Carcinogenesis, 2016, Vol. 37, No. 5, 443–451

doi:10.1093/carcin/bgw024
Advance Access publication February 19, 2016
Original Manuscript

OXFORD

ORIGINAL MANUSCRIPT

A functional polymorphism in *lnc-LAMC2-1:1* confers risk of colorectal cancer by affecting miRNA binding

Jing Gong, Jianbo Tian, Jiao Lou, Juntao Ke, Lu Li[1], Jiaoyuan Li, Yang Yang, Yajie Gong, Ying Zhu, Yi Zhang, Rong Zhong, Jiang Chang[2] and Xiaoping Miao*

Department of Epidemiology and Biostatistics, State Key Laboratory of Environment Health (Incubation), MOE (Ministry of Education) Key Laboratory of Environment and Health, Ministry of Environmental Protection Key Laboratory of Environment and Health (Wuhan), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, PR China, [1]Research Center for Medicine and Biology, Zunyi Medical University, Zunyi, Guizhou, China and [2]Department of Nutrition and Food Hygiene, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, PR China

*To whom correspondence should be addressed. Tel: +86 27 8365 0744; Fax: +86 27 8369 3673; Email: miaoxp@mail.hust.edu.cn

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

Abstract

Genome-wide association studies (GWASs) have identified multiple susceptibility loci of colorectal cancer (CRC), however, causative polymorphisms have not been fully elucidated. Long non-coding RNAs (lncRNAs) are a recently discovered class of non-protein coding RNAs that involved in a wide variety of biological processes. We hypothesized that single nucleotide polymorphisms (SNPs) in lncRNA may associate with the CRC risk by influencing lncRNA functions. To evaluate the effects of SNPs on CRC susceptibility in Chinese populations, we first screened out all potentially functional SNPs in exons of lncRNAs located in CRC susceptibility loci identified by GWAS. Eight SNPs were selected and genotyped in 875 CRC cases and 855 controls and replicated in an independent case–control study consisting of 768 CRC cases and 768 controls. Analyses showed that CG and GG genotypes of the rs2147578 were significantly associated with increased risk for CRC occurrence in both case–control studies [combined analysis OR = 1.29; 95% confidence interval (CI) = 1.11–1.51, $P = 0.001$] compared to the rs2147578 CC genotype. Bioinformatics analyses showed that rs2147578 is located in the transcript of *lnc-LAMC2-1:1* and could influence the binding of *lnc-LAMC2-1:1*/miR-128-3p. Further luciferase reporter assays demonstrated that the construct with the risk rs2147578G allele had relatively high expression activity compared with that of the rs2147578C allele. Expression quantitative trait loci analyses also showed that rs2147578 is correlated with the expression of a well established oncogene *LAMC2* (laminin subunit gamma 2). These findings indicated that rs2147578 in *lnc-LAMC2-1:1* might be a genetic modifier for the development of CRC.

Introduction

Colorectal cancer (CRC) is a common malignant disease and the fourth leading cause of cancer-related death in the world. In China, CRC has been one of the cancers with the most rapidly increasing incidence and mortality during the past 10 years (1–3). Targeted prevention and early detection programs should be carried out to reverse this trend. It is well known that the occurrence of CRC is a complicated process and that multiple risk factors are involved in its initiation, promotion, and progression (4–6). Of these risk factors, genetics play an important role and may partially explain why only a fraction of individuals actually develop CRC when exposed to the same environment (7). Multiple lines of evidence have reported genetic alterations in patients with CRC. In particular, the association between single nucleotide polymorphisms (SNPs) and the risk of CRC has been widely reported (8,9).

Over the past decades, the development of high throughput technology has boosted the application of genome-wide association studies (GWAS). GWASs have identified several

Received: November 8, 2015; **Revised:** January 18, 2016; **Accepted:** February 12, 2016

© The Author 2016. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com.

Abbreviations

CRC	colorectal cancer
CI	confidence interval
eQTL	expression quantitative trait loci
GWAS	genome-wide association studies
LD	linkage disequilibrium
lncRNA	long non-coding RNA
OR	odds ratio
SNP	single nucleotide polymorphism

susceptibility loci related to CRC, including SNPs, insertions and deletions, and copy number variations (10–19). However, several of these identified risk SNPs are located in non-coding regions (20), and the functional or causative SNPs in these susceptibility loci remains to be studied.

Long non-coding RNA (lncRNA) refers to a diverse class of transcripts that are larger than 200 nucleotides and do not serve as templates for proteins (21). Advances in deep sequencing are giving rise to a fast accumulation of lncRNAs. Increasing evidence demonstrates that up to 90% of the non-coding transcripts in the human genome have important and diverse biological roles (22). Many of the lncRNAs show abnormal expression in cancers, with potentially carcinogenic or tumor suppressor roles (23,24). For example, next-generation sequencing has identified several of abnormal expressed lncRNAs associated with vincristine resistance in colon cancer cells (23). The presence and overexpression of lncRNA *HNF1A-AS1* has been detected in oesophageal adenocarcinoma patients (24). Among the lncRNAs, a subset of the lncRNAs can interact with miRNAs by complementary sequence to act as a miRNA decoy or sponge, which indirectly influence the miRNA regulation of other protein-coding genes. For example, lncRNA *PTENP1* function as a decoy for *PTEN*-targeting miRNAs in tumor suppression (25). Currently, starBase has collected more than 10 000 miRNA-lncRNA interaction pairs from high-throughput studies (26).

Since the lncRNAs are not 'junk' genes, the SNPs in lncRNAs may disturb the function of lncRNAs, therefore promoting or inhibiting the occurrence and development of diseases. And if SNPs locate in miRNA target sites of lncRNAs, they may influence (destroy or create) the miRNA-lncRNA interactions, thereby directly impact the expression of lncRNA and/or indirectly regulate other protein-coding genes. Recently, several SNPs in lncRNAs have been found to be associated with different diseases or traits. For example, SNP rs6983267 is a well validated risk locus for CRC (17,27). Lin *et al.* (28) have demonstrated that the function of the SNP rs6983267 might be mediated by inducing the differential lncRNA CCAT2 expression. Zhang *et al.* (29) have identifed an ESCC susceptibility SNP rs920778 which may exert its function by regulating the expression of lncRNA *HOTAIR* via a novel intronic enhancer. Tao *et al.* (30) have validated an association between indel polymorphism in the promoter region of lncRNA GAS5 and the risk of hepatocellular carcinoma. Wu*et al.* found a genetic polymorphism in lincRNA-uc003opf.1 is associated with susceptibility to esophageal squamous cell carcinoma in Chinese populations (31). In addition, several GWASs identified trait-associated SNPs are in or nearby lncRNAs rather than protein-coding genes, such as *FAL1* (32). These SNPs are more likely to exert their roles through lncRNAs.

