



SNP rs2071095 in LincRNA H19 is associated with breast cancer risk

Ping Cui¹ · Yanrui Zhao¹ · Xinlei Chu¹ · Na He² · Hong Zheng¹ · Jiali Han³ · Fengju Song¹ · Kexin Chen¹

Received: 3 January 2018 / Accepted: 3 May 2018 / Published online: 8 May 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Purpose An increasing number of long intergenic non-coding RNAs (lincRNAs) appear to play critical roles in cancer development and progression. To assess the association between SNPs that reside in regions of lincRNAs and breast cancer risk, we performed a large case-control study in China.

Methods We carried out a two-stage case-control study including 2881 breast cancer cases and 3220 controls. In stage I, we genotyped 17 independent ($r^2 < 0.5$) SNPs located in 6 tumor-related lincRNAs by using the TaqMan platform. In stage II, SNPs potentially associated with breast cancer risk were replicated in an independent population. Quantitative real-time PCR was used to measure H19 levels in tissues from 228 breast cancer patients with different genotypes.

Results We identified 2 SNPs significantly associated with breast cancer risk in stage I ($P < 0.05$), but not significantly replicated in stage II. We combined the data from stage I and stage II, and found that, compared with the rs2071095 CC genotype, AA and CA + AA genotypes were associated with significantly decreased risk of breast cancer (adjusted OR 0.83, 95% CI 0.69–0.99; adjusted OR 0.88, 95% CI 0.80–0.98, respectively). Stratified analyses showed that rs2071095 was associated with breast cancer risk in estrogen receptor (ER)-positive patients ($P = 0.002$), but not in ER-negative ones ($P = 0.332$). Expression levels of H19 in breast cancer cases with AA genotype were significantly lower than those with CC genotype.

Conclusions We identified that rs2071095 may contribute to the susceptibility of breast cancer in Chinese women via affecting H19 expression. The mechanisms underlying the association remain to be investigated.

Keywords Breast cancer · H19 · LincRNA · Genetic susceptibility · SNP

✉ Fengju Song
songfengju@163.com

✉ Kexin Chen
chenkexin1963@yahoo.com

¹ Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy in Ministry of Education, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, People's Republic of China

² Department of Cancer Biobank, Key Laboratory of Cancer Prevention and Therapy of Tianjin, Tianjin's Clinical Research Center for Cancer, National Clinical Research Centre of Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, People's Republic of China

³ Department of Epidemiology, Fairbanks School of Public Health, Indiana University Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, IN, USA

Introduction

Breast cancer is the most frequently diagnosed cancer that accounts for 15% of all new cancer cases among women in China, and is the leading cause of cancer death in women younger than 45 years [1]. Over the last decades, genome-wide association studies (GWAS) have linked 3800 SNPs to 427 diseases and traits, only 7% of the SNPs are located in protein-coding regions, but with 93% located in non-coding regions [2–4].

LincRNAs are non-coding transcripts longer than 200 nucleotides, originating from the region between two protein-coding genes and have been described as the largest subclass in the non-coding transcriptome in human [5]. LincRNAs are emerging as key regulators of diverse biological and cellular processes through mechanisms such as transcriptional and post-transcriptional processing [6–11]. Gupta, R. A. et al. found that HOTAIR could reprogram chromatin state to promote breast cancer metastasis [12]. Studies from Subhrangsu S. Mandal's laboratory show that

HOTAIR is crucial for the viability of breast cancer cells and its expression is transcriptionally regulated in an estradiol (E2)-dependent manner via coordination of estrogen receptors (ER) and ER coregulators [13, 14]. Maite Huarte et al. identified that lincRNA-p21 could serve as a repressor in p53-dependent transcriptional responses; the inhibition of lincRNA-p21 could affect the expression of hundreds of gene targets enriched for genes normally repressed by p53 [15]. Recently, increasing number of researchers identified that genetic variation in lincRNAs may confer susceptibility to cancer by causing aberrant expressions and dysregulations in a cell-type or tissue-specific manner. Bayram S et al. first found that CC genotype of HOTAIR rs920778 polymorphism might play important roles in genetic susceptibility to breast cancer development and aggressiveness in a Turkish population [16]. Peng R et al. identified that tag SNPs (rs3200401, rs619586) of lincRNA MALAT1 were associated with the susceptibility of breast cancer via the alteration of mRNA expression level [17]. MALAT1 overexpression predicted poor survival in tamoxifen-treated ER-positive breast cancer patients, which might serve as a potential biomarker to predict endocrine treatment sensitivity [18]. Genetic variants in the INK4 locus serve as a hotspot for various diseases [13, 19], Royds JA et al. reported that rs11515 was associated with aggressive breast tumors with increased ANRIL and reduced p16 (INK4a) expression [20]. Riaz M et al. found that tag SNP rs2107425 was significantly associated with shorter metastasis-free survival but not through altering H19 mRNA expression [21]. However, most of these studies included only a few SNPs of the candidate lincRNA, and several lincRNAs, such as lincRNA-p21, HULC, and HOTTIP, have not been studied for their association with breast cancer susceptibility.

Systematically identifying such loci involved in breast cancer will be important in understanding breast cancer pathogenesis and in establishing new targets for cancer screening, prevention, and therapy. Therefore, we performed a case-control study to analyze the association between lincRNA SNPs and breast cancer risk.

Materials and methods

Ethic statement

The research protocol was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital, and written informed consents were obtained from all subjects participated in the study.

Study subjects

We conducted a two-stage case-control study in a combined dataset of 6101 individuals (2881 breast cancer cases and 3220 controls), who were randomly assigned to stage I and stage II according the ratio of 1:1. The sample size was calculated by “QUANTO” software under the condition of the minor allele frequency (MAF) ≥ 0.1 using parameters: $\alpha = 0.05$ (two-side test), $1 - \beta = 80\%$, $OR \geq 1.2$. All the patients were newly diagnosed and histologically confirmed breast cancer between 2006 and 2009 in Tianjin Medical University Cancer Institute and Hospital. Controls (genetically unrelated women) were enrolled from the nearby community without personal malignant tumor history and were frequency matched to the cases by age (± 5 year) at the same time. The detailed description of cases and controls can be found in our previous study [22]. Participants were excluded if they had a personal history of cancer, or if they had history of blood transfusion within 6 months as they donated blood. We also excluded male patients. The study comprised two stages. In stage I, 1492 patients and 1677 controls were randomly selected for SNP screening. In stage II, the candidate SNPs identified in the first stage were validated and evaluated; the validation set consists of 1389 cases and 1543 controls.

