OncoTargets and Therapy

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Genetic variation *rs7930* in the miR-4273-5p target site is associated with a risk of colorectal cancer

This article was published in the following Dove Press journal:
OncoTargets and Therapy
7 November 2016
Number of times this article has been viewed

Ah-Reum Lee[1]
Jongkeun Park[1]
Keum Ji Jung[2]
Sun Ha Jee[2]
Sungjoo Kim-Yoon[1]

[1]Department of Medical Life Sciences, The Catholic University of Korea, Seoul, South Korea; [2]Department of Epidemiology and Health Promotion, Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Republic of Korea

Purpose: MicroRNAs (miRNAs) are noncoding RNAs that play roles as tumor suppressors or oncogenes by regulating the expression of target genes via binding to seed-match sequences. Polymorphisms in the miRNA-binding site of a target gene can alter miRNA binding and potentially affect the risk of cancer. The objective of this study was to identify single-nucleotide polymorphisms (SNPs) in miRNA-binding sites and assess their involvement in the risk of colorectal cancer (CRC).

Materials and methods: SNPs in the 3′ untranslated regions of genes were selected and assessed for their effects on CRC risk in Korean population using participants in Korean Cancer Prevention Study-II. A detailed study was carried out with the SNP *rs7930* in the 3′ untranslated region of the translocase of outer mitochondrial membrane 20 (*TOMM20*) gene. A case–control study (1,545 controls and 620 CRC cases) was conducted to analyze the relationship between polymorphism at *rs7930* and the risk of CRC. An interacting miRNA was predicted using web-based software programs, and its interaction with *rs7930* in CRC cell lines was investigated by using a luciferase assay.

Results: Individuals carrying the *rs7930* AG genotype (G allele) had a 1.721-fold increased risk for CRC in comparison with those with the AA genotype (A allele). The miRNA miR-4273-5p was found to specifically interact with the A allele of *rs7930* and to suppress the expression of the target gene (*TOMM20*) in CRC cell lines.

Conclusion: *rs7930* is an independent genetic risk factor for CRC susceptibility. Our study suggests a mechanism of how this SNP contributes to CRC carcinogenesis.

Keywords: colorectal cancer, miR-4273-5p, SNP, *rs7930*, frequency, susceptibility

Introduction

Colorectal cancer (CRC) is the third most common cancer and the fourth leading cause of cancer-related deaths, accounting for 1–2 million new cases and 600,000 deaths per year worldwide.[1,2] Like most other cancers, no single risk factor accounts for most CRC cases.[3,4] A significant proportion of cases are attributed to heritable genetic conditions that influence not only cancer susceptibility but also its clinical course.[5]

Single-nucleotide polymorphisms (SNPs) can be used not only as markers for high-throughput screening in genome-wide association studies but also as markers for personalized diagnosis and prediction of disease risks. Furthermore, SNPs can have functional consequences, thus providing a means to understand the mechanism(s) of how genetic factors contribute to disease development.

Functional SNPs can be found in regions where they affect gene structure or expression; such regions include coding sequences, splice junctions, 5′ and 3′ untranslated regions (UTRs), and promoters. SNPs located in the 3′ UTRs of the genes have recently attracted attention because they may affect interactions with microRNAs (miRNAs),

Correspondence: Sungjoo Kim-Yoon
Department of Medical Life Science, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 137-701, Republic of Korea
Tel +822 2258 7474
Fax +82 2594 2385
Email sjkyoon@catholic.ac.kr

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/OTT.S108787

whose relationship with tumorigenesis is currently a hot topic of research.[6,7]

miRNAs affect gene expression by binding to the 3′ UTR of the target mRNA through a seed-match region, leading to translational repression or cleavage of the target mRNA. The same mRNA can be regulated by more than one miRNA, and one miRNA can target and regulate the levels of numerous mRNAs. Thus, functional variations such as SNPs located in the miRNA seed-match region can affect binding affinity of miRNAs to target mRNAs causing differential regulation of target gene expression and alter many molecular pathways simultaneously. As some of these pathways may be associated with tumorigenesis,[8] the interaction between miRNAs and SNP-bearing seed-match sequences can eventually affect individual's cancer risk and can be used for prognostic prediction. A number of such miRNA–SNP pairs have been found for various cancers and some of those are as follows: pre-miR-146a–*rs2910164* for thyroid cancer;[9] miR-107–*rs2296616* and miR-146a–*rs2910164* for gastric cancer;[10,11] miR-199a–*rs2057482* for pancreatic cancer;[12] miR-196a-2–*rs11614913*, miR-520a–*rs141178472*, and miR-27a–*rs895819* for CRC;[13–15] miR-196a–*rs11614913* for lung cancer;[16] and miR-125b–*rs1434536* for breast cancer.[17]

In this study, we sought to find the SNPs that are located in the 3′ UTRs of genes and to assess their effects on susceptibility to CRC. Among the identified SNPs, we further investigated the relationship between *rs7930* and the putative interacting miRNA miR-4273-5p. We demonstrated that this SNP located in the 3′ UTR changed the expression of the target gene, which may affect CRC risk of an individual.

Materials and methods

Cell culture

Human colorectal carcinoma cell lines (DLD1, HCT-116, SW620, SNUC4, SNUC5, and LoVo) were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS) and 100 U/mL penicillin and streptomycin. CCD18CO, a human colon fibroblast cell line, was cultured in minimum essential medium with 10% FBS and 100 U/mL penicillin and streptomycin. Cells were grown in a 37°C incubator with 5% CO_2. The biospecimens for this study were provided by the Ajou Human Bio-Resource Bank, a member of the National Biobank of Korea, which is supported by the Ministry of Health and Welfare. All samples derived from the National Biobank of Korea were obtained with written informed consent under the institutional review board of each member of the National Biobank of Korea.

