Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer

Debora Landi[1,2], Federica Gemignani[1], Alessio Naccarati[3], Barbara Pardini[3], Pavel Vodicka[3], Ludmila Vodickova[3,4], Jan Novotny[5,6], Asta Försti[2,7], Kari Hemminki[2,7], Federico Canzian[2] and Stefano Landi[1,*

[1]Dipartimento di Biologia, University of Pisa, Via Derna, 1, 56126 Pisa, Italy, [2]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, [3]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Academy of Science of Czech Republic, Videnska 1083, 14220 Prague 4, Czech Republic, [4]Centre of Occupational Health, National Institute of Public Health, Srobarova, 100 42 Prague 10, Czech Republic, [5]Department of Oncology, First Faculty of Medicine, Charles University, 12000 Prague, Czech Republic, [6]Oncology Centre, Hospital 26205 Pribram, Pribram, Czech Republic and [7]Center for Family and Community Medicine, Karolinska Institute, SE-17177 Huddinge, Sweden

*To whom correspondence should be addressed. Tel: +39 050 2211505;
Fax: +39 0502211527;
Email: slandi@biologia.unipi.it

Recent evidence indicate that small non-coding RNA molecules, called micro-RNAs (miRNAs), can bind to the 3′ untranslated regions (UTRs) of messenger RNAs and interfere with their translation, thereby regulating cell growth, differentiation, apoptosis and tumorigenesis. Genetic polymorphisms can reside on miRNA-binding sites. Thus, it is conceivable that the miRNA regulation may be affected by polymorphisms on the 3′ UTRs. Since gene deregulation is one of the key mechanisms by which cells can progress to cancer, we hypothesize that common polymorphisms within miRNA-target binding sites could play a role in the individual risk of cancer. In the present study, we selected the 3′ UTRs of 104 genes candidate for colorectal cancer (CRC) and we identified putative miRNA-binding sites by specialized algorithms (PicTar, DianaMicroT, miRBase, miRanda, TargetScan and microInspector). Fifty-seven single-nucleotide polymorphisms (SNPs) were identified in miRNA-binding sites. We evaluated the SNPs for their ability to affect the binding of the miRNA with its target, by assessing the variation of Gibbs free energy between the two alleles of each SNP. We found eight common polymorphisms that were further investigated by a case–control association studies. The study was carried out on a series of cases and controls from Czech Republic, a population with the highest worldwide incidence of CRC. We found statistically significant associations between risk of CRC and variant alleles of *CD86* [odds ratio (OR) = 2.74; 95% confidence interval (CI) = 1.24–6.04, for the variant homozygotes] and *INSR* genes (OR = 1.94; 95% CI = 1.03–3.66, for the variant homozygotes). These results are the first reporting positive association between miRNA-binding SNPs sequences and cancer risk.

Introduction

In developed countries, colorectal cancer (CRC) is one of the most common types of tumor (1). Genetic background is thought to play a role in modulating individual risk (2). Among the genes investigated and associated with the risk of CRC are the genes involved in xenobiotic metabolism (3), DNA repair (4), insulin resistance, obesity and glucose levels (5) and inflammation (6).

Gene deregulation is one of the key mechanisms by which cells can progress to cancer (7). Recently, new mechanisms governing the expression of genes related to cancer development have been identified.

Abbreviations: CRC, colorectal cancer; CI, confidence interval; mRNA, messenger RNA; miRNA, micro-RNA; MAF, minor allele frequency; nt, nucleotide; OR, odds ratio; SNP, single-nucleotide polymorphism; UTR, untranslated region.

Among them, the post-transcriptional regulation through micro-RNAs (miRNAs) is one of the most interesting and powerful in the maintenance and management of the cellular differentiation. miRNAs are endogenous non-coding RNAs of ~22 nucleotides (nts), regulating genes in animals and plants by pairing to the 3′ untranslated regions (UTRs) of messenger RNAs (mRNAs) of target genes and specifying mRNA cleavage or repression of protein synthesis (8). There is increasing evidence that miRNAs have an important regulatory role in a broad range of biological processes, including developmental timing, cellular differentiation, proliferation, apoptosis, gene regulation, cancer development, insulin secretion and cholesterol biosynthesis (8). miRNA genes represent only a small part of the genome, but they regulate ~20–30% of all human genes and there is an average of 200 predicted targets per miRNA (9).

The maturation of a miRNA is a multistep process that begins in the nucleus and ends in the cytoplasm. Briefly, miRNAs are transcribed by RNA polymerase II as part of a long RNA precursor, defined as 'primary miRNA'. Within the primary miRNA, the ~22 nt mature miRNA forms part of one arm of a ~80 nt imperfect stem–loop sequence (10). miRNA processing involves the nuclear cleavage of this RNA stem–loop structure. This liberates a 60 nt RNA hairpin intermediate, bearing a characteristic 2 nt 3′ overhang, defined as 'pre-miRNA' (11). Pre-miRNA is then exported into the cytoplasm (12) and further processed by an RNase III enzyme called 'Dicer' (8). Dicer removes the terminal loop, leaving a second 2 nt 3′ overhang. The result is a miRNA duplex intermediate, comprising the mature miRNA and its opposing strand. This short duplex intermediate is separated and the single-stranded mature miRNA is incorporated into the RNA-induced silencing complex (13). The composition of RNA-induced silencing complex remains incompletely defined, but a key component is an Argonaute protein (8).