Based on the above description, we speculate that there may exist functional SNPs in lncRNAs, especially in miRNA:lncRNA binding sites, which are associated with the occurrence of CRC. To test the above speculation, we first systematically screened functional SNPs in the CRC GWAS susceptibility regions, which may influence lncRNA:miRNA binding. Then two stage case-control studies in Chinese were conducted to explore the association between these candidate SNPs and the risk of CRC. Finally, the effect of SNPs on lncRNA:miRNA was tested by luciferase reporter assays.

Materials and methods

Selection of candidate SNPs

CRC risk tagSNPs identified by GWASs were downloaded from the NHGRI GWAS Catalog up to December 31, 2013 (20). We also searched the literature via PubMed using the search terms of 'GWAS', 'Genome-wide association studies' and 'colorectal cancer' to retrieve CRC risk tagSNPs. Then, Haploview software 4.2 (33) was used to calculate the linkage disequilibrium (LD) blocks of each tagSNP by analyzing the Chinese Han Beijing (CHB) genotype information of ±500 kb around the tagSNP (setting $R^2 \geq 0.8$), and these LD blocks were defined as CRC susceptibility loci. Next, SNPs in these CRC susceptibility loci and lncRNAs were screened out. SNP data were downloaded from dbSNP of NCBI (dbSNP v138), while lncRNA information was obtained from the LNCipedia database (34). Further to narrow down the potential functional SNPs, we only selected SNPs which may impact miRNA:lncRNA interaction according to the lncRNASNP database (35) and limited the minimum allele frequency of SNPs in the Chinese population (CHB) ≥ 0.05. Finally, eight SNPs were selected as candidate SNPs.

Study populations

A two-stage case–control study was conducted to comprehensively evaluate the association between the SNPs in lncRNAs and the risk of CRC. Stage Iconsisted of 875 CRC patients and 855 cancer-free controls. All subjects were unrelated ethnic Han Chinese. The CRC patients were consecutively recruited between 1 January 2010 and 31 November 2013 at the Tongji Hospital of Huazhong University of Science and Technology (HUST), Wuhan, Central China and were pathologically confirmed. Cancer-free subjects were recruited in the same hospitals from individuals receiving routine physical examinations or in the communities from individuals participating in screening of chronic diseases, part of which were also involved in our previous epidemiological studies (36,37). The stage II study consisted of 738 cases and 738 cancer-free controls. Patients were recruited between 1 January 2009 and 31 August 2012 at the Cancer Hospital, Chinese Academy of Medical Sciences (Beijing) in northern China. Controls came from a community cancer screening program for early detection of cancer conducted in the same regions. In each stage, controls were frequency matched to cases by gender and age (±5 years). At recruitment, a written informed consent was obtained from each participant; peripheral blood samples and demographic information including gender, age, smoking and drinking habits were also collected. Participants who had never smoked or who had smoked <1 cigarette/day and for <1 year before the date of cancer diagnosis for cases or before the date of the interview for controls, were defined as nonsmokers. Otherwise, they were considered as smokers (including current smokers and ever smokers). Individuals were classified as drinkers if they drank at least twice a week and continuously for at least 1 year during their lifetimes; otherwise, they were defined as non-drinkers. This study was performed under the approval of the institutional review boards of Tongji Medical College of HUST and the Chinese Academy of Medical Sciences Cancer Institute.

Genotype

Genomic DNA was extracted from a 5ml sample of peripheral blood by using the RelaxGene Blood DNA System DP319-02 (TIANGEN, Beijing, China). Candidate SNPs were genotyped by the TaqMan Openarray assay in the first stage and TaqMan real-time polymerase chain reaction (PCR) assay (Applied Biosystems, Foster city, CA) in the second stage. Quality control was monitored as follows: (i) case and control samples were mixed and genotyped without knowing the case or control status; (ii) including 5% duplicate and negative control. The concurrence rate of the duplicate sets in this study was 100%. The average call rate for the candidate SNPs genotyped was >90%.

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

Copy number of the locus *lnc-LAMC2-1:1*

The copy numbers of the locus *lnc-LAMC2-1:1* in the LoVo and SW480 cells were tested by real-time quantitative PCR, respectively. Total genomic DNA was extracted from SW480 and LoVo cell lines by TIANamp Genomic DNA Kit (TIANGEN, Beijing, China). Real-time Quantitative PCR was performed by Power SYBR Green PCR mix (Applied Biosystems, Foster city, CA). All primers used in quantitative PCR are shown as follows, *lnc-LAMC2-1:1* primer 5′- GGCCCAAGGAAGAACTAAGG -3′ and 5′- ATCCAAACCAACATCCACCC-3′; β-globin primer: 5′-GGTGAGCCAGGCCATCACTA-3′ and 5′-GGCAACCC TAAGGTGAAGGC-3′. The copy number of *lnc-LAMC2-1:1* gene was normalized to that of β-globin (known as two).

Plasmid construction

The wild-type transcript sequence of *lnc-LAMC2-1:1* (524bp) was downloaded from the LNCipedia database and synthesized by Genewiz Company (Suzhou, China). The sequence was cloned into the psiCHECK™ -2 vector (Promega, Madison, WI). The mutation type of *lnc-LAMC2-1:1* was generated by site-specific mutagenesis at the rs2147578 site (G>C). The primer sequences for site-specific mutagenesis were 5′-CATAGTCCCTCAGTGTGGGTCATTTTCATTAG-3′ and 5′-CTAATGAAAAT GACCCACACTGAGGGACTATG-3′. The mimics of miR-128-3p were chemically synthesized by Shanghai GenePharma Company (Shanghai, China).

