition of data; Yao Xue and Gaoxiang Ma were involved in analysis and interpretation of data; Yao Xue and Meilin Wang were involved in drafting of the manuscript; Zhengdong Zhang and Meilin Wang were involved in critical reversion of the manuscript; Haiyan Chu and Na Tong were involved in statistical analysis; Zhengdong Zhang and Meilin Wang were involved in obtaining the funds; sample support was given by Lingjun Zhu, Dongying Gu and Jingfei Chen; and supervised by Zhengdong Zhang.
Conflict of interest statement: None declared.

References

1. Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E. and Forman, D. (2011) Global cancer statistics. *CA Cancer J. Clin*., 61, 69–90.
2. Long, N., Moore, M. A., Chen, W. *et al*. (2010) Cancer epidemiology and control in north-East Asia—past, present and future. *Asian Pac. J. Cancer Prev*., 11(Suppl. 2), 107–148.
3. Black, R. E., Cousens, S., Johnson, H. L. *et al*. (2010) Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet*, 375, 1969–1987.
4. Moran, V. A., Perera, R. J. and Khalil, A. M. (2012) Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res*., 40, 6391–6400.
5. Costa, F. F. (2010) Non-coding RNAs: meet thy masters. *Bioessays*, 32, 599–608.
6. Huarte, M. and Rinn, J. L. (2010) Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet*., 19, R152–R161.
7. Ji, P., Diederichs, S., Wang, W., *et al*. (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, 22, 8031–8041.
8. Bussemakers, M. J., van Bokhoven, A., Verhaegh, G. W., Smit, F. P., Karthaus, H. F., Schalken, J. A., Debruyne, F. M., Ru, N. and Isaacs, W. B. (1999) DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res*., 59, 5975–5979.
9. Kogo, R., Shimamura, T., Mimori, K. *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.
10. Zhang, J., Zhang, P., Wang, L., Piao, H. L. and Ma, L. (2014) Long non-coding RNA *HOTAIR* in carcinogenesis and metastasis. *Acta Biochim. Biophys. Sin. (Shanghai)*., 46, 1–5.
11. Gupta, R. A., Shah, N., Wang, K. C. *et al*. (2010) Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature*, 464, 1071–1076.
12. Gaffney, D. J. (2013) Global properties and functional complexity of human gene regulatory variation. *PLoS Genet*., 9, e1003501.
13. Xue, Y., Wang, M., Kang, M. *et al*. (2013) Association between lncrna PCGEM1 polymorphisms and prostate cancer risk. *Prostate Cancer Prostatic Dis*., 16, 139–44, S1.
14. Liu, Y., Pan, S., Liu, L. *et al*. (2012) A genetic variant in long non-coding RNA HULC contributes to risk of HBV-related hepatocellular carcinoma in a Chinese population. *PLoS One*, 7, e35145.
15. Ma, L., Zhu, L., Gu, D., Chu, H., Tong, N., Chen, J., Zhang, Z. and Wang, M. (2013) A genetic variant in miR-146a modifies colorectal cancer susceptibility in a Chinese population. *Arch. Toxicol*., 87, 825–833.
16. Zhu, L., Du, M., Gu, D., Ma, L., Chu, H., Tong, N., Chen, J., Zhang, Z. and Wang, M. (2013) Genetic variant rs7758229 in 6q26-q27 is not associated with colorectal cancer risk in a Chinese population. *PLoS One*, 8, e59256.
17. Du, M., Liu, S., Gu, D. *et al*. (2014) Clinical potential role of circulating microRNAs in early diagnosis of colorectal cancer patients. *Carcinogenesis*. doi:10.1093/carcin/bgu189.
18. Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21, 263–265.
19. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. and Hofacker, I. L. (2008) The Vienna RNA websuite. *Nucleic Acids Res*., 36, W70–W74.
20. Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E. and Chang, H. Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 329, 689–693.
21. Spitale, R. C., Tsai, M. C. and Chang, H. Y. (2011) RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics*, 6, 539–543.
22. Geng, Y. J., Xie, S. L., Li, Q., Ma, J. and Wang, G. Y. (2011) Large intervening non-coding RNA *HOTAIR* is associated with hepatocellular carcinoma progression. *J. Int. Med. Res*., 39, 2119–2128.
23. Kim, K., Jutooru, I., Chadalapaka, G., Johnson, G., Frank, J., Burghardt, R., Kim, S. and Safe, S. (2013) *HOTAIR* is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. *Oncogene*, 32, 1616–1625.
24. Nakagawa, T., Endo, H., Yokoyama, M. *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.
25. Li, L., Sun, R., Liang, Y. *et al*. (2013) Association between polymorphisms in long non-coding RNA PRNCR1 in 8q24 and risk of colorectal cancer. *J. Exp. Clin. Cancer Res*., 32, 104.
26. Wu, H., Zheng, J., Deng, J., Hu, M., You, Y., Li, N., Li, W., Lu, J. and Zhou, Y. (2013) A genetic polymorphism in lincRNA-uc003opf.1 is associated with susceptibility to esophageal squamous cell carcinoma in Chinese populations. *Carcinogenesis*, 34, 2908–2917.
27. Zhang, X., Zhou, L., Fu, G. *et al*. (2014) The identification of an ESCC susceptibility SNP rs920778 that regulates the expression of lncRNA *HOTAIR* via a novel intronic enhancer. *Carcinogenesis*, 35, 2062–2067.
28. Compton, C. C. (2003) Colorectal carcinoma: diagnostic, prognostic, and molecular features. *Mod. Pathol*., 16, 376–388.