The complex miRNA–RNA-induced silencing complex can inhibit the translation when it binds the 3′ UTR-mRNA targets with an imperfect complementarity. In this case, there is a reduced level of protein without reductions in the mRNA levels. The maximal complementarity is usually restricted to the nucleotides 2–8 in the 5′ end of the miRNA and this region is called the 'seed sequence' (14). When the miRNA has a perfect complementarity with its mRNA target, there is a reduced protein level accompanied to a reduced mRNA level because the mRNA is cleaved (phenomenon known as 'RNA silencing') (15).

Recent studies have shown a link between altered miRNA expression patterns and cancer (16). Calin *et al.* (17,18) were the first to show a link with chronic lymphocytic leukemia. In 2004, Takamizawa *et al.* (19) first reported the role of let-7 in lung cancer. Moreover, altered miRNA expression patterns have been described in B-cell lymphomas (20), Burkitt's lymphoma (21), breast cancer (22), lung cancer (23), hepatocellular carcinoma (24), glioblastoma (25), follicular thyroid carcinoma (26) and CRC (27). Given the important role of miRNA in gene regulation and in carcinogenesis, we hypothesized that germ line polymorphisms in 3′ UTRs targeted by miRNAs might alter the strength of miRNA binding, with consequences on regulation of target genes thereby affecting the individual's cancer risk. To test this hypothesis, we evaluated 104 genes belonging to different pathways relevant for colorectal carcinogenesis. We searched for all the polymorphisms residing in the 3′ UTRs of the candidate genes and assessed *in silico* the abilities of each single-nucleotide polymorphism (SNP) to impact the binding between miRNAs and their target sites. In order to verify our initial hypothesis, we genotyped the resulting putatively functional polymorphisms in 697 CRC cases and 624 controls from the Czech Republic, where the incidence of colon cancer ranks the third highest worldwide and is the highest for rectal cancer (28,29).

Materials and methods

Study population

A hospital-based case–control study was conducted to assess gene–environment interactions in relation to CRC risk. Cases were incident patients with a new diagnosis of CRC attending nine oncological departments (two in Prague, one in Benesov, Brno, Liberec, Ples, Pribram, Usti nad Labem and Zlin) all over the Czech Republic from September 2004 to February 2006. All cases had histological confirmation of their tumor diagnosis. During the study period, a total of 968 cases were diagnosed with CRC in these hospitals. This study includes 697 (72%) patients who could be interviewed and provided biological samples of sufficient quality for genetic analysis. The lost cases were similar to those enrolled with respect to age, sex, tumor location and extent.

Controls were selected from among patients admitted to the same hospital during the same period and were subjects undergoing colonoscopy for various gastrointestinal complaints. The reasons for undergoing the colonoscopy were i) positive fecal occult blood test, ii) hemorrhoids, iii) abdominal pain of unknown origin and iv) macroscopic bleeding. Due to the high incidence of CRC in the Czech Republic, colonoscopy is largely recommended and practiced. To avoid selection bias, the criterion for inclusion of controls was that the reason for the current admittance to the hospital should be a new disease (not diagnosed previously) for that patient. This criterion was used to avoid inclusion of patients with chronic diseases, who might be repeatedly admitted to hospital and modify their habits because of their disease. This procedure paralleled the criterion for cases who were also newly diagnosed incident cases. Among 899 selected controls, a total of 624 (70%) were analyzed in this study. Twelve percent could not be interviewed because of refusal or mental or other impairment. Eighteen percent were interviewed but did not provide a blood sample. Controls included people who had benign lower abdominal pains, hemorrhoids with active bleeding, solitary rectal ulcer, lower gastrointestinal bleeding, diverticular disease, diarrhea and anemia. People with polyps, adenomas or other diagnoses related to cancer or to diseases known to predispose to cancer (such as, e.g. ulcerative colitis, inflammatory bowel disease and Crohn's disease) were excluded. Sex and broad age groups were used as stratifying criteria for frequency matching. Both cases and controls had to be in good mental condition, be able to see and hear and follow an interview and (for controls) not have diagnoses clearly related to cancer or chronic inflammatory diseases.

All subjects were informed and gave written consent to participate in the study to allow their biological samples to be genetically analyzed, according to the Helsinki declaration. The design of the study was approved by the Ethical Committee of the Institute of Experimental Medicine, Prague, Czech Republic.