Study population

The study population in the initial analysis has been described previously.[18] Briefly it is composed of 325 patients with confirmed CRC from Korean Cancer Prevention study II and 977 normal individuals as the control group. The genotypes of the patients and the control participants were previously determined using the Affymetrix Human SNP Array 5.0 (Affymetrix, Santa Clara, CA, USA),[19] and the demography of the study subjects have been described previously.[18]

For *rs7930* analysis, 297 confirmed CRC patients and 600 normal individuals were added to the already existing study population, resulting in a total of 620 CRC patients and 1,545 controls (Table S1). The cases were obtained from the Ajou Human Bio-Resource Bank (Suwon, Republic of Korea). CRC patients were diagnosed based on the International Classification of Disease for Oncology at Ajou Hospital. The control participants were selected from the Ansung–Ansan cohort study population[20] of the Biobank for Health Science of Center for Genome Science in Korea National Institute of Health, and the following criteria were applied: 1) age between 50 and 70 years old, 2) no past medical history of any carcinomas, and 3) no family history of cancers. The genotypes of the control participants for *rs7930* were obtained from the previously determined genotypes using the Affymetrix Human SNP Array 5.0 (Affymetrix) – the same platform as in the initial analysis.[21,22]

Genotyping

Genomic DNA was extracted from cells or tissues from CRC patients using the QIAamp DNA Mini Kit (QIAGEN, Valencia, CA, USA). Genomic DNA was used for genotyping of *rs7930* by the SNaPshot method using SNaPshot kit or by sequencing using 3730XL DNA analyzer (Applied Biosystems, Foster City, CA, USA).

Statistical analysis

The χ^2 test was used to examine the differences in the distribution of genotype and allele frequencies between cases and controls. Hardy–Weinberg equilibrium (HWE) was also tested in the total study participants using χ^2 analysis. A logistic regression analysis was used to evaluate the CRC risk associated with the genotype and alleles of SNPs in comparison to control participants. The odds ratio (OR) and 95% confidence intervals (CIs) of CRC associated with *rs7930* genotype was computed and adjusted for age, sex, and body mass index (BMI). Missing data were present only in genotypes, and the individuals with missing genotypes were

Table 1 Distribution of SNP genotype and allele frequencies in CRC and control participants in this study

SNP	Gene	Group (n)	Genotype			*P*-value[a]	Allele		*P*-value[a]	*P* for HWE
			Frequency				Frequency			
rs8117825	*ZNF831*		TT	TG	GG		T	G		
		Control (977)	0.999	0.001		**0.00001**	0.999	0.001	**0.00001**	0.98723
		Case (298)	0.913	0.087			0.956	0.044		0.43104
rs4985036	*ABAT*		GG	GA	AA		G	A		
		Control (976)	0.985	0.015		**0.00019**	0.992	0.008	0.01001	0.80884
		Case (324)	0.957	0.043			0.978	0.022		0.69102
rs9970671	*ST6GALNAC3*		GG	GA	AA		A	G		
		Control (977)	0.984	0.016		0.02682	0.988	0.012	0.56105	0.79636
		Case (325)	0.969	0.031			0.985	0.015		0.77819
rs16853287	*SLC9A9*		AA	AG	GG		A	G		
		Control (977)	0.966	0.033		0.02379	0.983	0.017	0.07345	0.70567
		Case (325)	0.945	0.052	0.003		0.971	0.029		0.48802
rs7930	*TOMM20*		AA	AG	GG		A	G		
		Control (945)	0.968	0.032		**0.00022**	0.984	0.016	0.00994	0.62002
		Case (323)	0.932	0.068			0.966	0.034		0.52632
rs11861556	*PDP2*		GG	GA	AA		G	A		
		Control (975)	0.954	0.045	0.001	0.28339	0.976	0.024	0.04207	0.47078
		Case (325)	0.92	0.08			0.96	0.04		0.45256
rs17500814	*ENAH*		AA	AG	GG		A	G		
		Control (951)	0.975	0.024	0.001	**0.00001**	0.987	0.013	0.00059	0.57133
		Case (323)	0.926	0.074			0.963	0.037		0.72154
rs12678	*PITRM1*		GG	GA	AA		G	A		
		Control (975)	0.914	0.083	0.003	**0.00001**	0.955	0.045	0.02063	0.42613
		Case (324)	0.864	0.13	0.006		0.931	0.069		0.75702
rs9129	*PITRM1*		GG	GC	CC		G	C		
		Control (977)	0.916	0.083	0.003	**0.00035**	0.955	0.045	0.0101	0.42396
		Case (325)	0.862	0.132	0.006		0.928	0.072		0.80358
rs2561819	*LIFR*		TT	TA	AA		T	A		
		Control (950)	0.944	0.055	0.001	**0.00001**	0.972	0.028	0.08849	0.7846
		Case (325)	0.92	0.077	0.003		0.958	0.042		0.54056

Notes: *P* for HWE; analysis on total study participants. [a]*P*-values adjusted for multiple comparison with Bonferroni correction. $P<0.0005$ was considered significant (α: 0.05, *R*: 10, *r*: 0). The statistical significance is denoted in bold.
Abbreviations: SNP, single-nucleotide polymorphism; HWE, Hardy–Weinberg equilibrium; CRC, colorectal cancer.

excluded from the analysis (Table 1). We applied Bonferroni correction to adjust the *P*-values for the multiple comparisons in all analysis. Using the online power and sample size calculators (http://powerandsamplesize.com), we analyzed the power of this study and found that this sample size had power of 0.7 with χ^2 test of two-sided and false discovery rate =0.05. Statistical analysis for the other experiments was carried out using a one-way analysis of variance. All tests were considered statistically significant at $P<0.05$.

Construction of reporter plasmid

To construct the luciferase reporter plasmids, the full-length 3′ UTR fragment of human *TOMM20* was amplified from genomic DNA of AG genotype cell line because *TOMM20* 3′ UTR was contained in a single exon. The amplified fragment was cloned in pGEM-T easy vector (Promega Corporation, Fitchburg, WI, USA) and further cloned by inserting the fragment into psiCEHCK-2 dual luciferase vector (Promega Corporation) using the Eco*R*I restriction site. The A or G allele was verified by sequencing, and the A allele bearing construct was named psiCHECK/h_TOMM20_3′ UTR_A and the G allele bearing construct psiCHECK/h_TOMM20_3′ UTR_G. The sequences for oligonucleotides used for cloning are as follows: F: 5′-GGC CCT TTT CAT TGG GTA CT3′; R: 5′GAA AGC CCA GCT CTC TCC TT-3′. The reporter construct in a dual luciferase reporter vector was used for luciferase assay in the presence of a miR-4273-5p mimic or a negative control RNA.