Cell cultures, transient transfections and luciferase reporter assays

SW480 and LoVo CRC cell lines were obtained from the China Center for Type Culture Collection (Wuhan, PR China) and were grown in DMEM supplemented with 10% fetal bovine serum (GIBCO) and 1% antibiotics (50 U/ml penicillin and 50 μg/ml streptomycin) in a humidified atmosphere of 5% CO_2 at 37°C. All cell lines have never been passaged longer than 3 months and tested routinely by DNA sequencing using the Applied Biosystems AmpF/STR Identifier kit and last checked in September 2015. For transfection assays, cells were seeded in 96 well plates and simultaneously transfected with psiCHECK-2 vector and miRNA mimics by using Lipofectamine 3000 (Invitrogen). After 48 h, luciferase activity was assessed by following manufacturer's instruction (dual luciferase assay system, Promega). Renilla luciferase and Firefly luciferase activities were detected. Renilla luciferase/Firefly luciferase was calculated to compare the differences between different alleles. All experiments were independently performed in triplicate.

Cis-eQTL analysis

For the eQTL analysis, we downloaded the germline genotypes, the somatic copy number, methylation and expression profiles of colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) from 'The Cancer Genome Atlas' (TCGA) data portal. Of them, we selected 254 individuals with genotypes that clustered with HapMap CEU controls by the EIGENSTRAT program (38) and both germline genotypes and expression data were available. The genotypes of TCGA were obtained by Affymetrix 6.0 array. For each interesting SNP, if the SNP was not designed in the Affymetrix 6.0 array, we first imputed its genotypes by impute2 (39) software using the genotypes of all SNPs on the Affymetrix 6.0 array within 2 Mb of either side of the SNP. Then, for each SNP, the association between the SNP and nearby genes was evaluated using a linear regression model with the effects of somatic copy number and CpG methylation being deducted. The detail of the algorithms is as previously reported in Li *et al.* (40).

Statistical analysis

In both case–control stages, Hardy–Weinberg equilibrium for the genotype distribution of each SNP was evaluated using the goodness-of-fit χ^2 test by comparing the observed genotype frequencies with those expected among the controls. The distribution of age, gender and smoking between the CRC patients and healthy controls, as well as alleles and genotypes, was appraised using the two-sided χ^2 test or the Student's t test. The associations between SNPs and susceptibility of CRC were demonstrated by calculating the odds ratios (ORs) and their 95% confidence intervals (CIs). Crude ORs and 95% CIs were calculated by the univariate logistic regression model; adjusted ORs and 95% CIs were calculated by the multivariate logistic regression model after adjusting for gender, age, smoking and drinking status. Statistical power was computed by applying the Power V3.0 software. All experiments data are presented as the mean ± SD and analyzed using a Student's t test. All the statistical analyses were conducted by SPSS Statistics 18.0 and PLINK software (41), and $P < 0.05$ was defined as statistically significant.

Results

Selection of candidate SNPs

By retrieving PubMed and NHGRI GWAS Catalog databases, a total of 47 CRC risk tagSNPs were collected by December of 2013, and their LD blocks were calculated by Haploview software. Through comparing the genomic coordinates of SNPs with CRC susceptibility loci, we identified 41 650 SNPs in these CRC susceptibility loci, and 5133 of these SNPs of them had a minimum allele frequency > 0.05. After intersecting these SNPs with SNPs in the lncRNASNP database, eight SNPs were identified with the potential to impact miRNA:lncRNA binding. These SNPs and corresponding lncRNA and miRNA information are shown in Table 1.

Subjects characteristics

The demographic characteristics of the two stage populations are summarized in Table 2. Of the 1643 cases, 678 cases are the patients with colon cancer and 885 cases are the patients with rectal cancer, while 80 patients are difficult to class because the mass located in the junction of the colon and rectum, or existed in both colon and rectum. The distributions of SNP alleles in combined population did not show any significantly heterogeneity between the colon and rectum in our study (Supplementary Table 1, available at *Carcinogenesis* Online). In both stages and combined population, no significant differences were observed between CRC cases and healthy controls in distribution of sex, age and drinking habit. CRC cases had significantly higher levels of smoking compared with controls in both the two stages and combined study populations. The ORs for smokers in stages 1, 2 and the combined population were 1.62 [95% confidence

Table 1. Candidate SNPs that can affect lncRNA:miRNA binding and are located in CRC risk loci

Chr	SNP position	SNP ID	Ref/alt	lncRNA	miRNA	Effect
chr1	183107699	rs2147578	G/C	lnc-LAMC2-1:1	miR-128-3p,216a-3p,3681-3p	gain
chr1	222139706	rs112462125	T/C	lnc-MIA3-3:1	miR-197-3p	loss
chr3	47421299	rs75511849	T/C	lnc-SCAP-1:1	miR-100-3p	loss
chr5	108833582	rs60719452	T/C	lnc-MAN2A1-1:1	miR-548-5p,548ab,548ak,548au-5p,548ay-5p,548b-5p,548d-5p,548i,548y	gain
chr8	117616931	rs7814028	G/A	lnc-UTP23-1:2	miR-5001-3p,miR-6819-3p	gain
chr8	117619136	rs12677572	G/A	lnc-UTP23-1	miR-891a-5p	Loss
chr8	117616805	rs7844527	C/A	lnc-UTP23-1:1	miR-146a-5p,146b-5p	Loss
chr12	4130184	rs61095617	A/G	lnc-CCND-2:1	miR-1307-5p	Gain

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

Table 2. Characteristics of participants in the two-stage case–control study

	Stage 1			Stage 2			Combined study		
	Cases	Controls		Cases	Controls		Cases	Controls	
	N (%)	*N* (%)	*P*	*N* (%)	*N* (%)	*P*	*N* (%)	*N* (%)	*P*
Total	875	855		768	768		1643	1622	
Gender			0.083			0.601			0.361
Male	512 (58.51)	465 (54.39)		465 (60.55)	475 (61.85)		977 (59.46)	939 (57.89)	
Female	363 (41.49)	390 (45.61)		303 (39.45)	293 (38.15)		666 (40.54)	683 (42.11)	
Age (mean ± SD)	59.81 ± 12.45	59.07 ± 12.59	0.219[a]	58.94 ± 11.73	59.43 ± 10.51	0.385[a]	59.40 ± 12.12	59.24 ± 11.65	0.703[a]
Age group			0.173			0.402			0.177
<50	179 (20.46)	169 (19.77)		158 (20.57)	141 (18.36)		337 (20.51)	310 (19.11)	
50–59	247 (28.23)	229 (26.78)		218 (28.39)	243 (31.64)		465 (28.30)	471 (29.04)	
60–69	232 (26.51)	267 (31.23)		234 (30.47)	239 (31.12)		466 (28.36)	506 (31.20)	
≥70	217 (24.80)	190 (22.22)		158 (20.57)	145 (18.88)		375 (22.82)	335 (20.65)	
Smoking			<0.0001			0.021			<0.0001
No	554 (64.34)	630 (73.68)		524 (68.23)	565 (73.57)		1078 (66.18)	1194 (73.61)	
Yes	307 (35.66)	225 (26.32)		244 (31.77)	203 (26.43)		551 (33.82)	428 (26.39)	
Drinking			0.284			0.862			0.53
No	614 (71.31)	629 (73.74)		570 (74.22)	566 (73.70)		1184 (72.68)	1195 (73.72)	
Yes	247 (28.69)	226 (26.26)		198 (25.78)	202 (26.30)		445 (27.32)	426 (26.28)	

SD, standard deviation.
[a]*P* value was calculated by the *t* test. Other *P* values were calculated by the χ^2 test.