Interviews

Cases and controls were personally interviewed by trained personnel using a structured questionnaire to determine demographic characteristics and potential risk factors for CRC. For each subject, age and sex were recorded. Study subjects provided information on their lifestyle habits, body mass index, diabetes, tentative occupational exposure to xenobiotics and family/personal history of cancer. A detailed dietary history questionnaire focused on average food consumption 1 year before the diagnosis of disease. Lifelong long-term (at least six consecutive months) drug use was included in the questionnaire. An initial open question was followed by a list of 20 chronic diseases that usually are treated pharmacologically and their treatments were recorded. No drug list was used. For each exposure, the ages at initial use and cessation were recorded and the cumulative duration was computed. Drugs were grouped using the anatomical therapeutic chemical classification. Other relevant risk factors explored were smoking, alcohol, physical activity and family history of cancer (Table I).

Selection of polymorphisms

In this study, we analyzed 104 candidate genes for CRC: 51 genes are involved in inflammatory processes, 37 belong to synthesis of prostaglandins and thromboxanes and 16 genes are connected with obesity and insulin resistance. The selected genes were among the most characterized and studied, according to the literature and the BioCarta and KEGG pathways (http://cgap.nci.nih.gov/Pathways). They are listed in Table II. For all of them, the 3′ UTRs were selected according to the University of California Santa Cruz genome browser (http://genome.ucsc.edu) and defined as transcribed sequences from the stop codon to the end of the last exon of each gene. Putative miRNA-binding sites within the 3′ UTR of each gene were identified by means of specialized algorithms (e.g. miRBase (30), miRanda (31), PicTar (32), MicroInspector (33), Diana-MicroT (34) and TargetScanS (35), using the default parameters included in the software, for each of them. The SNPs residing on the miRNA-binding sites were found by an extensive search in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP), using BLAST (http://www.ncbi.nlm.-nih.gov/BLAST/Blast.cgi) and BLAST-SNP algorithms (http://www.

Table I. Characteristics of CRC patients and control subjects

		Cases (n = 697)	Controls (n = 624)
Males/females (%)		57.2/48.2	53.6/46.4
Age	Average (years)	61.1	56.0
	Lower quartile (years)	55	47
	Upper quartile (years)	69	66
Rectal cancer (%)		38.4	
Colon cancer (%)		61.7	
Smoking	Non-smokers (%)	53.6	52.8
	Ex-smokers (%)	32.3	26.4
	Smokers (%)	14.1	20.9
Positive family history of CRC	No (%)	82.0	84.3
	Yes (%)	18.0	15.7
Percentage of strictly vegetarian (%)		10.9	7.8
Alcohol consumption	No (%)	47.3	40.2
	Yes (%)	52.7	59.8
Mean grams of alcohol/day		23.1	22.9
Living place	City (%)	57.2	56.2
	Suburbs (%)	15.5	21.2
	Country (%)	27.3	22.7
Education	Basic (%)	33.5	25.4
	High school (%)	51.6	54.8
	University (%)	14.9	19.8
Mean BMI (Kg/m^2) ± SD		26.7 ± 4.3	26.7 ± 4.5
Distribution of the populations according to BMI	<18.5 Kg/m^2 (%)	1.56	0.73
	18.5–24.9 Kg/m^2 (%)	37.0	37.9
	25–29.9 Kg/m^2 (%)	42.5	41.5
	30–34.9 Kg/m^2 (%)	15.1	15.3
	≥35 Kg/m^2 (%)	3.80	4.61

BMI, body mass index. No statistically significant differences were found for all the covariates, between cases and controls.

ncbi.nlm.nih.gov/SNP/snpblastByChr.html). For all the identified SNPs, miRanda was run again to assess the Gibbs binding free energy (ΔG, expressed in KJ/mol) both for the common and the variant alleles. The difference of the free energies between the two alleles was computed as 'variation of ΔG' (i.e. $\Delta\Delta G$). Because in some genes the same sequence is predicted to bind several miRNAs, and thus, a SNPs in these sequences could impact the binding site of more than one miRNA, we used the sum of all the $|\Delta\Delta G|_s$ for each SNP ($|\Delta\Delta G_{tot}|$) as parameter for predicting the biological impact of the polymorphism. Although each target can bind only one miRNA at a time in each tissue, this approach is based on the basic idea that the more miRNAs are predicted to bind to a given target, the more likely it is that at least one of them truly binds to the target. Thus, if the algorithms predict that only one miRNA binds to a given target, at least in theory, we should consider this target less important than others predicted to bind many different miRNAs. Paralogue miRNAs, when reported to have a different sequence, are considered as different miRNAs. When summing the $\Delta\Delta G$s, we did not account for the exactly duplicated miRNAs, which are, in this sense, considered as unique.

In order to give a priority list of polymorphisms having real impact on miRNA bindings, we ranked the values of $|\Delta\Delta G_{tot}|$ and we chose (arbitrarily) the upper tertile of the distribution ($|\Delta\Delta G_{tot}| \geq 5.27$ KJ/mol) as the significant cutoff. In other words, when the SNP had a $|\Delta\Delta G_{tot}| \geq 5.27$ KJ/mol was considered biologically relevant, whereas for $|\Delta\Delta G_{tot}| < 5.27$ KJ/mol the SNPs were considered biologically neutral and not included in further analyses. Moreover, as second criterion for SNP selection, we excluded the SNPs having the minor allele frequency (MAF) lower than 0.10 (for Caucasians). Thus, only polymorphisms falling in the upper tertile of $|\Delta\Delta G_{tot}|$ with MAF > 0.10 were selected for genotyping. According to Breslow and Day (36), this MAF cutoff gave a statistical power of 85% or greater to detect an odds ratio (OR) of at least 1.5 in our sample set (for the codominant model).