SNP selection

We used “lincRNA” as key word on PubMed in 2012, and we selected the well-known, functionally characterized and cancer-related lincRNAs; 40 papers were showed up with the inclusion of 7 lincRNAs (HOTAIR, lincRNA-p21, MALAT1, ANRIL, H19, and HOTTIP). The dbSNP database (<https://www.ncbi.nlm.nih.gov/projects/SNP/>) was used to select SNPs with the following criteria: SNPs have Chinese or Asian population frequency, the minor allele frequency (MAF) ≥ 0.1 . We then performed linkage disequilibrium analysis, and a total of 6 lincRNAs, 17 ($r^2 < 0.5$) independent SNPs were selected in this study.

SNP genotyping

Genomic DNA was extracted from peripheral blood using QIAGEN DNA Extraction Kit (QIAGEN Inc.). Taqman platform was used to perform genotyping. Primers and probes were supplied by Applied Biosystems. The 5- μ l reaction mixture contained 20 ng genomic DNA, 2.5 μ l $2 \times$ TaqMan Genotyping Master Mix, 0.1 μ l probe, and 1.7 μ l double distilled water. The PCR conditions used were as follows: 50 °C for 2 min, 95 °C for 10 min, and 60 °C for 1 min for 40 cycles. Amplifications were conducted using the 384-well plates, and

allelic discrimination was performed using SDS 2.4 software (Applied Biosystems, Foster City, CA, USA). We also performed negative controls (water) and positive controls with known genotype to ensure the accuracy. After 2 rounds of genotyping, the successful rate was 99%, and approximately 5% of the samples were randomly selected for replication; the final concordance rate for quality control was 100%.

mRNA expression profiling

Total RNA, isolated from 228 liquid nitrogen-frozen breast cancer tissues with known H19 genotypes (CC, CA or AA), was extracted and purified Trizol reagents (Invitrogen, USA) according to the manufacturer's protocol. M-MLV Reverse Transcriptase (Applied Biosystems, USA) was applied for reverse transcript. Quantitative real-time PCR was performed on ABI Prism 7900HT Fast Real-time PCR System (Applied Biosystems, USA). Probe for H19 used Hs00262142_g1 (TaqMan® Gene Expression Assay). GAPDH was used as an endogenous control. All samples were done in triplicate. After quality control (CV < 15%, $Ct_{GAPDH} \leq 28$), 213 breast cancer tissues were included in the expression analysis. The relative expression of H19 was calculated using the $2^{-\Delta\Delta Ct}$ method.

Statistical analysis

The Hardy–Weinberg equilibrium was assessed for controls by a χ^2 goodness of fit test. The distributions of the SNPs' genotypes, demographic variables, and other risk factors between breast cancer cases and controls were evaluated by performing χ^2 test. The association between the SNPs and breast cancer risk was evaluated by unconditional logistic regression, estimating the odds ratios (ORs) and 95% confident intervals (CIs), with and without adjustment of age, smoking status, menopause status, oral contraception use, history of benign breast diseases, and family history of cancer. We further analyzed the association of rs2071095 genotypes with demographic, lifestyle, and clinical characteristics by unconditional logistic regression. H19 relative expression in different genotypes was calculated using one-way ANOVA method. All statistical tests were two-sided, and $P < 0.05$ was considered significant; correction for multiple comparisons was not performed. We used SPSS 20.0 (SPSS Inc., Chicago, IL, USA) for all statistical analyses, ggplot2 and forestplot packages from R 3.4.2 for Figs. 1, 2, and 3 and GraphPad Prism 5.0 for Fig. 4.

Results

A total of 2881 patients with pathologically confirmed breast cancer and 3220 tumor-free controls were included in stage I and II. Their demographic characteristics are summarized in

Table 1. The ages of the participants were matched between cases and controls ($P = 0.808$). The differences between cases and controls in terms of smoking status ($P < 0.001$), oral contraceptive usage ($P < 0.001$), history of benign breast disease ($P < 0.001$), and family history of cancer ($P < 0.001$) were statistically significant. We further analyzed the effects of the rs2071095 genotypes on the risk of breast cancer among different subgroups of demographic characteristics (Figs. 1, 2, 3). For AA genotype, decreased risks of breast cancer were more likely to be evident in subgroups of younger patients (age ≤ 55 years, stage I+II: adjusted OR 0.86, 95% CI 0.68–0.98; stage I: adjusted OR 0.63, 95% CI 0.45–0.87), non-smoking women (stage I+II: adjusted OR 0.88, 95% CI: 0.74–0.98; stage I: adjusted OR 0.63, 95% CI 0.48–0.84), premenopausal women (stage I+II: adjusted OR 0.76, 95% CI 0.57–0.96; stage I: adjusted OR 0.57, 95% CI 0.39–0.84), non-use of oral contraception (stage I: adjusted OR 0.65, 95% CI 0.48–0.87), non-history of benign breast disease (stage I+II: adjusted OR 0.83, 95% CI 0.67–0.98; stage I: adjusted OR 0.65, 95% CI 0.48–0.86), and non-family history of cancer (stage I+II: adjusted OR 0.82, 95% CI: 0.65–0.97; stage I: adjusted OR 0.56, 95% CI 0.41–0.77). The results of stratified analysis in stage II were not significant. The interactions between genotype and demographic characteristics were not significant in stage I, stage II, nor stage I+II.

In stage I, the genotype and allele frequencies of the 17 candidate SNPs and their associations with breast cancer risk are shown in Table 2. The distributions of the genotype in controls were in line with the Hardy–Weinberg equilibrium. The significant associated SNPs were rs2071095 located in *H19* at 11p15.5 and rs2151280 located in *ANRIL* at 9p21.3.