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

Table 3. Association analyses between individual SNPs and CRC risk in the combined population

SNP	Genotypes	Stage 1			Stage 2			Combined study		
		Cases/controls	OR (95%CI)[a]	P[b]	Cases/Controls	OR (95%CI)[a]	P[b]	Cases/controls	OR (95%CI)[a]	P[b]
rs2147578	CC	225/259	1		206/256	1		431/515	1	
	CG	453/420	1.25 (1.00–1.56)	0.053	409/367	1.35 (1.07–1.71)	0.011	862/787	1.30 (1.10–1.52)	0.002
	GG	179/166	1.25 (0.95–1.66)	0.113	152/143	1.30 (0.97–1.76)	0.078	331/309	1.28 (1.05–1.56)	0.018
	Dominant		1.25 (1.01–1.55)	0.040		1.34 (1.07–1.67)	0.009		1.29 (1.11–1.51)	**0.001**
	Recessive		1.09 (0.86–1.38)	0.495		1.08 (0.83–1.39)	0.568		1.08 (0.91–1.29)	0.372
	Additive		1.13 (0.98–1.30)	0.083		1.16 (1.01–1.35)	0.042		1.15 (1.04–1.27)	0.008
rs112462125	TT	496/442	1		409/395	1		905/837	1	
	CT	277/318	0.77 (0.62–0.95)	0.014	299/305	0.93 (0.76–1.15)	0.440	576/623	0.85(0.73–0.98)	0.028
	CC	62/75	0.75 (0.52–1.07)	0.115	53/63	0.80 (0.54–1.19)	0.267	115/138	0.77(0.59–1.01)	0.058
	Dominant		0.77 (0.63–0.93)	0.008		0.91 (0.75–1.12)	0.370		0.83(0.72–0.96)	0.011
	Recessive		0.83 (0.58–1.18)	0.291		0.83 (0.56–1.21)	0.325		0.83(0.64–1.07)	0.153
	Additive		0.82 (0.71–0.96)	0.012		0.92 (0.78–1.07)	0.263		0.87(0.78–0.97)	0.010
rs7844527	CC	549/553	1		474/471	1		1023/1024	1	
	CA	273/268	1.00 (0.82–1.24)	0.974	253/259	0.88 (0.52–1.47)	0.617	526/527	0.99 (0.85–1.15)	0.853
	AA	42/26	1.60 (0.96–2.65)	0.074	34/30	1.02 (0.54–1.50)	0.695	76/56	1.33 (0.93–1.90)	0.122
	Dominant		1.06 (0.86–1.30)	0.598		0.89 (0.54–1.48)	0.661		1.02 (0.88–1.18)	0.800
	Recessive		1.59 (0.96–2.63)	0.072		1.02 (0.83–1.25)	0.875		1.33 (0.94–1.90)	0.112
	Additive		1.10 (0.93–1.30)	0.283		1.00 (0.84–1.20)	0.984		1.05 (0.93–1.19)	0.442
rs7814028	GG	539/550	1		476/474	1		1015/1024	1	
	GA	269/266	1.02 (0.83–1.26)	0.858	252/260	0.96 (0.78–1.20)	0.729	521/526	0.99 (0.85–1.15)	0.893
	AA	44/26	1.68 (1.02–2.81)	0.036	32/30	1.05 (0.62–1.75)	0.866	76/56	1.34 (0.94–1.92)	0.11
	Dominant		1.08 (0.88–1.32)	0.456		0.97 (0.79–1.20)	0.783		1.02 (0.89–1.18)	0.751
	Recessive		1.68 (1.02–2.77)	0.037		1.06 (0.64–1.77)	0.824		1.35 (0.94–1.92)	0.100
	Additive		1.12 (0.95–1.33)	0.177		0.99 (0.83–1.18)	0.875		1.05 (0.93-1.20)	0.399

OR, odds ratio; 95% CI, 95% confidence interval.

[a]ORs and 95%CIs calculation were conducted under assumption that variant alleles were risk alleles.

[b]All the *P* values were adjusted for gender, age group smoking and drinking status.

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

interval (CI) = 1.26–2.07], 1.41 (95% CI = 1.10–1.79) and 1.52 (95% CI = 1.28–1.80), respectively.

Association between candidate SNPs and CRC risk

In the first stage (except for rs61095617), the genotype distribution in controls of the other seven SNPs were all in Hardy–Weinberg equilibrium ($P>0.05$). The results of association analyses were presented in Table 3 and Supplementary Table 2, available at *Carcinogenesis* Online. As shown in Table 3, the CG and GG genotypes of rs2147578 were significantly associated with CRC (OR = 1.25, 95%CI = 1.01–1.55, $P = 0.040$), compared with the CC counterpart. The adjusted OR of carrying the rs112462125 CT and CC genotypes was 0.77 (95%CI = 0.63–0.93, $P = 0.008$), compared with the rs112462125 TT genotype. The P values of another three SNPs (rs7844527, rs7814028 and rs12677572) were around 0.05 when comparing the variant homozygote to the wild-type homozygote. Among them, the rs12677572 and rs7814028 located very close and showed high LD ($R^2 = 0.98$) in our results, so we chose rs7814028 which with a lower P value for the replication. Therefore, four SNPs were finally chosen as candidate SNPs for the replication study.