Laboratory techniques

Genotyping was carried out with the 5′ nuclease assay (TaqMan; Applied Biosystems, Foster City, CA). Two TaqMan probes were used, one for each allele. Analysis was performed using the ABI PRISM® 7900HT Sequence Detection System and SDS 2.2 software (Applied Biosystems).

The reaction employed the TaqMan Universal PCR Master Mix 2×, TaqMan primers and probes (20× or 40×), water and 5 ng of DNA. The thermal cycle program used included one cycle at 95°C for 10 min to activate the AmpliTaq

Table II. List of 104 candidate genes evaluated for the presence of polymorphic miRNA target sites

Gene symbol	Gene name	Gene symbol	Gene name
TLR1	Toll-like receptor-1	*FADS1*	Fatty acid desaturase 1
TLR2	Toll-like receptor-2	*FADS2*	Fatty acid desaturase 2
TLR3	Toll-like receptor-3	*FADS3*	Fatty acid desaturase 3
TLR4	Toll-like receptor-4	*HNF4A*	Hepatocyte nuclear factor 4, alpha
TLR5	Toll-like receptor-5	*NR1H2*	Nuclear receptor subfamily 1, group H, member 2
TLR6	Toll-like receptor-6	*NR1H3*	Nuclear receptor subfamily 1, group H, member 3
TLR7	Toll-like receptor-7	*PGDS*	Prostaglandin D2 synthase
TLR8	Toll-like receptor-8	*PLA2G10*	Phospholipase A2, group X
TLR9	Toll-like receptor-9	*PLA2G2A*	Phospholipase A2, group IIA
NOD1	Nucleotide-binding oligomerization domain 1	*PLA2G4A*	Phospholipase A2, group IVA
NOD2	Nucleotide-binding oligomerization domain 2	*PLA2G6*	Phospholipase A2, group VI
NFkB1	Nuclear factor -light polypeptide gene enhancer in B-cells	*PPARD*	Peroxisome proliferator-activated receptor, delta
TCRA	T-cell receptor alpha locus	*PPARG*	Peroxisome proliferator-activated receptor, gamma
TCRB	T-cell receptor beta locus	*PPARGC1A*	Peroxisome proliferator-activated receptor coactivator 1, alpha
CD3Z	CD3Z antigen, zeta polypeptide	*PRKCB1*	Protein kinase C, beta 1
CD3G	CD3G antigen, gamma polypeptide	*PTGDS*	Prostaglandin D2 synthase, brain
CD3E	CD3E antigen, epsilon polypeptide	*PTGER1*	Prostaglandin E receptor 1, EP1
CD3D	CD3D antigen, delta polypeptide	*PTGER2*	Prostaglandin E receptor 2, EP2
CD4	CD4 antigen (p55)	*PTGER3*	Prostaglandin E receptor 3, EP3
CD80	CD80 antigen	*PTGER4*	Prostaglandin E receptor 4, EP4
CD86	CD86 antigen	*PTGES*	Prostaglandin E synthase
CD28	CD28 antigen	*PTGES2*	Prostaglandin E synthase 2
CTLA4	Cytotoxic T-lymphocyte-associated protein 4	*PTGIS*	Prostaglandin I2 synthase
GMCSF	Colony-stimulating factor 2	*PTGS1*	Cyclooxygenase-1
IL1A	Interleukin 1, alpha	*PTGS2*	Cyclooxygenase-2
IL1B	Interleukin 1, beta	*RARA*	Retinoic acid receptor, alpha
IL2	Interleukin 2	*RXRB*	Retinoic X receptor, beta
IL3	Interleukin 3	*SCD*	Steoryl-CoA desaturase
IL4	Interleukin 4	*SREBF1*	Sterol regulatory element-binding transcription factor 1
IL5	Interleukin 5	*TBXAS1*	Thromboxane A synthase 1
IL6	Interleukin 6	*INS*	Insulin
IL7	Interleukin 7	*INSR*	Insulin receptor
IL8	Interleukin 8	*IRS1*	Insulin receptor substrate 1
IL10	Interleukin 10	*IRS2*	Insulin receptor substrate 2
IL12A	Interleukin 12	*PI3K*	Phosphoinositide-3-kinase, gamma pol
IL12B	Interleukin 12B	*IGF1*	Somatomedin C
IL15	Interleukin 15	*IGF2*	Somatomedin A
IL16	Interleukin 16	*ACDC*	Adiponectin
IL18	Interleukin 18	*LEP*	Leptin
IL23	Interleukin 23	*LEPR*	Leptin receptor
IL1R1	Interleukin 1 receptor, type I	*GH*	Growth hormone
IL1R2	Interleukin 1 receptor, type II	*GHR*	Growth hormone receptor
IL2RA	Interleukin 2 receptor, alpha	*IGFBP3*	Insulin-like growth factor-binding protein 3
IL2RB	Interleukin 2 receptor, beta	*IGFBP1*	Insulin-like growth factor-binding protein 1
IL2RG	Interleukin 2 receptor, gamma	*IGFALS*	IGF, acid labile subunit
IL12RB2	Interleukin 12 receptor, beta 2	*IGF1R*	Insulin-like growth factor 1 receptor
IFNG	Interferon gamma		
TGFB1	Transforming growth factor beta		
TNFA	Tumor necrosis factor alpha		
JAK2	Janus kinase 2		
STAT4	Signal transducer/activator of transcription 4		
ALOX12	Arachidonate 12-lipoxygenase		
ALOX12B	Arachidonate 12-lipoxygenase, 12R type		
ALOX15	Arachidonate 15-lipoxygenase		
ALOX15B	Arachidonate 15-lipoxygenase, second type		
ALOX5	Arachidonate 5-lipoxygenase		
ELOVL6	Long-chain fatty acyl elongase		
FABP5	Fatty acid-binding protein 5		