The above 2 SNPs identified in stage I were replicated in an independent set. Results for the SNPs association in stage II were not significant at $P < 0.05$. We combined the data from stage I and stage II, and found that, compared with rs2071095 CC genotype, AA and CA + AA genotypes were associated with significantly decreased breast cancer risk (OR 0.82, 95% CI 0.68–0.98; OR 0.87, 95% CI 0.79–0.97, respectively), which were still significantly associated with breast cancer risk after adjusted by age, menopausal status, smoking status, family history of the benign breast disease, and family history of cancer (AA vs CC: OR 0.83, 95% CI 0.69–0.99; CA + AA vs CC: OR 0.88, 95% CI 0.80–0.98). We further assessed the associations between rs2071095 and breast cancer risk stratified by ER and PR status. The association was still significant for ER-positive breast cancer (OR 0.83, 95% CI 0.74–0.94, adjusted OR 0.86, 95% CI 0.76–0.98, $P = 0.002$), but not for ER-negative breast cancer (OR 0.93, 95% CI 0.81–1.07, adjusted OR 0.94, 95% CI 0.81–1.09, $P = 0.332$), though p for heterogeneity was

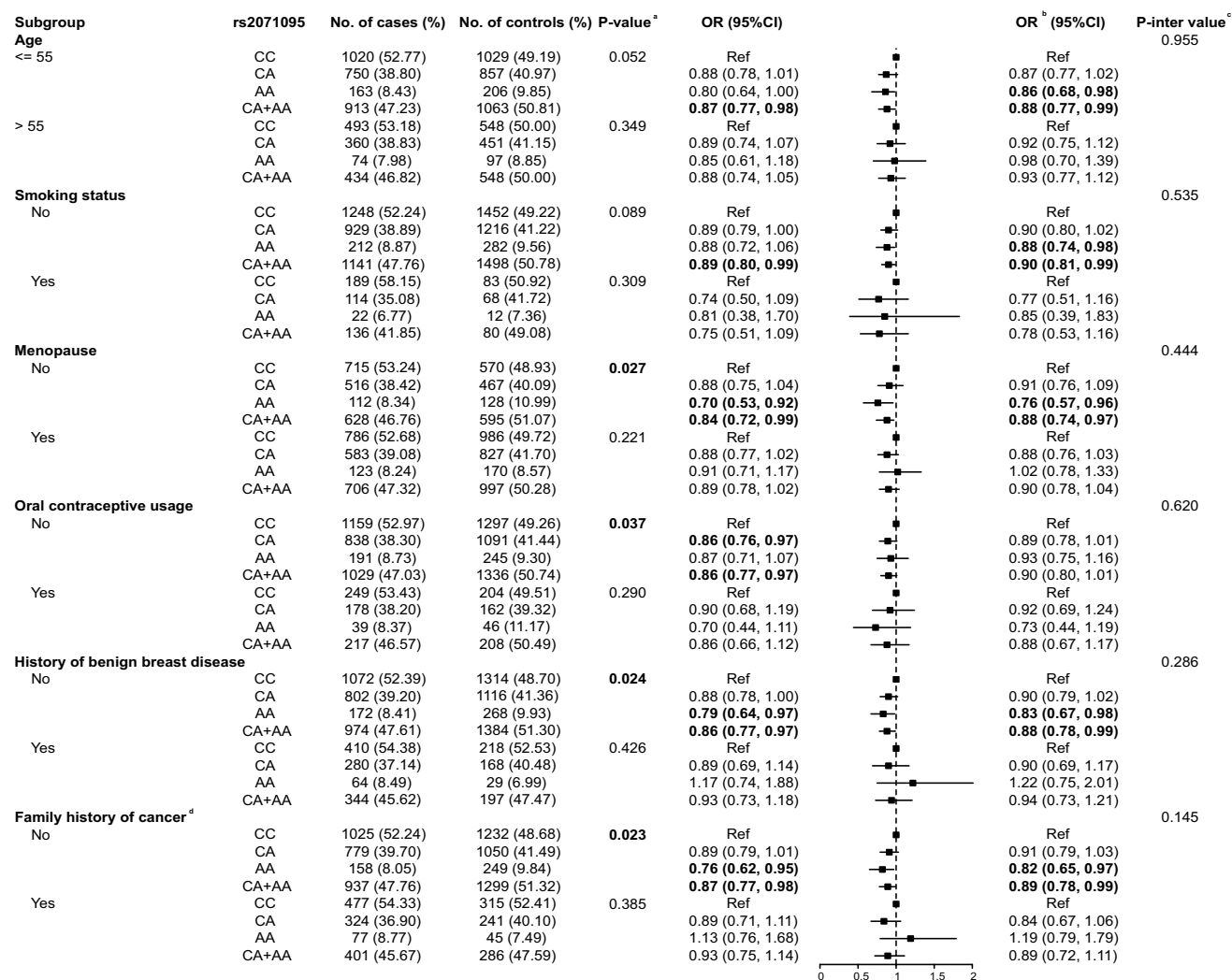


Fig. 1 Logistic regression analysis of associations between rs2071095 and the risk of breast cancer stratified by demographic characteristics (stage I and stage II). ^atwo-side χ^2 test for the difference in frequency distribution of the three genotypes (CC, CA, and AA) between cancers and controls and we did not show the difference in frequency distribution of the genotypes (CC, CA+AA) between cancers and

controls; $P < 0.05$ was considered statistically significant; ^badjusted by age, menopause status, oral contraceptive usage, smoking status, history of benign breast disease, and family history of cancer; ^cinteraction analysis between demographic variables and rs2071095; ^dfirst- and second-degree relatives with history of cancer

not significant. For PR status, both PR+ and PR- were significantly associated among those with the CA+AA genotypes (unadjusted OR for PR+: 0.87 and 95% CI 0.76–0.99; unadjusted OR for PR-: 0.87 and 95% CI 0.77–0.99) although the adjusted OR/95% CIs were not significant (Table 3).

We further explored the H19 expression by rs2071095 genotypes among 213 breast cancer tissues. As shown in Fig. 4, rs2071095 CA and AA genotypes were found to be significantly associated with decreased expression of H19 in breast cancer patients compared with the CC genotype ($P = 0.027$ and $P < 0.001$, respectively).

Discussions

Through a two-stage case-control study including 2881 breast cancer cases and 3220 controls, we identified a significant association with breast cancer risk for rs2071095 (A allele) in *H19*, which was significantly associated with decreased expression of H19 in breast cancer patients, indicating a novel susceptibility locus for breast cancer.