In the second stage, the average genotyping call rate of the four SNPs was 99.2%, and all SNPs in the controls were in the Hardy–Weinberg equilibrium ($P>0.05$). Consistent with the results in the first stage, rs2147578 was still significantly associated with CRC risk and had an OR of 1.34 (95%CI = 1.07–1.67, $P = 0.009$) after adjusting for age, sex, smoking and drinking status in the dominant model. The other SNPs did not show any association with CRC risk.

The combined analysis also confirmed a significant association between rs2147578 and CRC risk (OR = 1.15, 95% CI = 1.04–1.27, $P = 0.008$, Table 3), under the additive model with the adjustment for gender, age, smoking and drinking status, which suggested that each additional copy of the effect allele (G) of rs2147578 conferred a 15% increase in the susceptibility of CRC compared to those carrying the C allele. Further logistic regression analyses showed no interactions of rs2147578 with smoking status or other SNPs. Although there were no associations between smoking and rs2147578, we found that smokers with rs2147578 CG and GG genotypes were significantly related to the susceptibility of CRC with a higher OR of 1.94 (95% CI = 1.54–2.44, $P < 0.0001$) compared with non-smokers with CC genotypes (Table 4), which implied that this SNP might play an important role in the development of CRC for the smoking population.

We also conducted a meta-analysis to assess the effects of the four SNPs on the CRC risk by combining the results of stages I and II. The fixed-effects model was applied due to no significant heterogeneity ($P>0.1$) was observed between studies. The results also showed that the rs2147578 was significant associated with CRC risk ($P=0.007$) (Supplementary Table 3, available at *Carcinogenesis* Online).

The effect of rs2147578 on miR-128-3p: *lnc-LAMC2-1:1* binding

First, the quantitative PCR and Taqman genotyping technology were used to obtain the copy number and the genotype of the locus *lnc-LAMC2-1:1* in the LoVo and SW480 cells, respectively. The copy numbers of the locus *lnc-LAMC2-1:1* in the LoVo and SW480 cells are 1.779 ± 0.010 and 1.625 ± 0.002, respectively. The *lnc-LAMC2-1:1* located in the *LAMC1* gene. The genotypes of the locus *lnc-LAMC2-1:1* in the LoVo and SW480 cells are both CC.

Then, we have constructed two luciferase reporter containing 524bp of the *lnc-LAMC2-1:1* exon regions, with a G or C allele at rs2147578 polymorphic sites in psiCHECK-2. Luciferase reporter assays were conducted in SW480 and LoVo cell lines, respectively. As revealed in Figure 1, compared with the rs2147578G allele, the construct containing rs2147578C allele demonstrated ~12.6 and 12.3% decrease in luciferase activity in the absence of miR-128-3p mimics in SW480 and LoVo cells, respectively ($P < 0.001$). When cotransfected with miR-128-3p mimics, there was a more significant decrease in luciferase activity, approximately 31.5% in SW480 cells and 24.9% in LoVo cells ($P < 0.001$), respectively. When additional inhibitor was co-transfected into cells, the differences in luciferase activity between the G and C alleles disappeared. These results suggested that miR-128-3p might bind tightly to *lnc-LAMC2-1:1* containing the rs2147578C allele, thereby negatively regulating the level of *lnc-LAMC2-1:1*.

Table 4. Interaction analysis between smoking and the rs2147578 in the combined study

		Cases/controls	OR (95% CI)	P^a
rs2147578				
Non-smokers	CC	283/381	1	
	CG+GG	783/806	1.30 (1.08-1.56)	0.005
Smokers	CC	145/134	1.52 (1.14-2.04)	0.005
	CG+GG	399/290	1.94 (1.54-2.44)	<0.0001
P_{mult}, FDR-P_{mult} [b]				0.895, 0.895
P_{add}, FDR-P_{add} [b]				0.369, 1

[a]The P values were adjusted by gender and age group except the additive model.
[b]The P values were adjusted by Benjamini–Hochberg correction.

eQTL analysis

There are 17 genes within rs2147578 ± 1 Mb of flanking regions. The expressions of 14 of these genes can be obtained from TCGA RNAseq datasets. The genotypes of rs2147578 were imputed by impute2 software with an imputation quality of 0.997, which represents a good certainty of the guess genotypes. Cis-eQTL analyses between rs2147578 and nearby genes are performed by linear regression. The results showed that there were no significant associations between rs2147578 and nearby genes (Table 5). However, the P value of the correlation between rs2147578 and *LAMC2* is 0.09. *LAMC2* has 23 exons and the expression profiles of these exons can also be obtained from TCGA RNAseq datasets. Further Cis-eQTL analysis between rs2147578 and these exons showed that some P values of the correlations are under cutoff 0.05 (Table 5). As shown in Figure 2, with the increase of G genotype, the average expression of *LAMC2* is gradually increased. This suggested that the expression of *LAMC2* may be slightly influenced by rs2147578, however, this conclusion requires validation by enlarging the sample size. In addition, TCGA RNAseq data also show that *LAMC2* is highly expressed in tumors compared with adjacent normal tissues (4593 ± 3762 RPKM versus 2246 ± 1578 RPKM), while no expression difference of *LAMC1* was observed (3587 ± 1800 RPKM versus 3480 ± 2549 RPKM).

Discussion

In this study, we screened out eight lncRNA-related SNPs in the CRC GWAS susceptibility regions and designed two stage case–control studies in the Chinese population to explore the associations between these candidate SNPs and the risk of CRC. We confirmed that smoking is a risk factor for CRC and found that the CG and GG genotype of the rs2147578 in *lnc-LAMC2-1:1* exon had a significantly higher risk compared with the CC genotype. In the additive model, the rs2147578 G allele significantly