Gold enzyme activation, 40 cycles at 92°C for 15 s to denature the DNA and at 60°C for 1 min for the stage of annealing/extension. The assays ID of probes are C___441625_10 (for *CD86*, rs17281995), C___7537839_10 (for *IL12B*, rs1368439), C___1844363_10 (for *IL16*, rs1131445), C___27465694_10 (for *NOD2*, rs3135500), C___8356120_10 (for *INSR*, rs1051690), C___8685434_10 (for *ALOX15*, rs916055) and C___362670_20 (for *PLA2-G2A*, rs11677). The SNP rs16870224 (*PTGER4*) was genotyped according to an ABI assay-by-design. The sequences of primers and probes for this polymorphism are 5′-AGTGTCTCACTAAAGCATGAAATGTGAA (forward primer), 5′-CCTTGTATTAACAATAAACCTTCTCACAGAGA (reverse

primer), VIC-5′-CATACGATTTAAGGTATTTAA (reporter1) and FAM-5′-CATACGATTTAAAGTATTTAA (reporter2).

Statistical analysis

Each polymorphism was tested to ensure the fitting with Hardy–Weinberg equilibrium with alpha threshold of 0.05. We used a multivariate logistic regression analysis to examine associations between these polymorphisms and CRC risk by estimating the ORs and 95% confidence intervals (CIs), adjusted for sex and age (as linear variable) as covariates. Genotypes have been divided into three groups: major allele homozygous (also defined as +/+),

Table III. Candidate SNPs with $|\Delta\Delta G_{tot}|^a > 5.27$ KJ/mol (upper tertile of the distribution of $|\Delta\Delta G_{tot}|$) and MAF > 0.10 (validated in Caucasians), evaluated in the case–control association study

Gene	dbSNP ID	Variation	Frequency in Caucasians	miRNA	$|\Delta\Delta G|$	$|\Delta\Delta G_{tot}|^a$
CD86	rs17281995[b]	G/C	0.12	miR-337	5.76	20.81
				miR-582	4.02	
				miR-200a*	4.32	
				miR-184	4.51	
				miR-212	2.20	
IL12B	rs1368439[b]	A/C	0.32	miR-513	6.50	15.10
				miR-210	4.20	
				miR-27b	2.20	
				miR-27a	2.20	
NOD2	rs3135500[b]	G/A	0.33	miR-158	6.87	14.69
				miR-215	6.87	
				miR-98	0.64	
				miR-573	0.31	
PLA2G2A	rs11677[b]	C/T	0.11	miR-187	7.90	14.09
				miR-638	2.81	
				miR-154	1.81	
				miR-453	1.37	
				miR-296	0.20	
INSR	rs1051690[b]	G/A	0.14	miR-618	3.84	11.25
				miR-612	7.41	
PTGER4	rs16870224[c]	G/A	0.11	miR-9	4.40	10.30
				miR-30a-3p	3.10	
				miR-30e-3p	2.80	
IL16	rs1131445[b]	T/C	0.31	miR-135a	3.93	9.53
				miR-135b	3.93	
				miR-143	0.66	
				miR-18	0.51	
				miR-18a	0.50	
ALOX15	rs916055[b]	T/C	0.27	miR-588	4.29	7.50
				miR-183	3.21	

SNPs are showed with decreasing $|\Delta\Delta G_{tot}|$ values.
[a]The expected $|\Delta\Delta G_{tot}|$ for each given polymorphism is calculated as the sum of all the $|\Delta\Delta G|$s.
[b]The SNPs were genotyped according to the ABI predesigned SNP genotyping assays.
[c]The SNP was genotyped according to an ABI assay-by-design.

heterozygous (+/−) and homozygous variant (−/−). This analysis had been done using a codominant model: ORs and 95% CI were calculated for each genotype compared with the group +/+. Test for linear trend of ORs were calculated using the categorized variable as quantitative, after assigning a linear score to each ordered category. When associations were detected, also the recessive (homozygotes variant versus heterozygotes + homozygotes common allele) and dominant (homozygotes variant + heterozygotes versus homozygotes common allele) models were evaluated. The statistical P-level of significance was also evaluated after Bonferroni's correction for multiple testing. All statistical analyses were conducted using STATA software version 8.0 (Stata Corp. LP, College Station, TX).