The *H19* gene, located in human chromosome 11p15.5 locus, is one of the first discovered paternally imprinted and maternally expressed genes [23]. It generates a highly

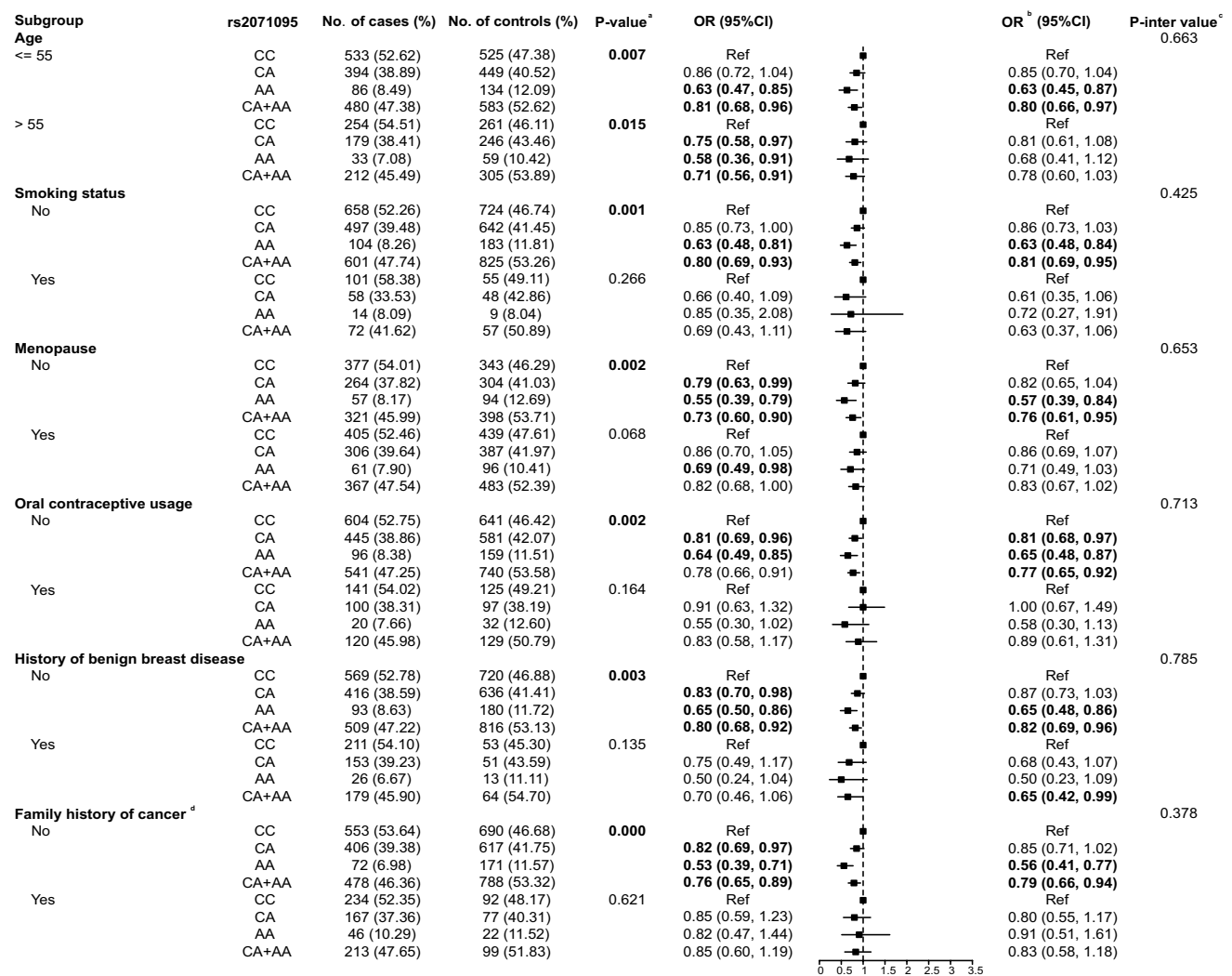


Fig. 2 Logistic regression analysis of associations between rs2071095 and the risk of breast cancer stratified by demographic characteristics (stage I). ^atwo-side χ^2 test for the difference in frequency distribution of the three genotypes (CC, CA, and AA) between cancers and controls and we did not show the difference in frequency distribution of the genotypes (CC, CA + AA) between cancers and controls; $P < 0.05$

conserved 2.3 kb capped, spliced and polyadenylate non-coding RNA, which location is mainly cytoplasmic [24–26]. *H19* abundantly expresses RNA during embryogenesis, playing a pivotal role in embryonic development, and is downregulated postnatally [27]. *H19* can function as a primary miR-675 precursor and suggest that *H19* expression could post-transcriptionally downregulate specific mRNAs [28]. *H19* might also act as competitive endogenous RNA [29]. *H19* overexpression could result in the activation of genes involved in angiogenesis, cell viability, and proliferation, triggering malignancies such as breast, lung, gastric, bladder, and pancreatic carcinomas [26, 30–32].

was considered statistically significant; ^badjusted by age, menopausal status, oral contraceptive usage, smoking status, history of benign breast disease and family history of cancer; ^cinteraction analysis between demographic variables and rs2071095; ^dfirst- and second-degree relatives with history of cancer

Curgy JJ et al. reported that overexpression of the *H19* gene was significantly correlated with the presence of steroid receptors and controlled by steroid hormones [33]. A recent study also identified that *H19* expression had a positive correlation with ER α expression in breast cancer, the estrogen-ER α -*H19* signaling axis plays a role in regulating the proliferation and differentiation potentials of the normal luminal progenitors and that this signaling network may also be important in the development of ER(+) breast cancer tumors [34]. Our results add another aspect to the above findings that the “A” allele of *H19* had an ER-positive-specific association with breast cancer risk. In recent years, studies have shown that trait-associated SNPs are concentrated in