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

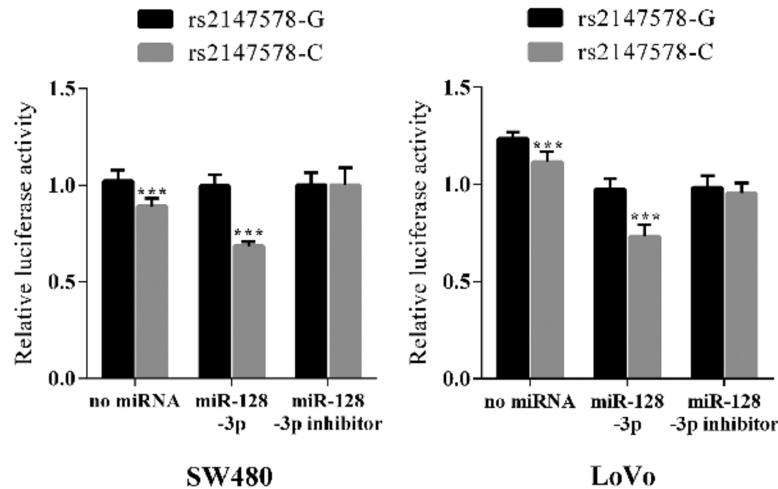

Figure 1. The effect of rs2147578 on miR-128-3p: lnc-LAMC2-1:1 binding. Bar graphs of (**A**) and (**B**) show the relatively luciferase activity of vectors containing the rs2147578G or C allele in SW480 and LoVo cells, respectively. Luciferase expression vectors containing the *lnc-LAMC2-1:1* with the rs2147578G or C allele were constructed and co-transfected with miRNA mimics (miR-128-3p) in SW480 and LoVo cells, respectively. Renilla luciferase/Firefly luciferase were calculated and normalized to blank or NC controls as relatively luciferase activity. Six replicates for each group were conducted and all transfection experiments were repeated at least three times. Data are presented as the mean ± SD and asterisk indicates a significant change ($P < 0.001$).

Table 5. Expression correlation between rs2147578 and flanking 1 Mb genes

Gene	Correlation P	Correlation R^2
LAMC2:chr1:183194743:183194855:+	0.043	0.016
LAMC2:chr1:183189960:183190096:+	0.048	0.015
LAMC2:chr1:183212282:183214262:+	0.048	0.015
LAMC2:chr1:183195833:183196051:+	0.046	0.016
APOBEC4	0.966	0.000
ARPC5	0.670	0.001
DHX9	0.896	0.000
GLUL	0.475	0.002
LAMC1	0.813	0.000
LAMC2	0.091	0.011
NCF2	0.332	0.004
NMNAT2	0.855	0.000
NPL	0.235	0.006
RGL1	0.554	0.001
RGS16	0.422	0.003
RNASEL	0.846	0.000
SMG7	0.431	0.002
TSEN15	0.332	0.004

increased the risk of CRC incidence, compared with the C allele. Luciferase reporter assays showed that rs2147578C can mildly reduce the expression of *lnc-LAMC2-1:1* by binding miR-128-3p. Moreover, eQTL analysis showed that rs2147578 can slightly impact the expression of the protein coding gene *LAMC2*, which is a well-established oncogene (42). These results suggest that rs2147578 may be a functional SNP in colorectal carcinogenesis.

Rs2147578 locates in the first exon of *lnc-LAMC2-1:1* on chromosome 1q25.3, and also in the 26th intron of the *LAMC1* gene and close to the *LAMC2* gene. This CRC risk locus was first identified by Peters *et al.* (12) by a meta-analysis including 12 696 cases of colorectal tumors (11 870 cancer, 826 adenoma) and 15 113 controls of European descent. The most significant SNP was rs10911251 with an OR of 1.10 per risk allele and $P = 9.5 \times 10^{-8}$. As shown in the Supplementary Figure 3, available at *Carcinogenesis* Online of Peters *et al.* study, there were dozens of SNPs associated with CRC ($P < 0.05$), even with a low LD ($R^2 < 0.2$) versus rs10911251. By using Haploview software, we have defined the LD block ($R^2 \geq 0.8$) of this locus as chromosome 1: 182964441-183117430. The LD between rs2147578 and rs10911251 is relatively high ($R^2 = 0.7$). Although they predicted that rs10911205 would be a potential functional candidate in this locus, the causal variants are still remained to be studied. The newly identified rs2147578 may be an individual or synergetic functional SNP in this region.

The biological role of *lnc-LAMC2-1:1* has not yet been studied, however, it overlap with the *LAMC1* gene and is close to the *LAMC2* gene. According to previous studies, a subset of lncRNA genes are involved in the cis-regulation of target genes located at or near the same genomic locus (43–45), which suggests that *lnc-LAMC2-1:1* may exert its function by interacting with nearby genes. Our cis-eQTL analyses did not reveal any correlation between rs2147578 and *LAMC1*, but did show that the expression of *LAMC2* was slightly influenced by rs2147578, with the risk allele G slightly increasing the expression of *LAMC2*. Furthermore, elevated expression of *LAMC2* was observed in TCGA CRC tissues. *LAMC2*, known as laminin gamma 2, is one member of the Laminin family and shows elevated expression in multiple cancers (46–48). The elevated expression of *LAMC2* has been demonstrated to drive tumorigenesis through its interactions with several cell-surface receptors (46,49). Accumulating evidence indicates that *LAMC2* plays an important role in the progression, migration and invasion of multiple types of cancer (50–52), suggesting that it might be a potential therapeutic anticancer target for inhibiting tumorigenesis (42). Therefore, we propose the existence of some interaction between *lnc-LAMC2-1:1* and *LAMC2*. The elevated expression of *lnc-LAMC2-1:1* may increase the expression of *LAMC2*, consequently enhancing cancer progression. And according to our luciferase assays, the function of rs2147578 may be mediated by miR-128-3p. The binding between the rs2147578G construct and miR-128-3p was weaker than that of the rs2147578C construct, thereby increasing the expression of *lnc-LAMC2-1:1*.

In our bioinformatics analyses, rs2147578C not only bind to miR-128-3p, but can also bind to miR-216a-3p and miR-3681-3p. However, the expression of miR-216a-3p and miR-3681-3p are