Transfection and luciferase assay

For transfection of a reporter plasmid, cells were plated on a six-well plate at a density of 1.5×10^5 cells/well for DLD1, and 2×10^5 cells/well for HCT116, SW620, and SNUC4 cell lines. Cells were grown overnight and transfected with 25 nM of miRNA mimic using DharmaFECT transfection reagent (Thermo Fisher Scientific, Waltham, MA, USA)

following the manufacturer's instructions. After 24 hours, 1 μg of the reporter plasmid was introduced to the cells using lipofectamine2000 transfection reagent (Thermo Fisher Scientific). The firefly luciferase and renilla luciferase activities were measured at 48 hours after reporter plasmid transfection using the Dual-Luciferase Reporter Assay System (Promega Corporation) and a Luminometer (SpectraMax L, Sunnyvale, CA, USA). Activity of constitutive firefly luciferase in a separate operon served as a transfection control. The renilla luciferase activity was normalized against firefly luciferase activity. The activities are the average of three experiments performed in triplicates.

Western blot analysis

Cells were grown and transfected with 25 nM of miRNA mimic or negative control RNA using DharmaFECT reagent. At 72 hours after transfection, cells were harvested and proteins extracted using radio immunoprecipitation assay lysis buffer (150 mM NaCl, 1% NP-40, 0.5% Na-DCA, 0.1% SDS, 50 mM Tris-HCl [pH 8.0], and protease inhibitor) and quantified by the Bradford assay method (Quick Start™ Bradford Protein Assay; Bio-Rad Laboratories Inc., Hercules, CA, USA). Thirty micrograms of protein was separated on 12% SDS-polyacrylamide gel and then transferred to nitrocellulose membrane (GE Healthcare Life Sciences Whatman™, Little Chalfont, Buckinghamshire, UK). The membrane was incubated with a mouse anti-TOMM20 monoclonal antibody (dilution, 1:5,000, Abcam, Cambridge, UK), or a mouse anti-β-actin monoclonal antibody (dilution, 1:5,000, ABM, Vancouver, BC, Canada,) overnight at 4°C and further incubated with a HRP-conjugated secondary antibody (goat anti-mouse 1:5,000, Santa Cruz Biotechnology Inc., Dallas, TX USA) for 1 hour at room temperature. Signals were detected using a enhanced chemiluminescence solution (Thermo Fisher Scientific). The β-actin gene (*ACTB*) was used as a loading control. The band intensity was analyzed using ImageJ software (version 1.49; National Institutes of Health, Bethesda, MD, USA) and defined as the ratio of target protein relative to β-actin gene.

Results

Association between SNPs in 3′ UTRs and CRC

To search for SNPs associated with CRC, we selected SNPs located in the 3′ UTRs of genes in the Genome-Wide Human SNP Array 5.0 (Affymetrix), and the χ^2 and logistic regression analyses were carried out using the genotype data from the Korean Cancer Prevention Study-II.[23] We found 123 SNPs in 111 genes with statistically significant OR ($P<0.05$) for CRC risk without adjustment for confounding factors. Of these, the top ten risk SNPs were found in nine genes, which encode zinc finger protein 831 (*ZNF831*), 4-aminobutyrate aminotransferase (*ABAT*), ST6 (α-*N*-acetylneuraminyl-2,3-β-galactosyl-1,3)-*N*-acetylgalactosaminide α-2,6-sialyltransferase 3 (*ST6GALNAC3*), solute carrier family 9, subfamily A, member 9 (*SLC9A9*), translocase of outer mitochondrial membrane 20 (*TOMM20*), pyruvate dehydrogenase phosphatase catalytic subunit 2 (*PDP2*), enabled homologue (*ENAH*), pitrilysin metallopeptidase 1 (*PITRM1*), and leukemia inhibitory factor receptor α (*LIFR*). With the exception of *rs4985036* whose minor allele frequency was 0.6%, minor allele frequencies of all the other SNPs were higher than 1% in the study population. All SNPs were in HWE in both controls and cases. The distributions of genotypes and allele frequency for these SNPs in the study population are summarized in Table 1. The genotypes of *rs8117825*, *rs4985036*, *rs7930*, *rs17500814*, *rs12678*, *rs9129*, and *rs2561819* SNPs showed significant association with CRC.

Because homozygous individuals were either not found or rarely present for these SNPs in the study population, the logistic regression analysis was performed in dominant model. As listed in Table 2, adjusted logistic regression analyses of ten SNPs showed that only six SNPs (*rs8117825*, *rs7930*, *rs17500814*, *rs12678*, *rs9129*, and *rs2561819*) were associated with elevated risk for CRC. Affected genotypes of six SNPs showed increased risks ranging from 30.12 to 1.262 compared with the other homozygote genotype (Table 2). However, after Bonferroni correction, only *rs8117825* and *rs7930* remained significant. After adjustment for age, sex, and BMI (the variables that affect the risk of CRC and that hold in our laboratory), *rs8117825* TG and GG genotype showed a 30-fold increased risk (95% CI: 4.22–310.2, $P=0.0011$) compared with the TT genotype. Also, the AG genotype of *rs7930* showed a 1.72-fold increased risk for CRC (95% CI: 1.11–3.24, $P=0.00623$) compared with the AA genotype.

To identify miRNAs interacting with these ten SNPs, two online software programs (http://www.bioguo.org/miRNASNP/ and http://compbio.uthsc.edu/miRSNP/) were used. Among the ten SNPs, *rs8117825*, *rs1685328*, and *rs7930* were the only ones for which both programs predicted the same interacting miRNAs (Table S2). In this study, we focused only on *rs7930* because it is located in the 3′ UTR of *TOMM20*, the role of which in tumorigenesis has not been studied extensively.