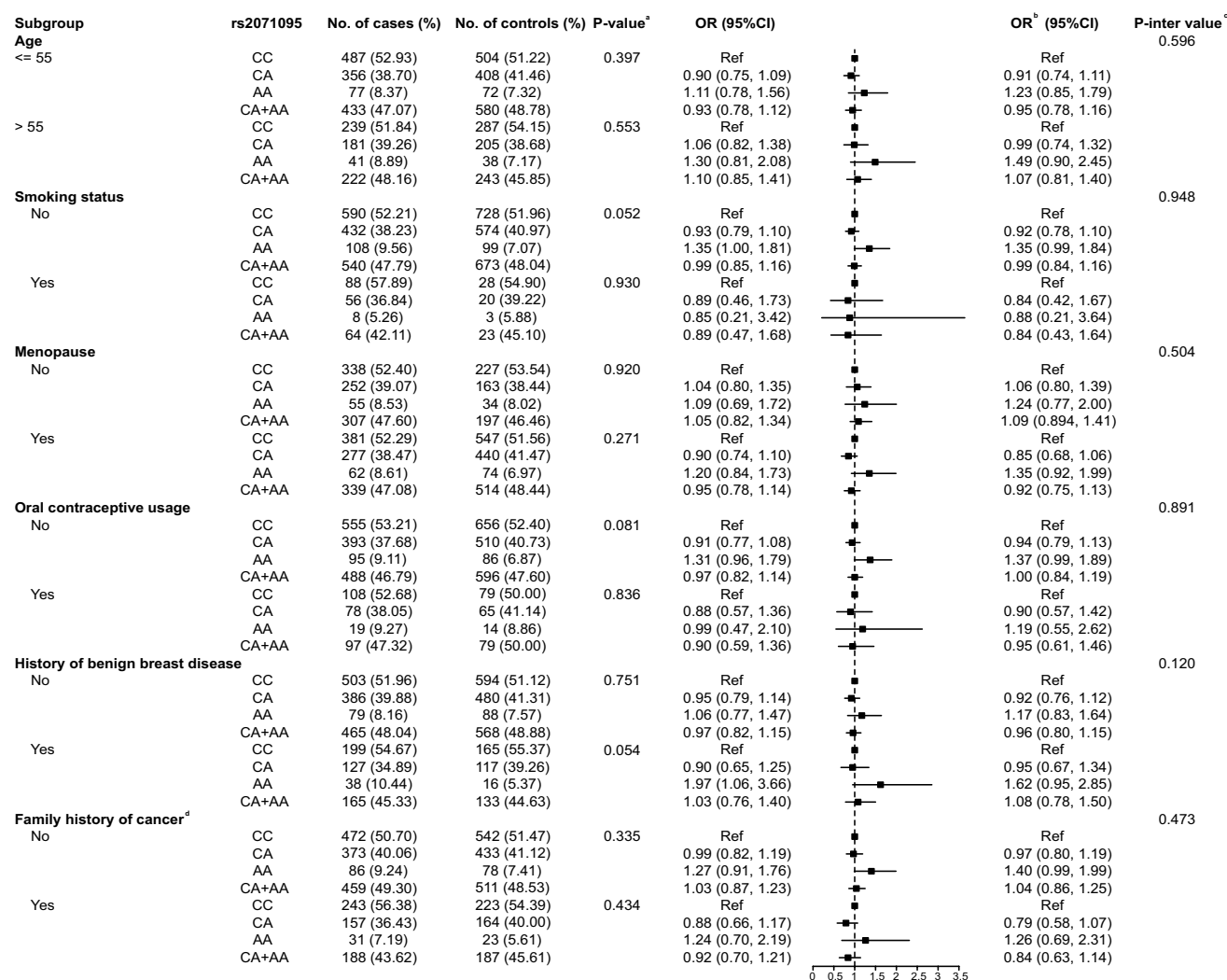


Fig. 3 Logistic regression analysis of associations between rs2071095 and the risk of breast cancer stratified by demographic characteristics (stage II). ^atwo-side χ^2 test for the difference in frequency distribution of the three genotypes (CC, CA, and AA) between cancers and controls and we did not show the difference in frequency distribution of the genotypes (CC, CA + AA) between cancers and controls; $P < 0.05$

regulatory regions and can perturb transcription factor recognition to these regions and thus conferring allele-specific dysregulation of the SNP-associated gene [35–39]. SNP rs2071095 is located about 2 kb upstream of H19 transcriptional start site, and is likely to modulate transcription factor binding to this region, resulting in “A” allele-specific down-regulation of the H19 expression in breast cancer tissues in our study. However, the underlying mechanism remains to be elucidated and may be exploited for therapeutic benefit.

In the present study, we included SNP rs2151280 in *ANRIL*, which has been reported to affect susceptibility to basal cell carcinoma [40]. *ANRIL* (officially known as *CDKN2B* antisense RNA 1 (*CDKN2B-AS1*)) is encoded in the chr9:p21 region at the *INK4* locus. Genetic

was considered statistically significant; ^badjusted by age, menopausal status, oral contraceptive usage, smoking status, history of benign breast disease, and family history of cancer; ^cinteraction analysis between demographic variables and rs2071095; ^dfirst- and second-degree relatives with history of cancer

susceptibility locus in *ANRIL* could serve as a hotspot for multiple diseases like cardiovascular disease, cancer, and diabetes [41–46]. *ANRIL* is overexpressed in a variety of cancers including leukemia, breast cancer, and prostate cancer [13], and controls cell proliferation and senescence via interacting with CBX7 (a PRC1 component) and SUZ12 (a PRC2 component) to form heterochromatin surrounding the *INK4b-ARF-INK4a* locus, leading to induce gene silencing [47–50]. The knockdown of *ANRIL* could lower multidrug resistance [51], decrease proliferation [52], and inhibit invasiveness [53]. However, we identified rs2151280 as a genetic susceptibility to breast cancer only in stage I, and could not validate in stage II.

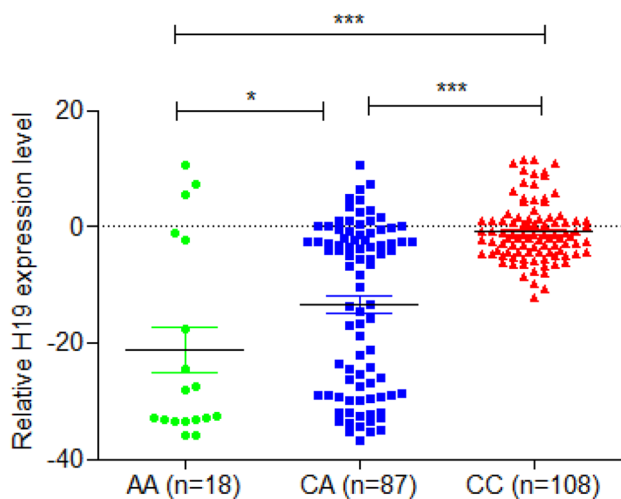


Fig. 4 Different genotype at rs2071095 affects H19 expression level. H19 levels determined by qRT-PCR in breast cancer tissues in subjects with the AA ($n=18$), CA ($n=87$) or CC ($n=108$) genotype at rs2071095. Results are shown as means \pm s.e.m. relative to GAPDH levels, and all P values are from two-sided t tests. * $P < 0.05$, *** $P < 0.001$

Although we take a large systematic case-control study to detect the breast cancer risk associated SNPs in lincRNAs. Our study has several limitations. Firstly, we only selected independent ($r^2 < 0.5$) SNPs for stage I, because of limitation of scope, we need to perform fine-mapping of a larger region centered on rs2071095 to analyze in depth. Secondly, we only included the SNPs with minor allele frequency (MAF) ≥ 0.1 , inevitably miss the low frequency breast cancer risk associated SNPs. Thirdly, we did not perform multiple comparisons correction, although we performed the large sample size and an independent validation set to ensure the replicability of our finding, and the expression data by genotype lend another support to the association.