Results

Among the 104 starting candidate genes, 46 did not present any SNP in their 3′ UTRs and 21 genes did not show SNPs in miRNA target sequences. The remaining 37 genes showed 57 SNPs in the predicted miRNA-binding sites. However, among them, only eight SNPs were positive for both the selection criteria ($|\Delta\Delta G_{tot}| \geq 5.27$ KJ/mol and MAF > 0.10) and were further investigated in the case–control study. Table III shows the nucleotide variation, the predicted binding miR-NAs, variation of ΔG ($|\Delta\Delta G|$) for each miRNA and the overall $|\Delta\Delta G_{tot}|$ for the eight selected SNPs. These SNPs were in the 3′ UTR of *CD86* and *NOD2* (receptors involved in the immune response); *IL16* and *IL12B* (cytokines), *ALOX15*, *PLA2G2A* and *PTGER4* (biosynthesis of prostaglandins and thromboxanes) and *INSR* (obesity and insulin resistance).

We tested the genotypes for Hardy–Weinberg equilibrium, using $P = 0.05$ as significance threshold and all SNPs were in equilibrium in controls. Table IV shows the statistical analyses, for the codominant model only. Both the heterozygotes and homozygotes for the variant allele rs17281995 within *CD86* were associated with an increased risk of CRC (OR = 1.33; 95% CI = 1.00–1.76 and OR = 2.93; 95% CI = 1.29–6.67, respectively). Moreover, the homozygotes for the variant allele rs1051690 within *INSR* showed an OR of 1.86 with a 95% CI of 0.99–3.50 ($P = 0.052$). This polymorphism reached the significance level using a recessive model (heterozygotes + homozygotes for common variant versus homozygotes rare variant) with an OR of 1.94 and 95% CI of 1.04–3.64 ($P = 0.03$). After Bonferroni's correction for multiple testing, the new threshold for the statistical significance was 0.0062. Under this corrected P-value, the trend test for *CD86* held statistical significance. It is important to note that the association found for *CD86* in the dominant model, with an OR of 1.44 (95% CI = 1.10–1.88; $P = 0.015$) was also very close to this tight limit of significance. Stratifying for the site of cancer, the increased risk for *CD86* was observed both for the colon and rectum; however, it was more evident in the rectum (OR = 1.5; 95% CI 1.05–2.15; $P = 0.028$, dominant model).

Discussion

In the present study, two polymorphisms (rs17281995 in *CD86* and rs1051690 in *INSR*) were found associated with an increased risk of CRC. Five different miRNAs bind to a target site that contains the same polymorphism within *CD86*: miR-337, miR-582, miR-200a*, miR-184 and miR-212 (Table III). When C substitutes G, miR-337, miR-582 and miR-200a* are predicted to bind less tightly to the *CD86* 3′ UTR. In contrast, miR-184 and miR-212 increase their binding affinity. Among all the SNPs analyzed, this is the only one that impacts the binding site of five different miRNAs at the same time. In other words, the SNP that produces the largest predicted effect is also the most strongly associated with risk in the present case–control association study. Moreover, this finding holds even after the Bonferroni's correction, which is generally considered very stringent (37). Furthermore, it should be considered that part of the selected miRNA target sites could have been the result of erroneous prediction by the algorithms (limited experimental information is available for most of the selected miRNAs, according to the latest update of Tarbase 4.0; http://www.diana.pcbi.upenn.edu/tarbase.html). It should be stressed that miR-582 was shown to be present in normal colon tissues, and its expression appeared to be reduced in CRC following measurements done with quantitative real-time polymerase chain reaction (38). We could speculate that this finding seems in agreement with the fact that the SNP predicted to cause a reduced binding of miR-582 to its target is associated with increased risks of CRC. Taken together, all these facts encourage further investigation, by replicating the results in other populations and by making *in vitro* experiments in the biological activity of these polymorphisms.

CD86, with CD80, is a costimulatory ligand expressed on the surface of the antigen presenting cells (dendritic cells, macrophages and B cells), in the immune system (39). Binding of these molecules to the T-cell costimulatory receptors, CD28 and CTLA-4 is essential for the activation and regulation of T-cell immunity (40). Functionally, the binding of CD28 with CD80 and CD86 ligands delivers a positive signal to T cells that culminates in T-cell proliferation, cytokine production and prevents induction of T-cell tolerance (41). In contrast, the interaction of CTLA-4 with CD80/CD86 results in negative signaling leading to attenuation of T-cell activation. Such processes are essential for the regulation of a T-cell response and maintaining T-cell homeostasis (42). CD80 and CD86 have been shown to modulate also the T_h1/T_h2 T-cell differentiation. While CD80 preferentially favors T_h1-type T-cell differentiation, CD86 augments IL4 production and overall T_h2-type T-cell responses (mostly producing anti-inflammatory cytokines) (43). Normal colonic tissue expresses CD80 and CD86 only sporadically, whereas in the pathogenesis of Crohn's and inflammatory bowel diseases, a sharp increase is observed paralleling the