Table 2 Association between CRC risks and top ten high-risk SNPs

SNP	Gene	Genotype	OR (95% CI)[a]	*P*-value[b]
rs8117825	*ZNF831*	TT	1.00	**0.00111**
		TG	30.12 (4.22–310.2)	
rs4985036	*ABAT*	GG	1.00	0.06786
		GA	3.11 (0.98–1.12)	
rs9970671	*ST6GALNAC3*	GG	1.00	0.09259
		GA	2.124 (0.99–2.48)	
rs16853287	*SLC9A9*	AA	1.00	0.07212
		AG + GG	1.997 (0.99–3.22)	
rs7930	*TOMM20*	AA	1.00	**0.00623**
		AG	1.721 (1.11–3.24)	
rs11861556	*PDP2*	GG	1.00	0.05141
		GA + AA	1.626 (0.97–2.18)	
rs17500814	*ENAH*	AA	1.00	0.04212
		AG + GG	1.525 (1.210–2.924)	
rs12678	*PITRM1*	GG	1.00	0.03149
		GA + AA	1.493 (1.23–2.69)	
rs9129	*PITRM1*	GG	1.00	0.04102
		GC + CC	1.422 (1.21–2.65)	
rs2561819	*LIFR*	TT	1.00	0.01209
		TA + AA	1.262 (1.12–3.17)	

Notes: [a]OR adjusted for age, sex, and BMI. [b]*P*-values adjusted for multiple comparisons with partial Bonferroni correction $P<0.00629$ for significance (α; 0.05, *R*: 10, *r*: 0.1, *df*, 1). The statistical significance is denoted in bold.
Abbreviations: CRC, colorectal cancer; SNP, single-nucleotide polymorphism; OR, odds ratio; CI, confidence interval; *df*, degrees of freedom; BMI, body mass index.

Rs7930 is an independent genetic factor for CRC risk

Initial analysis of the association of *rs7930* with CRC risk was carried out in 323 CRC patients and 945 normal individuals by the χ^2 test as previously reported.[18] A significant association was found between *rs7930* and CRC. The frequencies of *rs7930* AA and AG genotypes were 0.932 and 0.068 in CRC patients and 0.968 and 0.032 in controls, respectively (P=0.00022; Table 1).

Logistic regression analysis revealed that the AG and GG genotype was associated with a 1.721-fold increased CRC risk compared with the AA genotype after adjustment for age, sex, and BMI (95% CI: 1.11–3.24, P=0.0062; Table 2). There was no individual with the GG genotype in the study population.

We additionally analyzed 297 patients and 600 normal individuals to increase our confidence in this finding (thus, 620 CRC patients and 1,545 controls were evaluated in total) and found similar results (Table 3). Distribution of the *rs7930* genotypes and alleles was in HWE for both controls and cases. Individuals with the AG genotype had a 1.9-fold (95% CI: 1.191–3.034, P=0.0084) higher risk for CRC development after adjustment for age, sex, and BMI than individuals with the AA genotype. There were no individuals with the GG genotype in this expanded study population. Thus, the G allele in heterozygous individuals

Table 3 Distribution of genotype and allele frequencies of *rs7930* in CRC and control participants

Study genotype/allele	Control n (%)	Case n (%)	*P*-value[a]	OR[b]	95% CI	*P*-value[c]
Study set						
Total	1,545 (100)	620 (100)				
Genotype						
AA	1,502 (97.2)	588 (94.8)	0.0062	1.00	1.191–3.034	0.0084
AG	43 (2.8)	32 (5.2)		1.901		
Allele						
A	3,047 (98.6)	1,208 (97.4)	0.0067	1.00	1.501–3.126	0.0005
G	43 (1.4)	32 (2.6)		2.535		
P for HWE	0.579098	0.509512				

Notes: [a]χ^2 test. [b]OR adjusted for age, sex, and BMI. [c]Logistic regression analysis. $P<0.05$ is considered statistically significant.
Abbreviations: CRC, colorectal cancer; OR, odds ratio; CI, confidence interval; BMI, body mass index; HWE, Hardy–Weinberg equilibrium.

was sufficient to increase CRC susceptibility (OR =2.535, 95% CI: 1.501–3.126, *P*=0.0005) after adjustment for age, sex, and BMI (Table 3). These results suggest that the SNP *rs7930* is an independent genetic factor for CRC susceptibility and that G is the risk allele.

Allele-dependent interaction of miR-4273-5p with its target

The miRNA miR-4273-5p was predicted to interact with *rs7930* (Figure 1A). We investigated whether miR-4273-5p directly interacts with *rs7930* using a luciferase assay in the CRC cell lines SW620, SNUC4, DLD1, and HCT116. Although the negative control RNA had no effect on luciferase activity regardless of the allele of the reporter construct, the miR-4273-5p mimic affected luciferase activity differently depending on the allele of the reporter construct (Figure 1B): the G allele plasmid had no effect, whereas the A allele plasmid significantly suppressed luciferase activity in comparison with the control RNA in three cell lines (by 60%–80%; SW620, *P*<0.001; DLD1, *P*=0.019; HCT116, *P*=0.021) and tended to decrease luciferase activity in one cell line (SNUC4, *P*=0.074).

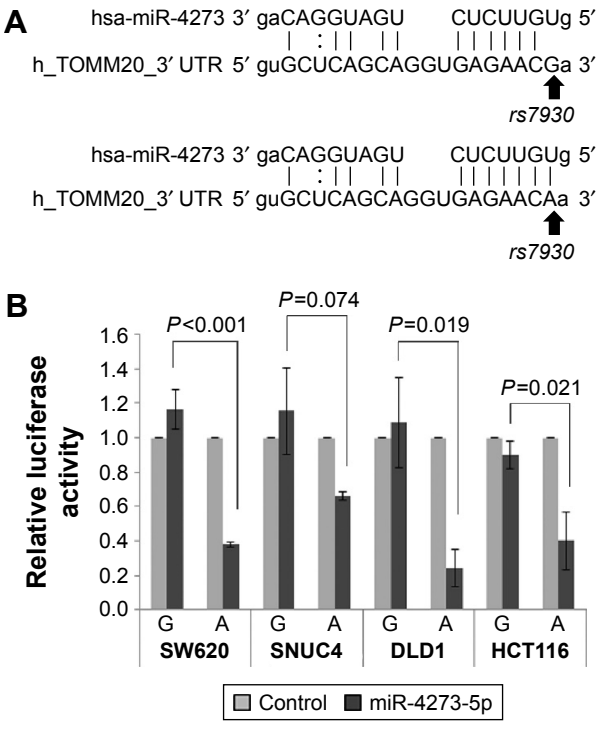

Figure 1 A decrease in the expression of the *TOMM20* A allele construct relative to that of the G allele construct by miR-4273-5p demonstrated by a luciferase assay.
Notes: (**A**) Schema of miR-4273-5p selection based on interaction prediction using two website-based software programs. (**B**) Relative renilla luciferase activity normalized against firefly luciferase activity. Each CRC cell line (SW620, SNUC4, DLD1, and HCT116) was transfected with the control mimic or miR-4273-5p mimic and then each of the reporter constructs with the G or A allele of *rs7930* in the *TOMM20* 3′ UTR. A statistically significant decrease in expression was seen with psiCHECK/h_TOMM20_3′ UTR_A allele in all cell lines.
Abbreviations: TOMM20, translocate of outer mitochondria membrane; UTR, untranslation region; CRC, colorectal cancer.