In conclusion, our study has identified a common variant in H19 that is implicated in susceptibility to breast cancer among Chinese women, which remains to be prioritized for future analysis in tumorigenicity assays.

Table 1 Demographic characteristics of breast cancer cases and controls (Stage I and II)

Variables	No. (%)		P value ^a	OR (95% CI)
	Controls ($n=3220$)	Cases ($n=2881$)		
Age				
≤ 55	2117 (65.75)	1903 (66.05)	0.808	
> 55	1103 (34.25)	978 (33.95)		
Smoking status ^b				
No	2979 (94.81)	2393 (88.01)	< 0.001	1.00
Yes	163 (5.19)	326 (11.99)		2.49 (2.05, 3.03)
Menopause ^b				
No	1174 (36.94)	1347 (47.43)	< 0.001	1.00
Yes	2004 (63.06)	1493 (52.57)		0.65 (0.59, 0.72)
Oral contraceptive usage ^b				
No	2659 (86.50)	2193 (82.48)	< 0.001	1.00
Yes	415 (13.50)	466 (17.52)		1.36 (1.18, 1.57)
History of benign breast disease ^b				
No	2720 (86.57)	2049 (73.07)	< 0.001	1.00
Yes	422 (13.43)	755 (26.93)		2.38 (2.08, 2.71)
Family history of cancer ^{b, c}				
No	2548 (80.61)	1966 (69.10)	< 0.001	1.00
Yes	613 (19.39)	879 (30.90)		1.86 (1.65, 2.09)

^aTwo-side χ^2 test, $P < 0.05$ was considered statistically significant

^bDue to the missing values, the number of cases and controls was less than 2881 and 3220, respectively

^cFirst- and second-degree relatives with history of cancer

Table 2 Logistic regression analysis of associations between lincRNAs' SNPs and the risk of breast cancer (Stage I)

SNP/gene	Genotype	No. (%)		P value ^a	OR (95% CI)	Adjusted OR (95% CI) ^b
		Controls (n = 1677)	Cases (n = 1492)			
rs217727 <i>H19</i>	GG	685 (40.90)	611 (41.06)	0.907	1.00	1.00
	GA	773 (46.15)	692 (46.51)		1.01 (0.87, 1.17)	0.98 (0.83, 1.15)
	AA	217 (12.96)	185 (12.43)		0.96 (0.77, 1.21)	0.93 (0.73, 1.19)
rs545226 <i>ANRIL</i>	GG	549 (32.85)	503 (33.74)	0.214	1.00	1.00
	GA	814 (48.71)	684 (45.88)		0.92 (0.78, 1.08)	0.95 (0.80, 1.13)
	AA	308 (18.43)	304 (20.39)		1.07 (0.88, 1.31)	1.08 (0.87, 1.34)
rs1328867 <i>HULC</i>	AA	476 (28.45)	437 (29.35)	0.318	1.00	1.00
	AC	823 (49.19)	752 (50.50)		1.00 (0.85, 1.18)	0.97 (0.81, 1.16)
	CC	374 (22.36)	300 (20.15)		0.89 (0.73, 1.08)	0.85 (0.69, 1.06)
rs1537375 <i>ANRIL</i>	CC	746 (44.51)	677 (45.47)	0.841	1.00	1.00
	CT	746 (44.51)	655 (43.99)		0.96 (0.83, 1.12)	0.97 (0.82, 1.15)
	TT	184 (10.54)	157 (10.57)		0.93 (0.74, 1.18)	0.85 (0.67, 1.07)
rs2023844 <i>HOTTIP</i>	AA	566 (33.79)	520 (34.88)	0.627	1.00	1.00
	AG	837 (49.97)	746 (50.03)		0.98 (0.84, 1.15)	0.97 (0.82, 1.15)
	GG	272 (16.24)	225 (15.09)		0.90 (0.73, 1.12)	0.86 (0.68, 1.09)
rs2038540 <i>HULC</i>	CC	984 (58.71)	871 (58.50)	0.977	1.00	1.00
	CG	608 (36.28)	541 (36.33)		0.99 (0.86, 1.15)	1.01 (0.86, 1.19)
	GG	84 (5.01)	77 (5.17)		1.04 (0.75, 1.43)	1.09 (0.76, 1.56)
rs2071095 <i>H19</i>	CC	786 (46.95)	792 (53.08)	0.000*	1.00	1.00
	CA	695 (41.52)	580 (38.87)		0.82 (0.71, 0.95)	0.84 (0.71, 0.99)
	AA	193 (11.53)	120 (8.04)		0.62 (0.48, 0.79)	0.64 (0.48, 0.83)
rs2106120 <i>ANRIL</i>	TT	682 (40.79)	579 (38.86)	0.223	1.00	1.00
	TG	691 (41.33)	610 (40.94)		1.04 (0.89, 1.22)	1.03 (0.87, 1.23)
	GG	299 (17.88)	301 (20.20)		1.18 (0.97, 1.44)	1.20 (0.97, 1.49)
rs2151280 <i>ANRIL</i>	AA	653 (38.96)	545 (36.58)	0.021*	1.00	1.00
	AG	801 (47.79)	696 (46.71)		1.04 (0.89, 1.21)	1.04 (0.88, 1.24)
	GG	222 (13.25)	249 (16.71)		1.33 (1.08, 1.65)	1.34 (1.06, 1.70)
rs2251375 <i>H19</i>	CC	569 (33.93)	504 (33.80)	0.854	1.00	1.00
	CA	778 (46.39)	704 (47.22)		1.02 (0.87, 1.20)	1.06 (0.89, 1.26)
	AA	330 (19.68)	283 (18.98)		0.97 (0.79, 1.18)	0.96 (0.77, 1.19)
rs2839698 <i>H19</i>	GG	875 (52.18)	801 (53.72)	0.489	1.00	1.00
	GA	673 (40.13)	568 (38.10)		0.91 (0.79, 1.06)	0.96 (0.81, 1.13)
	AA	129 (7.69)	122 (8.18)		1.03 (0.79, 1.35)	1.08 (0.81, 1.44)
rs2839701 <i>H19</i>	CC	801 (47.76)	762 (51.14)	0.138	1.00	1.00
	CG	732 (43.65)	600 (40.27)		0.91 (0.70, 1.18)	0.86 (0.73, 1.02)
	GG	144 (8.59)	128 (8.59)		1.07 (0.83, 1.38)	0.97 (0.73, 1.28)
rs3200401 <i>MALATI</i>	CC	1167 (69.67)	1052 (70.75)	0.223	1.00	1.00
	CT	446 (26.63)	396 (26.63)		0.99 (0.84, 1.16)	1.00 (0.84, 1.20)
	TT	62 (3.70)	39 (2.62)		0.70 (0.47, 1.06)	0.70 (0.44, 1.10)
rs3217992 <i>ANRIL</i>	CC	444 (26.48)	429 (28.77)	0.120	1.00	1.00
	CT	850 (50.69)	762 (51.11)		0.93 (0.79, 1.10)	0.90 (0.75, 1.08)
	TT	383 (22.84)	300 (20.12)		0.82 (0.67, 1.00)	0.79 (0.64, 0.99)
rs3741219 <i>H19</i>	AA	832 (49.61)	782 (52.45)	0.210	1.00	1.00
	AG	706 (42.10)	582 (39.03)		0.86 (0.75, 1.00)	0.88 (0.75, 1.04)
	GG	139 (8.29)	127 (8.52)		0.97 (0.75, 1.26)	1.03 (0.78, 1.37)
rs4930098 <i>H19</i>	GG	872 (52.12)	801 (53.72)	0.535	1.00	1.00
	GC	672 (40.17)	570 (38.23)		0.91 (0.79, 1.06)	0.97 (0.83, 1.14)
	CC	129 (7.71)	120 (8.05)		1.01 (0.78, 1.32)	1.04 (0.78, 1.39)