Table IV. Logistic regression analysis for the selected SNPs in the miRNA target sequences

Gene name	Distance from stop codon	dbSNP ID number	Controls[a]			Cases[a]			+/− versus +/+		P_{ass}	−/− versus +/+		P_{ass}	P_{trend}
			+/+	+/−	−/−	+/+	+/−	−/−	OR[b]	95% CI		OR[b]	95% CI		
CD86	+1257G>C	rs17281995	434	114	8	475	161	24	1.33	(1.00, 1.76)	**0.049**	2.93	(1.29, 6.67)	**0.009**	**0.0043**
NOD2	+1156G>A	rs3135500	209	265	81	243	303	120	0.94	(0.73, 1.22)	0.765	1.22	(0.87, 1.74)	0.065	0.2547
IL16	+1263T>C	rs1131445	251	240	53	308	287	65	0.97	(0.77, 1.24)	0.830	0.99	(0.67, 1.50)	0.998	0.9115
IL12B	+1095T>G	rs1368439	388	164	15	465	188	21	0.96	(0.75, 1.23)	0.726	1.17	(0.59, 2.30)	0.652	0.9762
ALOX15	+61T>C	rs916055	218	287	55	284	311	70	0.83	(0.65, 1.06)	0.130	0.98	(0.65, 1.45)	0.908	0.4067
PLA2G2A	+230C>T	rs11677	440	107	8	541	121	10	0.92	(0.69, 1.23)	0.570	1.02	(0.40, 2.60)	0.973	0.6440
PTGER4	+460G>A	rs16870224	439	116	4	523	130	11	0.94	(0.71, 1.25)	0.669	2.31	(0.73, 7.30)	0.143	0.7820
INSR	+104G>A	rs1051690	366	186	15	441	196	33	0.87	(0.68, 1.12)	0.465	1.86	(0.99, 3.50)	0.052	0.7526

Only the codominant model is shown.
[a]Numbers may not add up to 100% of subjects due to genotyping failure. All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to three additional rounds of genotyping. Data points that were still not filled after this procedure were left blank.
[b]ORs were adjusted for sex and age. Only results for the codominant model are shown. Statistically significant results ($P < 0.05$) are shown in bold.

impaired and/or excessive activation of mucosal macrophages and T lymphocytes (44). It should be considered that even a normal colonic mucosa undergoes a continuous state of subchronic inflammation due to the normal bacterial activity present within its lumen (45). Thus, our results support the hypothesis that the regulation of the immune response and its consequent level of local inflammatory microenvironment could play an important role for the modulation of the individual risk. CD86 is probably one of the key molecules involved in these processes. If we assume the role of miR-582 for CD86, one could speculate that the rare variant of rs17281995 is linked to increased levels of CD86, thereby favoring an inflammatory microenvironment predisposing to CRC.

Although less strongly associated with the risk of CRC, it is worth to note that also the variation G/A within the 3′ UTR of *INSR* has a predicted strong impact in the biology of the miRNA target, as it can be observed by the fifth strongest $|\Delta\Delta G_{tot}|$ (11.25 KJ/mol, see Table III). It is important to stress that this value is due largely by the predicted alteration with miR-612 (7.41 KJ/mol), the second strongest individual $|\Delta\Delta G|$ observed in this study.

The levels of glucose and insulin increase after each meal. Following binding of insulin to the extracellular portion of the insulin receptors, a second messenger system diverges into two separate pathways that regulate distinct biological effects: the phosphoinositide-3-kinase pathway or the mitogen-activated protein kinase pathway (46). In insulin resistance, the ability of insulin to initiate these phosphorylation cascades is diminished (47). Epidemiologic studies indicate that the insulin resistance, elevated fasting plasma insulin, glucose and free fatty acids, glucose intolerance, increased body mass index and visceral adiposity are all associated with adenomas, a precursor lesion for CRC (48). Several mechanisms exist whereby insulin signaling can be modulated at the level of the *INSR*. Generally, these processes involve either the modification of *INSR* contained in the cells at the level of transcription or protein degradation or by modification of the enzymatic activity of individual *INSR* (49). Puig and Tijan (50) have demonstrated that the FOXO1 transcription factor mediates a down-regulation of *INSR* transcription in response to insulin stimulation of insulin receptor signaling (51). Insulin also reduces the effective concentration of cellular insulin receptor by increasing internalization and degradation of the protein as well (51). Activation of this system produces a down-regulation of multiple components of the insulin signaling system (52).

We were aware that there could be a potential bias in this study because controls were selected among people undergoing colonoscopy. However, in terms of diet, socioeconomic status, ethnic and familial origins and all the variables collected with in questionnaires, there is no evidence that the controls are not representative of the general population of the Czech Republic. Moreover, in previous studies on the same series of colonoscopized controls, the frequencies of polymorphic alleles within DNA repair genes were shown to be comparable with those of the general healthy Czech population

(53,54), making it reasonable to think that our controls are representative of the general population.