Genotype-dependent regulation of *TOMM20* expression by miR-4273-5p in CRC cell lines

To investigate the effect of the *rs7930* genotype on the regulation of *TOMM20* expression by miR-4273-5p, we first determined the *rs7930* genotypes of various cell lines by polymerase chain reaction on genomic DNA, followed by sequencing. The cell lines CCD18CO, SW620, SNUC4, and SNUC5 had the AA genotype, whereas DLD, HCT116, and LoVo had the AG genotype. Unfortunately, we did not find a CRC cell line with the GG genotype (Table S3). TOMM20 expression was determined at the protein level by Western blot analysis. The TOMM20 levels were higher in all CRC cell lines than in CCD18CO (Figure 2).

Regulation of TOMM20 expression by the miR-4273-5p mimic depended on *rs7930* genotype (*P*=0.043). Cell lines with the AA genotype showed a considerably stronger reduction in TOMM20 levels (SNUC4 and SNUC5, both by 60%) than those with the AG genotype (DLD1, by 24.3%; HCT116, 37%; LoVo, 31%) in comparison with the control (Figure 3A). Transfection with the miR-4273-5p inhibitor resulted in a more than 1.76-fold increase in TOMM20 expression in all CRC cell lines in comparison with that in cells transfected with the control RNA (Figure 3B).

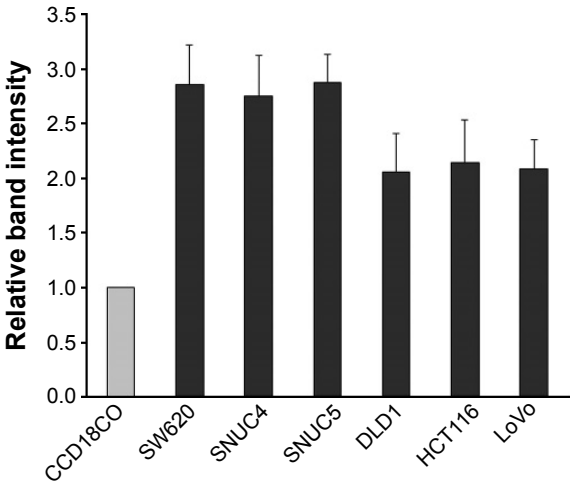

Figure 2 TOMM20 protein expression levels in CRC cell lines.
Note: Quantifications of Western blot analysis of TOMM20 expression relative to β-actin in CRC cells.
Abbreviations: CRC, colorectal cancer; TOMM20, translocate of outer mitochondria membrane.

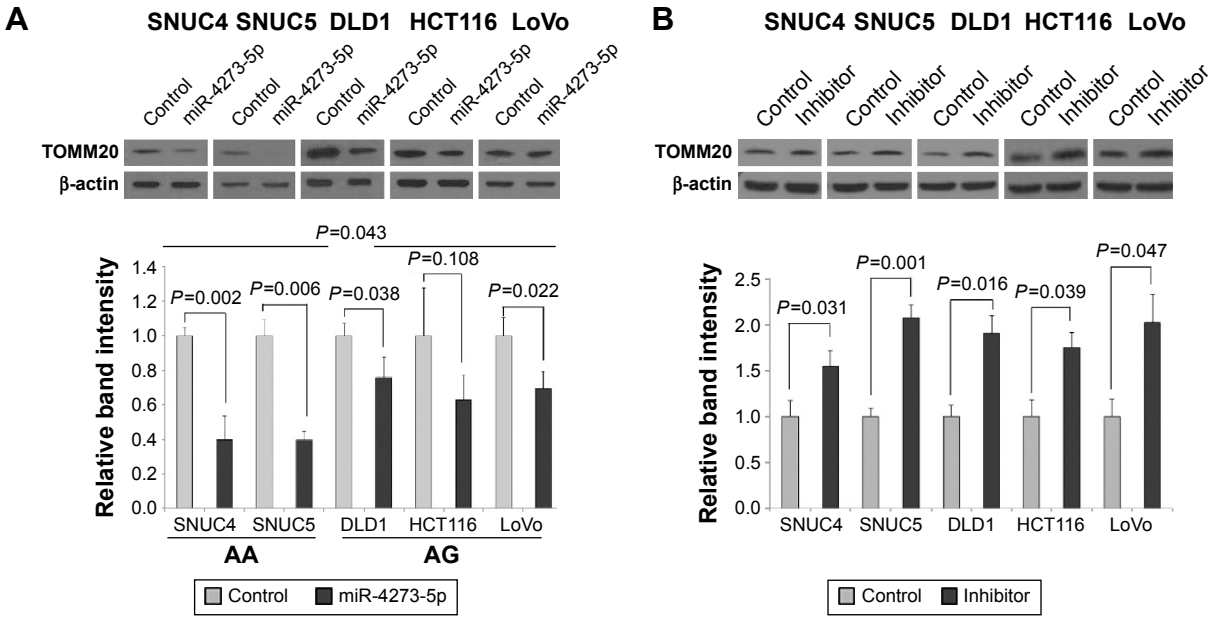

Figure 3 SNP *rs7930* variations in the *TOMM20* 3′ UTR affects *TOMM20* expression via interaction with miR-4273-5p.
Notes: (**A**) Regulation of the TOMM20 protein expression level according to the *rs7930* genotype by miR-4273-5p in CRC cell lines. (**B**) The level of the TOMM20 protein was reversed by a miR-4273-5p inhibitor.
Abbreviations: SNP, single-nucleotide polymorphism; TOMM20, translocate of outer mitochondria membrane; CRC, colorectal cancer; UTR, untranslation region.