Table 2 (continued)

SNP/gene	Genotype	No. (%)		<i>P</i> value ^a	OR (95% CI)	Adjusted OR (95% CI) ^b
		Controls (<i>n</i> = 1677)	Cases (<i>n</i> = 1492)			
rs7958904	GG	986 (58.87)	881 (59.17)	0.738	1.00	1.00
<i>HOTAIR</i>	GC	583 (34.81)	505 (33.92)		0.97 (0.83, 1.12)	0.93 (0.79, 1.10)
	CC	106 (6.33)	103 (6.92)		1.07 (0.80, 1.43)	1.13 (0.83, 1.54)

The definition for the significance of bold is $P < 0.05$

CI confidence interval, OR odds ratio

^a $P < 0.05$

^aTwo-side χ^2 test for the difference in frequency distribution of the three genotypes between cancers and controls, $P < 0.05$ was considered statistically significant

^badjusted by age, menopausal status, oral contraceptive usage, smoking status, history of benign breast disease, and family history of cancer

Table 3 Logistic regression analysis of associations between rs2071095 and the risk of breast cancer (stage I and stage II)

Polymorphisms	No. (%)		<i>P</i> value ^a	OR (95%CI)	Adjusted OR (95%CI) ^b
	Cases (<i>n</i> = 2876)	Controls (<i>n</i> = 3190)			
Rs2071095 (C/A) (stage I)					
CC	792 (53.08)	786 (46.95)		1.00	1.00
CA	580 (38.87)	695 (41.52)	0.010	0.82 (0.71, 0.96)	0.84 (0.71, 0.99)
AA	120 (8.04)	193 (11.53)	0.000	0.62 (0.48, 0.79)	0.64 (0.48, 0.83)
Rs2071095 (C/A) (stage II)					
CC	727 (52.53)	792 (52.24)		1.00	1.00
CA	539 (38.95)	614 (40.50)	0.552	0.95 (0.82, 1.11)	0.92 (0.78, 1.09)
AA	118 (8.53)	110 (7.26)	0.273	1.17 (0.89, 1.54)	1.26 (0.94, 1.68)
Rs2071095 (C/A) (combined)					
CC	1519 (52.82)	1578 (49.47)		1.00	1.00
CA	1119 (38.91)	1309 (41.03)	0.024	0.89 (0.80, 0.98)	0.89 (0.79, 0.99)
AA	238 (8.28)	303 (9.50)	0.030	0.82 (0.68, 0.98)	0.83 (0.69, 0.99)
CA + AA	1358 (47.22)	1612 (50.53)	0.009	0.87 (0.79, 0.97)	0.88 (0.80, 0.98)
Rs2071095 (C/A) (ER-)					
CC	535 (51.20)	1577 (49.47)	0.332	1.00	1.00
CA + AA	510 (48.80)	1611 (50.53)		0.93 (0.81, 1.07)	0.94 (0.81, 1.09)
Rs2071095 (C/A) (ER+)					
CC	861 (54.12)	1577 (49.47)	0.002	1.00	1.00
CA + AA	730 (45.88)	1611 (50.53)		0.83 (0.74, 0.94)	0.86 (0.76, 0.98)
Rs2071095 (C/A) (PR-)					
CC	643 (52.92)	1577 (49.47)	0.040	1.00	1.00
CA + AA	572 (47.08)	1611 (50.53)		0.87 (0.76, 0.99)	0.88 (0.76, 1.01)
Rs2071095 (C/A) (PR+)					
CC	750 (52.85)	1577 (49.47)	0.034	1.00	1.00
CA + AA	669 (47.15)	1611 (50.53)		0.87 (0.77, 0.99)	0.90 (0.79, 1.03)

The definition for the significance of bold is $P < 0.05$

^aTwo-side χ^2 test for difference in frequency distribution of the genotypes between cancers and controls, $P < 0.05$ was considered statistically significant

^badjusted by age, menopausal status, smoking, family history of the benign breast disease and family history of cancer

Acknowledgements This study was supported by the National Natural Science Foundation of China (No.81473039), and the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (No.20140141).