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

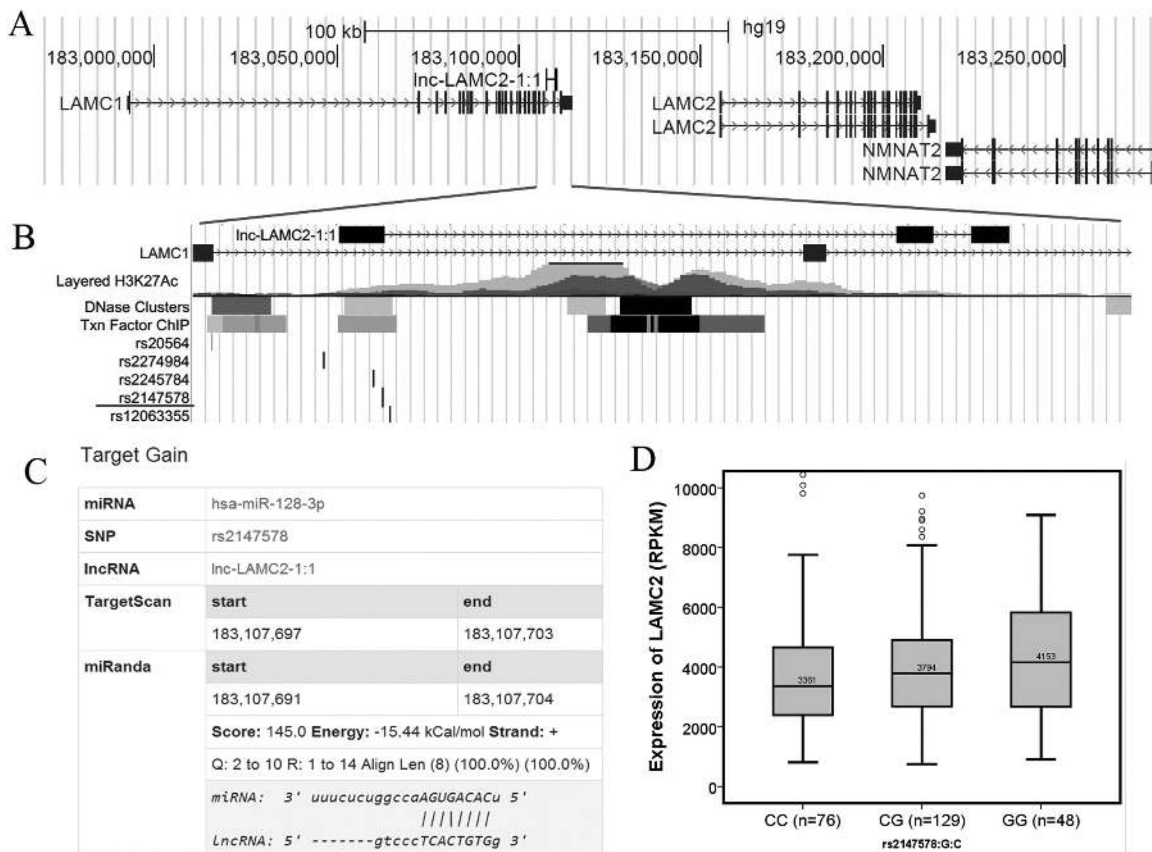

Figure 2. The functional prediction of rs2147578 at the 1q25.3 locus. (**A**) The *lnc-LAMC2-1:1* locates on chromosome 1, overlaps with the protein-coding gene *LAMC1* and is close to *LAMC2*. (**B**) The rs2147578 is in the first exon of *lnc-LAMC2-1:1* and 26th intron of *LAMC1*. (**C**) The rs2147578 C allele gain the binding of miR-128-3p as predicted by lncRNASNP database. (**D**) Expression correlation between rs2147578 and *LAMC2*.

at a very low level (< 1 RPKM), while the average expression of miR-128-3p is 138 RPKM in CRC tissues by miRNASNP2 annotation (53). The function of miR-216a-3p and miR-3681-3p in CRC may be very limited, therefore we only validated the effect of rs2147578 on the binding of *lnc-LAMC2-1:1*:miR-128-3p. In addition, several studies have reported that miR-128 is involved in CRC through different targets or networks (54,55). lncRNA can function as an miRNA decoy or sponge, thereby affecting miRNA regulation of other protein-coding genes (56). Therefore, the function of rs2147578 may also indirectly affect miR-128-3p binding with other genes, consequently affecting tumorgenesis.

In conclusion, our study highlighted that *lnc-LAMC2-1:1* rs2147578G allele may increase the risk of CRC in Chinese populations by losing the binding of miR-128-3p. The results of our study further support assumptions that miRNAs may regulate lncRNA expression by the combining with lncRNA exons, thereby affecting tumor susceptibility. However, several limitations should be acknowledged here. First, in the stages I and II, no SNP can survive after the multiple comparisons, suggesting that the sample sizes of our case–control studies were relatively small. Independent replication studies with large sample sizes are warranted to verify our results. Second, the function of *lnc-LAMC2-1:1* is not fully elucidated. Additional experiments *in vivo* and *in vitro* of *lnc-LAMC2-1:1* will be the next step towards determining the functionality of this lncRNA and its associated miRNAs in CRC.

Supplementary material

Supplementary Tables 1–3 can be found at http://carcin.oxfordjournals.org/

Funding

This work was supported by National Program for Support of Top-notch Young Professionals, National Natural Science Foundation of China (NSFC-81001275, 81171878, 81222038, 81402744 and 81502875); the Fok Ying Tung Foundation for Young Teachers in the Higher Education Institutions of China (131038); Natural Science Foundation of Hubei (2012FFA011) to X.M. and Special Financial Grant from the China Postdoctoral Science Foundation (2015T80806) to J.G.

Acknowledgements

We are grateful to all members of the Miao lab for helpful suggestions.

Conflict of interest statement. None declared.

References

1. ChinaLiu, S. et al. (2015) Incidence and mortality of colorectal cancer in China, 2011. Chin. J. Cancer Res., 27, 22–28.
2. Chen, W. et al. (2015) Annual report on status of cancer in China, 2011. Chin. J. Cancer Res., 27, 2–12.
3. Dai, Z. et al. (2012) Analysis and prediction of colorectal cancer incidence trend in China. Zhonghua Yu Fang Yi Xue Za Zhi, 46, 598–603.
4. Ladabaum, U. et al. (2014) Colorectal cancer incidence in Asian populations in California: effect of nativity and neighborhood-level factors. Am. J. Gastroenterol., 109, 579–588.
5. Walter, V. et al. (2014) Smoking and survival of colorectal cancer patients: systematic review and meta-analysis. Ann. Oncol., 25, 1517–1525.
6. Zhu, B. et al. (2014) Allium vegetables and garlic supplements do not reduce risk of colorectal cancer, based on meta-analysis of prospective studies. Clin. Gastroenterol. Hepatol., 12, 1991–2001.