Discussion

The SNPs and associated genes identified in this study may function in CRC and provide opportunities to understand carcinogenesis, to develop diagnostic and prognostic markers, and to discover therapeutic approaches. The six risk SNPs in five genes we detected have not been previously implicated in any cancers, including CRC. However, some of these genes associated with high risk for CRC have been implicated in cancers. *ENAH* and *LIFR* have been previously shown to be related to CRC. Increased expression of *ENAH*, a transcriptional target of the Wnt/β-catenin pathway, was correlated with advanced CRC stages.[24] Downregulated *LIFR* expression was shown in primary CRC tissues. Both genes have also been implicated in many other carcinomas such as myeloid leukemia or breast cancer.[25–27] Therefore, the SNPs *rs17500814* and *rs2561819* can be candidate susceptibility markers, and interactions of the *ENAH* and *LIFR* mRNAs with miRNAs should be further investigated.

Although *rs4985036* (in *ABAT*) and *rs16853287* (in *SLC9A9*) were not significantly associated with CRC risk, nevertheless they show tendency for increased CRC risk. It is worthy to note that decreased *ABAT* expression was observed in breast cancer;[28] *SLC9A9* expression was downregulated in prostate cancer, and this gene was suggested as a prognostic marker for esophageal squamous cell cancer.[29] *TOMM20* is suggested as a prognostic marker for patients with gastric cancer[30] and as a specific therapeutic target for anaplastic thyroid cancer cells.[31] Therefore, the relevance of these genes to CRC should be established. Establishment of interactions of *ABAT* and *SLC9A9* mRNAs with miRNAs may lead to new pairs of SNP–miRNA indicator for CRC risk. *ZNF831*, *ST6GALNAC3*, *PDP2*, and *PITRM1* have not been specifically associated with any cancers. *rs8117825* (in *ZNF831*) and *rs12678* and *rs9129* (in *PITRM1*) were found significantly associated with increased CRC risk. However, *rs9970671* (in *ST6GALNAC3*) and *rs11861556* (in *PDP2*) only showed marginal tendency for CRC risk. Interestingly, these genes show variable expression in tumors (from the public data posted by the Broad Institute; http://firebrowse.org/). In CRC, the expression of *ZNF831* and *ST6GALNAC3* was found to be lower than in normal tissues (Table S4). *ABAT* expression was slightly decreased in colon cancer but not in rectal cancer, whereas the expression of *SLC9A9* was markedly lower in CRC than in normal tissues. Therefore, the SNPs in these genes are attractive candidates for the development of biomarkers for diagnosis and prognosis of CRC. However, currently, the functional relevance of *PITRM1* and *PDP2* to CRC is not clear. The expression of *PITRM1* was slightly higher (by 20%) in CRC than in normal tissues, whereas *PDP2* expression was similar in normal and tumor tissues, as shown in the public data (Table S4: data from http://firebrowse.org/). Thus, SNPs in the 3′ UTRs of these genes may act simply as surrogate markers of genetic contributors to CRC rather than play a role in CRC carcinogenesis. It may

be that r*s11861556* in PDP2 was not associated with CRC risk at all. Obviously, further study is required to understand their relationship (if any) to CRC.

Accumulating evidence has explained the specificity of miRNAs for different cancers and suggested that miRNAs can potentially be used as biomarkers for the diagnosis and prognosis of cancers. SNPs in the seed-match regions of miRNAs are a hot topic of research because of their potential usefulness as markers of noninvasiveness, their availability, and the ease of their detection. miRNA–SNP pairs have been implicated in cancer risk and suggested as prognostic markers. For CRC, an SNP in the let-7 miRNA-binding site that is present in the *KRAS* 3′ UTR is a prognostic biomarker, and *rs895819* in the miR-27a binding site is associated with increased genetic susceptibility in Han Chinese.[32,33] *rs2910164* in miR-146a binding site is associated with CRC susceptibility in Han Chinese,[34] and *rs141178472* in the *PIK3CA* 3′ UTR and miR-520a was implicated in CRC risk in Han Chinese population.[35]

This study identified *rs7930*–miR-4273-5p, a new SNP–miRNA pair, as an indicator of CRC risk; this pair has never been studied in any cancers. Our study clearly showed that miR-4273-5p interacted differentially with *rs7930* alleles and regulated expression of the target gene *TOMM20* in CRC cells. Thus, both *rs7930* and miR-4273-5p and their interaction itself have potential to be genetic contributors of CRC risks.

Our study has a few limitations. First, variables were limited to age, sex, and BMI, and so only limited analyses were possible. Since the analysis was carried out using a data set merged from various sources whose common variables were limited to age, sex, and BMI, further multivariable analysis was not possible. Although this presents a limitation for our study, the direction or the size of effect would not be affected considerably even if environmental factors are added because this study was a genetic analysis. Second, the power of our study is slightly less than optimal condition, although fairly good number of samples were analyzed. The sample size of our study gives the power of 0.7 with the genotype frequencies of controls and CRC cases, which warrants cautious interpretation of the results. Our study result suggested that individuals with the AG genotype of *rs7930* have a 1.9-fold higher risk for CRC development. This OR was calculated with 95% CI of 1.19–3.03, which gives best guess for the size of the population effect 95% of time. With regard to clinical consideration, an individual with AG genotype may have CRC risk of anywhere between 1.19 and 3.03, which is rather wide. Thus, a larger sample size is needed to obtain more reliable result with an improved estimate, or a narrower CI.

Since *TOMM20* expression is essential for cell viability, it is critically important to dissect its function in CRC carcinogenesis. Although a rigorous validation process with a larger sample size must precede its application, this SNP has a potential to be utilized for detection of high-risk group as well as early detection of CRC development, which may provide a practical means for preventive medicine.