Data availability statement The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval We declare that all experiments that we have performed comply with the current laws of our country. All procedures performed in our study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Chen W, Zheng R, Baade PD et al (2016) Cancer statistics in China, 2015. *CA Cancer J Clin* 66:115–132
- Pennisi E (2011) The Biology of Genomes. Disease risk links to gene regulation. *Science* 332:1031
- Zheng J, Huang X, Tan W et al (2016) Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat Genet* 48:747–757
- Kumar V, Westra HJ, Karjalainen J et al (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet* 9:e1003201
- Cabili MN, Trapnell C, Goff L et al (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 25:1915–1927
- Iyer MK, Niknafs YS, Malik R et al (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47:199–208
- Geisler S, Collier J (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 14:699–712
- Khalil AM, Guttman M, Huarte M et al (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 106:11667–11672
- Rinn JL, Kertesz M, Wang JK et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129:1311–1323
- Nagano T, Mitchell JA, Sanz LA et al (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322:1717–1720
- Guttman M, Donaghey J, Carey BW et al (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477:295–300
- Gupta RA, Shah N, Wang KC et al (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071–1076
- Bhan A, Soleimani M, Mandal SS (2017) Long noncoding RNA and cancer: a new paradigm. *Cancer Res* 77:3965–3981
- Bhan A, Hussain I, Ansari KI et al (2013) Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. *J Mol Biol* 425:3707–3722
- Huarte M, Guttman M, Feldser D et al (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142:409–419
- Bayram S, Sumbul AT, Batmaci CY et al (2015) Effect of HOTAIR rs920778 polymorphism on breast cancer susceptibility and clinicopathologic features in a Turkish population. *Tumour Biol* 36:3863–3870
- Peng R, Luo C, Guo Q et al (2018) Association analyses of genetic variants in long non-coding RNA MALAT1 with breast cancer susceptibility and mRNA expression of MALAT1 in Chinese Han population. *Gene* 642:241–248
- Huang NS, Chi YY, Xue JY et al (2016) Long non-coding RNA metastasis associated in lung adenocarcinoma transcript 1 (MALAT1) interacts with estrogen receptor and predicted poor survival in breast cancer. *Oncotarget* 7:37957–37965
- Timofeeva MN, Hung RJ, Rafnar T et al (2012) Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* 21:4980–4995
- Royds JA, Pilbrow AP, Ahn A et al (2015) The rs11515 polymorphism is more frequent and associated with aggressive breast tumors with increased ANRIL and decreased p16 (INK4a) expression. *Front Oncol* 5:306
- Riaz M, Berns EM, Sieuwerts AM et al (2012) Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes. *Breast Cancer Res Treat* 133:843–851
- Song F, Zheng H, Liu B et al (2009) An miR-502-binding site single-nucleotide polymorphism in the 3'-untranslated region of the SET8 gene is associated with early age of breast cancer onset. *Clin Cancer Res* 15:6292–6300
- Bartolomei MS, Zemel S, Tilghman SM (1991) Parental imprinting of the mouse H19 gene. *Nature* 351:153–155
- Seidl CI, Stricker SH, Barlow DP (2006) The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J* 25:3565–3575
- Brannan CI, Dees EC, Ingram RS et al (1990) The product of the H19 gene may function as an RNA. *Mol Cell Biol* 10:28–36
- Adriaenssens E, Dumont L, Lottin S et al (1998) H19 overexpression in breast adenocarcinoma stromal cells is associated with tumor values and steroid receptor status but independent of p53 and Ki-67 expression. *Am J Pathol* 153:1597–1607
- Gabory A, Jammes H, Dandolo L (2010) The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays* 32:473–480
- Cai X, Cullen BR (2007) The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* 13:313–316
- Kallen AN, Zhou XB, Xu J et al (2013) The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell* 52:101–112
- Barsyte-Lovejoy D, Lau SK, Boutros PC et al (2006) The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res* 66:5330–5337
- Ariel I, Sughayer M, Fellig Y et al (2000) The imprinted H19 gene is a marker of early recurrence in human bladder carcinoma. *Mol Pathol* 53:320–323
- Sorin V, Ohana P, Gallula J et al (2012) H19-promoter-targeted therapy combined with gemcitabine in the treatment of pancreatic cancer. *ISRN Oncol* 2012: 351750

33. Adriaenssens E, Lottin S, Dugimont T et al (1999) Steroid hormones modulate H19 gene expression in both mammary gland and uterus. *Oncogene* 18:4460–4473
34. Basak P, Chatterjee S, Weger S et al (2015) Estrogen regulates luminal progenitor cell differentiation through H19 gene expression. *Endocr Relat Cancer* 22:505–517
35. Guo H, Ahmed M, Zhang F et al (2016) Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nat Genet* 48:1142–1150
36. Maurano MT, Humbert R, Rynes E et al (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337:1190–1195
37. Huang Q, Whittington T, Gao P et al (2014) A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet* 46:126–135
38. Zhang X, Cowper-Sal LR, Bailey SD et al (2012) Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res* 22:1437–1446
39. Cowper-Sal LR, Zhang X, Wright JB et al (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* 44:1191–1198
40. Stacey SN, Sulem P, Masson G et al (2009) New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet* 41:909–914
41. Iacobucci I, Sazzini M, Garagnani P et al (2011) A polymorphism in the chromosome 9p21 ANRIL locus is associated to Philadelphia positive acute lymphoblastic leukemia. *Leuk Res* 35:1052–1059
42. Healy J, Belanger H, Beaulieu P et al (2007) Promoter SNPs in G1/S checkpoint regulators and their impact on the susceptibility to childhood leukemia. *Blood* 109:683–692
43. Cunnington MS, Santibanez KM, Mayosi BM et al (2010) Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet* 6:e1000899
44. Motterle A, Pu X, Wood H et al (2012) Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Hum Mol Genet* 21:4021–4029
45. Holdt LM, Beutner F, Scholz M et al (2010) ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler Thromb Vasc Biol* 30:620–627
46. Pasmant E, Sabbagh A, Vidaud M et al (2011) ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J* 25:444–448
47. Aguilo F, Zhou MM, Walsh MJ (2011) Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer Res* 71:5365–5369
48. Kotake Y, Nakagawa T, Kitagawa K et al (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30:1956–1962
49. Yap KL, Li S, Munoz-Cabello AM et al (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 38:662–674
50. Yu W, Gius D, Onyango P et al (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451:202–206
51. Lan WG, Xu DH, Xu C et al (2016) Silencing of long non-coding RNA ANRIL inhibits the development of multidrug resistance in gastric cancer cells. *Oncol Rep* 36:263–270
52. Zhang EB, Kong R, Yin DD et al (2014) Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a. *Oncotarget* 5:2276–2292
53. Sun Y, Zheng ZP, Li H et al (2016) ANRIL is associated with the survival rate of patients with colorectal cancer, and affects cell migration and invasion in vitro. *Mol Med Rep* 14:1714–1720