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022

7. Lichtenstein, P. et al. (2000) Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N. Engl. J. Med., 343, 78–85.
8. Pardini, B. et al. (2015) Polymorphisms in microRNA genes as predictors of clinical outcomes in colorectal cancer patients. Carcinogenesis, 36, 82–86.
9. Geva, R. et al. (2015) FCGR polymorphisms and cetuximab efficacy in chemorefractory metastatic colorectal cancer: an international consortium study. Gut, 64, 921–928.
10. Jia, W.H. et al. (2013) Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat. Genet., 45, 191–196.
11. Fernandez-Rozadilla, C. et al. (2014) A genome-wide association study on copy-number variation identifies a 11q11 loss as a candidate susceptibility variant for colorectal cancer. Hum. Genet., 133, 525–534.
12. Peters, U. et al. (2013) Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. Gastroenterology, 144, 799–807.
13. Broderick, P. et al. (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat. Genet., 39, 1315–1317.
14. Houlston, R.S. et al. (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat. Genet., 42, 973–977.
15. Houlston, R.S. et al. (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat. Genet., 40, 1426–1435.
16. Tenesa, A. et al. (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat. Genet., 40, 631–637.
17. Tomlinson, I. et al. (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat. Genet., 39, 984–988.
18. Tomlinson, I.P. et al. (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat. Genet., 40, 623–630.
19. Zanke, B.W. et al. (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat. Genet., 39, 989–994.
20. Welter, D. et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res., 42, D1001– D1006.
21. Rinn, J.L. et al. (2012) Genome regulation by long noncoding RNAs. Annu. Rev. Biochem., 81, 145–166.
22. Geisler, S. et al. (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat. Rev. Mol. Cell Biol., 14, 699–712.
23. Sun, Q.L. et al. (2015) Expression profile analysis of long non-coding RNA associated with vincristine resistance in colon cancer cells by next-generation sequencing. Gene, 572, 79–86.
24. Yang, X. et al. (2014) Long non-coding RNA HNF1A-AS1 regulates proliferation and migration in oesophageal adenocarcinoma cells. Gut, 63, 881–890.
25. Poliseno, L. et al. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature, 465, 1033–1038.
26. Li, J.H. et al. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res., 42, D92–D97.
27. Berndt, S.I. et al. (2008) Pooled analysis of genetic variation at chromosome 8q24 and colorectal neoplasia risk. Hum. Mol. Genet., 17, 2665–2672.
28. Ling, H. et al. (2013) CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. Genome Res., 23, 1446–1461.
29. Zhang, X. 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.
30. Tao, R. et al. (2015) Association between indel polymorphism in the promoter region of lncRNA GAS5 and the risk of hepatocellular carcinoma. Carcinogenesis, 36, 1136–1143.
31. Wu, H. et al. (2013) A genetic polymorphism in lincRNA-uc003opf.1 is associated with susceptibility to esophageal squamous cell carcinoma in Chinese populations. Carcinogenesis, 34, 2908–2917.
32. Hu, X. et al. (2014) A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. Cancer Cell, 26, 344–357.
33. Barrett, J.C. et al. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, 21, 263–265.
34. Volders, P.J. et al. (2015) An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res., 43, 4363–4.
35. Gong, J. et al. (2015) lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. Nucleic Acids Res., 43, D181–D186.
36. Zhu, B. et al. (2015) Genetic variants in the SWI/SNF complex and smoking collaborate to modify the risk of pancreatic cancer in a Chinese population. Mol. Carcinog., 54, 761–768.
37. Ke, J. et al. (2015) Identification of a potential regulatory variant for colorectal cancer risk mapping to chromosome 5q31.1: a post-GWAS study. PLoS One, 10, e0138478.
38. Price, A.L. et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet., 38, 904–909.
39. Howie, B.N. et al. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet., 5, e1000529.
40. Li, Q. et al. (2013) Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell, 152, 633–641.
41. Purcell, S. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet., 81, 559–575.
42. Garg, M. et al. (2014) LAMC2 as a therapeutic target for cancers. Expert Opin. Ther. Targets, 18, 979–982.
43. Bhan, A. et al. (2013) Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. J. Mol. Biol., 425, 3707–3722.
44. Sehgal, L. et al. (2014) FAS-antisense 1 lncRNA and production of soluble versus membrane Fas in B-cell lymphoma. Leukemia, 28, 2376–2387.
45. Guil, S. et al. (2012) Cis-acting noncoding RNAs: friends and foes. Nat. Struct. Mol. Biol., 19, 1068–1075.
46. Garg, M. et al. (2014) Laminin-5gamma-2 (LAMC2) is highly expressed in anaplastic thyroid carcinoma and is associated with tumor progression, migration, and invasion by modulating signaling of EGFR. J. Clin. Endocrinol. Metab., 99, E62–E72.
47. Kosanam, H. et al. (2013) Laminin, gamma 2 (LAMC2): a promising new putative pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues. Mol. Cell Proteomics, 12, 2820–2832.
48. Shou, J.Z. et al. (2008) Overexpression of CDC25B and LAMC2 mRNA and protein in esophageal squamous cell carcinomas and premalignant lesions in subjects from a high-risk population in China. Cancer Epidemiol. Biomarkers Prev., 17, 1424–1435.
49. Degen, M. et al. (2012) MAPK/ERK-dependent translation factor hyperactivation and dysregulated laminin gamma2 expression in oral dysplasia and squamous cell carcinoma. Am. J. Pathol., 180, 2462–2478.
50. Moon, Y.W. et al. (2015) LAMC2 enhances the metastatic potential of lung adenocarcinoma. Cell Death Differ., 22, 1341–1352.
51. Korbakis, D. et al. (2015) Serum LAMC2 enhances the prognostic value of a multi-parametric panel in non-small cell lung cancer. Br. J. Cancer, 113, 484–491.
52. Kwon, O.H. et al. (2011) Aberrant up-regulation of LAMB3 and LAMC2 by promoter demethylation in gastric cancer. Biochem. Biophys. Res. Commun., 406, 539–545.
53. Gong, J. et al. (2015) An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database (Oxford), 2015, bav029.
54. Takahashi, Y. et al. (2014) Up-regulation of NEK2 by microRNA-128 methylation is associated with poor prognosis in colorectal cancer. Ann. Surg. Oncol., 21, 205–212.
55. Wu, L. et al. (2015) MicroRNA-128 suppresses cell growth and metastasis in colorectal carcinoma by targeting IRS1. Oncol. Rep., 34, 2797–2805.
56. Bak, R.O. et al. (2014) miRNA sponges: soaking up miRNAs for regulation of gene expression. Wiley Interdiscip. Rev. RNA, 5, 317–333.

Downloaded from https://academic.oup.com/carcin/article/37/5/443/1744566 by guest on 20 August 2022