Conclusion

In summary, we identified SNPs that have not been previously implicated in etiology of any tumors including CRC. We also demonstrated that miR-4273-5p specifically interacts with the A allele of the SNP *rs7930*, which suggests a mechanism for the association of the G allele (which is not targeted by miR-4273-5p) with an increased CRC risk. Thus, the A allele is protective, whereas G allele is a risk allele, that is, patients with the G allele have a higher risk for CRC development. Therefore, we provide a diagnostic biomarker for CRC with the *rs7930* SNP_G allele. Our study found SNPs that may be useful for CRC risk assessment and provided opportunity to elucidate the biological importance of these SNPs, which may be accomplished by dissecting the relationship between the SNPs and miRNAs.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A5A2047939 and 2015R1D1A1A2061749).

Author contributions

ARL and SJKY designed the experiments. ARL, JP, KJ, SHJ, and SJKY conducted the experiments. ARL and SJKY wrote the paper. All authors contributed toward data analysis, drafting and critically revising the paper and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interests in this work.

References

1. Cyhaniuk A, Coombes ME. Longitudinal adherence to colorectal cancer screening guidelines. *Am J Manag Care*. 2016;22(2):105–111.
2. Moravkova P, Kohoutova D, Rejchrt S, Cyrany J, Bures J. Role of S100 proteins in colorectal carcinogenesis. *Gastroenterol Res Pract*. 2016;2016:2632703.

3. Hsueh YM, Cheng GS, Wu MM, Yu HS, Kuo TL, Chen CJ. Multiple risk factors associated with arsenic-induced skin cancer: effects of chronic liver disease and malnutritional status. *Br J Cancer*. 1995;71(1):109–114.
4. Kendzor DE, Costello TJ, Li Y, et al. Race/ethnicity and multiple cancer risk factors among individuals seeking smoking cessation treatment. *Cancer Epidemiol Biomarkers Prev*. 2008;17(11):2937–2945.
5. Mohammed AA, El-Tanni H, El-Khatib HM, Mirza AA, El-Kashif AT. Molecular classification of colorectal cancer: current perspectives and controversies. *J Egypt Natl Cancer Inst*. 2016. doi:10.1016/j.jnci.2015.11.004. Epub ahead of print.
6. Song T, Zhang X, Yang G, Song Y, Cai W. Decrement of miR-199a-5p contributes to the tumorigenesis of bladder urothelial carcinoma by regulating MLK3/NF-kappaB pathway. *Am J Transl Res*. 2015;7(12):2786–2794.
7. Zheng N, Yang P, Wang Z, Zhou Q. OncomicroRNAs-mediated tumorigenesis: implication in cancer diagnosis and targeted therapy. *Curr Cancer Drug Targets*. 2016. Epub ahead of print.
8. Yin Z, Li H, Cui Z, et al. Polymorphisms in pre-miRNA genes and cooking oil fume exposure as well as their interaction on the risk of lung cancer in a Chinese nonsmoking female population. *OncoTargets Ther*. 2016;9:395–401.
9. Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc Natl Acad Sci U S A*. 2008;105(20):7269–7274.
10. Fu B, Song P, Lu M, Wang B, Zhao Q. The association between miR-146a gene rs2910164 polymorphism and gastric cancer risk: a meta-analysis. *Biomed Pharmacother*. 2014;68(8):923–928.
11. Wang S, Lv C, Jin H, et al. A common genetic variation in the promoter of miR-107 is associated with gastric adenocarcinoma susceptibility and survival. *Mutat Res*. 2014;769:35–41.
12. Wang X, Ren H, Zhao T, et al. Single nucleotide polymorphism in the microRNA-199a binding site of HIF1A gene is associated with pancreatic ductal adenocarcinoma risk and worse clinical outcomes. *Oncotarget*. 2016;7(12):13717–13729.
13. Ding L, Jiang Z, Chen Q, Qin R, Fang Y, Li H. A functional variant at miR-520a binding site in PIK3CA alters susceptibility to colorectal cancer in a Chinese Han population. *Biomed Res Int*. 2015;2015:373252.
14. Uvirova M, Simova J, Kubova B, et al. Comparison of the prevalence of KRAS-LCS6 polymorphism (rs61764370) within different tumour types (colorectal, breast, non-small cell lung cancer and brain tumours). A study of the Czech population. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*. 2015;159(3):466–471.
15. Xu L, Tang W. Associations of polymorphisms in mir-196a2, mir-146a and mir-149 with colorectal cancer risk: a meta-analysis. *Pathol Oncol Res*. 2016;22(2):261–267.
16. Hu Z, Chen J, Tian T, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J Clin Invest*. 2008;118(7):2600–2608.
17. Saetrom P, Biesinger J, Li SM, et al. A risk variant in an miR-125b binding site in BMPR1B is associated with breast cancer pathogenesis. *Cancer Res*. 2009;69(18):7459–7465.
18. Park J, Kim I, Jung KJ, Kim S, Jee SH, Yoon SK. Gene-gene interaction analysis identifies a new genetic risk factor for colorectal cancer. *J Biomed Sci*. 2015;22:73.
19. Jo J, Nam CM, Sull JW, et al. Prediction of colorectal cancer risk using a genetic risk score: the Korean Cancer Prevention Study-II (KCPS-II). *Genomics Inform*. 2012;10(3):175–183.
20. Cho NH, Kim JY, Kim SS, Shin C. The relationship of metabolic syndrome and constitutional medicine for the prediction of cardiovascular disease. *Diabetes Metab Syndr*. 2013;7(4):226–232.
21. Cha S, Park AY, Kang C. A genome-wide association study uncovers a genetic locus associated with thoracic-to-hip ratio in Koreans. *PloS One*. 2015;10(12):e0145220.
22. Cha SW, Choi SM, Kim KS, et al. Replication of genetic effects of FTO polymorphisms on BMI in a Korean population. *Obesity (Silver Spring, Md.)*. 2008;16(9):2187–2189.
23. Jung KW, Park S, Kong HJ, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2008. *Cancer Res Treat*. 2011;43(1):1–11.
24. Toyoda A, Kawana H, Azuhata K, et al. Aberrant expression of human ortholog of mammalian enabled (hMena) in human colorectal carcinomas: implications for its role in tumor progression. *Int J Oncol*. 2009;34(1):53–60.
25. Bruce AG, Hoggatt IH, Rose TM. Oncostatin M is a differentiation factor for myeloid leukemia cells. *J Immunol*. 1992;149(4):1271–1275.
26. Estrov Z, Samal B, Lapushin R, et al. Leukemia inhibitory factor binds to human breast cancer cells and stimulates their proliferation. *J Interferon Cytokine Res*. 1995;15(10):905–913.
27. Hudson KR, Vernallis AB, Heath JK. Characterization of the receptor binding sites of human leukemia inhibitory factor and creation of antagonists. *J Biol Chem*. 1996;271(20):11971–11978.
28. Jansen MP, Sas L, Sieuwerts AM, et al. Decreased expression of ABAT and STC2 hallmarks ER-positive inflammatory breast cancer and endocrine therapy resistance in advanced disease. *Mol Oncol*. 2015;9(6):1218–1233.
29. Chen J, Wen J, Zheng Y, et al. Prognostic significance of SLC9A9 in patients with resectable esophageal squamous cell carcinoma. *Tumour Biol*. 2015;36(9):6797–6803.
30. Zhao Z, Han F, He Y, et al. Stromal-epithelial metabolic coupling in gastric cancer: stromal MCT4 and mitochondrial TOMM20 as poor prognostic factors. *Eur J Surg Oncol*. 2014;40(10):1361–1368.
31. Johnson JM, Lai SY, Cotzia P, et al. Mitochondrial metabolism as a treatment target in anaplastic thyroid cancer. *Sem Oncol*. 2015;42(6):915–922.
32. Jiang Y, Lin DH, Xu JP, Chen WX, Zheng SJ, Song L. Genotype GG of rs895819 functional polymorphism within miR-27a might increase genetic susceptibility to colorectal cancer in Han Chinese population. *J Clin Lab Analysis*. 2016;30(4):351–355.
33. Smits KM, Paranjape T, Nallur S, et al. A let-7 microRNA SNP in the KRAS 3′UTR is prognostic in early-stage colorectal cancer. *Clin Cancer Res*. 2011;17(24):7723–7731.
34. Hu X, Li L, Shang M, et al. Association between microRNA genetic variants and susceptibility to colorectal cancer in Chinese population. *Tumour Biol*. 2014;35(3):2151–2156.
35. Ding L, Jiang Z, Chen Q, Qin R, Fang Y, Li H. A functional variant at miR-520a binding site in PIK3CA alters susceptibility to colorectal cancer in a Chinese Han population. *Biomed Res Int*. 2015;2015:373252.

Supplementary materials

Table S1 Demographic characteristics of CRC and normal participants of this study

	Total		*P*-value
	Normal participants (n=1,545)	CRC patients (n=620)	
Male (n)	956	406	0.116271[a]
Female (n)	589	214	
Age (years) (± standard deviation)	60.77 (±4.77)	65.42 (±14.1)	0.042277[b]
BMI (kg/m^2) (± standard deviation)	24.15 (±2.83)	23.20 (±3.34)	0.03692[b]

Notes: [a]χ^2 test. [b]Unpaired *t*-test.
Abbreviations: BMI, body mass index; CRC, colorectal cancer.

Table S2 miRNAs predicted for interaction with top ten SNPs associated with CRC risk

SNP	Gene	Prediction A	Prediction B
rs817825	*ZNF831*	hsa-miR-3126-5p hsa-miR-337-3p	hsa-miR-135a-3p
			hsa-miR-337-3p
			hsa-miR-7845-5p
			hsa-miR-3126-5p
			hsa-miR-520g-5p
			hsa-miR-6873-5p
			hsa-miR-6875-5p
rs4985036	*ABAT*		
rs9970671	*ST6GALNAC3*		hsa-miR-3663-5p
			hsa-miR-6801-5p
			hsa-miR-7152-3p
			hsa-miR-29b-5p
			hsa-miR-329-5p
			hsa-miR-4464 hsa-miR-4748
rs16853287	*SLC9A9*	hsa-miR-128-3p hsa-miR-140-3p	hsa-miR-128-3p
			hsa-miR-140-3p
			hsa-miR-216a-3p
			hsa-miR-3681-3p
rs7930	*TOMM20*	hsa-miR-4273	hsa-miR-129-5p
			hsa-miR-2117
			hsa-miR-375
			hsa-miR-4273 hsa-miR-4677-5p
			hsa-miR-6739-3p
			hsa-miR-7156-5p
rs11861556	*PDP2*		hsa-miR-1207-3p
			hsa-miR-3617-3p
			hsa-miR-3927-5p
rs17500814	*ENAH*	hsa-miR-3663-5p	
rs12678	*PITRM1*		hsa-miR-6736-3p
rs9129	*PITRM1*		hsa-let-7a-5p
			hsa-let-7b-5p
			hsa-let-7c-5p
			hsa-let-7d-5p
			hsa-let-7e-5p
			hsa-let-7f-5p
			hsa-let-7g-5p
			hsa-let-7i-5p
			hsa-miR-202-3p
			hsa-miR-4458
			hsa-miR-4500
			hsa-miR-98-5p
rs2561819	*LIFR*		

Abbreviations: SNP, single-nucleotide polymorphism; CRC, colorectal cancer; miRNAs, microRNAs.

Table S3 *rs7930* genotypes of CRC cell lines

	A/A	A/G	G/G
CCD18CO	O		
CoLo320DM	O		
DLD1		O	
HCT116		O	
HT29		O	
SNUC2B	O		
SNUC4	O		
SNUC5	O		
SW620	O		
LoVo		O	

Abbreviation: CRC, colorectal cancer.

Table S4 Relative expression of the genes in CRC compared to normal control tissues obtained from the GDAC database

	COAD			READ			COADREAD		
	Tumor (n)	Normal (n)	Fold change	Tumor (n)	Normal (n)	Fold change	Tumor (n)	Normal (n)	Fold change
ZNF831	447	41	0.303	164	10	0.274	611	51	0.299
ABAT	459	41	0.515	167	10	0.932	626	51	0.629
ST6GALNAC3	457	41	0.36	166	10	0.225	623	51	0.307
SLC9A9	457	41	0.19	167	10	0.119	624	51	0.181
PITRM1	459	41	1.21	167	10	1.21	626	51	1.2
PDP2	459	41	1.03	167	10	1.28	626	51	1.06

Abbreviations: COAD, colon adenocarcinoma; READ, rectum adenocarcinoma; COADREAD, colorectal adenocarcinoma; CRC, colorectal cancer.