This study has other strengths: i) an adequate sample size (in relation to the criteria of power we set) and ii) the representative character of the study subjects for the entire country and, basically, for central Europe, a very homogeneous area, with generally typical lifestyle and dietary habits, in particular in the last 50–60 years. It is of interest to note that our results are corroborated also by a recent research showing that some miRNA-binding SNPs exhibit significant different allele frequencies between the human cancer expressed sequence tag libraries and the dbSNP database. More importantly, using human cancer specimens against the dbSNP database for case–control association studies, it has been found that 12 miRNA-binding SNPs indeed display an aberrant allele frequency in human cancers. Hence, it can be concluded that SNPs located in miRNA-binding sites affect miRNA target expression and function and are potentially associated with cancers (55). In addition, recently it has also been shown that polymorphic miRNA-binding sites could affect the individual blood pressure levels, further stressing the importance of the polymorphisms involved in these mechanisms for several human traits (56).

In conclusion, we have analyzed 104 CRC-related genes and selected, with *in silico* methods, eight SNPs predicted to have an impact on miRNA binding. Two of them were associated with CRC risk, one remained statistically significant after Bonferroni's correction. This study provides evidence that these SNPs in miRNA-binding sites may be important to cancer risk and supports future work to validate the results in other well-characterized populations as well as to explore the biological significance of these particular SNPs. Our work is the first showing an association between cancer risk and polymorphic miRNA-binding sites. These results show also that a careful *a priori* selection of candidate SNPs could increase the signal-to-noise ratio in case–control association studies. The proposed approach could ease the identification of functionally relevant polymorphisms involved in carcinogenesis to be further analyzed by appropriate (but time consuming) experiments of molecular biology.

References

1. Potter,J.D. (1999) Colorectal cancer: molecules and populations. *J. Natl Cancer Inst.*, **91**, 916–932.

2. Strate,L.L. *et al.* (2005) Hereditary colorectal cancer syndromes. *Cancer Causes Control*, **16**, 201–213.

3. Kiyohara,C. (2000) Genetic polymorphism of enzymes involved in xenobiotic metabolism and the risk of colorectal cancer. *J. Epidemiol.*, **10**, 349–360.

4. Moreno,V. *et al.* (2006) Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. *Clin. Cancer Res.*, **12**, 2101–2108.

5. Michaud,D.S. *et al.* (2005) Dietary glycemic load, carbohydrate, sugar, and colorectal cancer risk in men and women. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 138–147.

6. Eaden,J.A. *et al.* (2001) The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*, **48**, 526–535.

7. Compagni,A. *et al.* (2000) Recent advances in research on multistage tumorigenesis. *Br. J. Cancer*, **83**, 1–5.

8. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

9. Carthew,R.W. (2006) Gene regulation by microRNAs. *Curr. Opin. Genet. Dev.*, **16**, 203–208.

10. Cullen,B.R. (2004) Transcription and processing of human microRNA precursors. *Mol. Cell*, **16**, 861–865.

11. Lee,Y. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.

12. Yi,R. *et al.* (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.*, **17**, 3011–3016.

13. Chendrimada,T.P. *et al.* (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740–744.

14. Zeng,Y. *et al.* (2002) Both natural and designed microRNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell*, **9**, 1327–1333.

15. Zeng,Y. *et al.* (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl Acad. Sci. USA.*, **100**, 9779–9784.

16. Calin,G.A. *et al.* (2006) MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res.*, **66**, 7390–7394.

17. Calin,G.A. *et al.* (2002) Frequent deletions and down-regulation of microRNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA.*, **99**, 15524–15529.

18. Calin,G.A. *et al.* (2004) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl Acad. Sci. USA.*, **101**, 11755–11760.

19. Takamizawa,J. *et al.* (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.*, **64**, 3753–3756.

20. Eis,P.S. *et al.* (2005) Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc. Natl Acad. Sci. USA.*, **102**, 3627–3632.

21. Metzler,M. *et al.* (2004) High expression of precursor microRNA-155/BIC RNA in children with Burkitt lymphoma. *Genes Chromosomes Cancer*, **39**, 167–169.

22. Iorio,M.V. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.*, **65**, 7065–7070.

23. Johnson,S.M. *et al.* (2005) RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635–647.

24. Murakami,Y. *et al.* (2006) Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. *Oncogene*, **25**, 2537–2545.

25. Chan,J.A. *et al.* (2005) MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res.*, **65**, 6029–6033.

26. Weber,F. *et al.* (2006) A limited set of human microRNA is deregulated in follicular thyroid carcinoma. *J. Clin. Endocrinol. Metab.*, **91**, 3584–3591.

27. Michael,M.Z. *et al.* (2003) Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.*, **1**, 882–891.

28. Parkin,D.M. *et al.* (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.

29. Boyle,P. *et al.* (2000) ABC of colorectal cancer: epidemiology. *BMJ*, **321**, 805–808.

30. Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.

31. John,B. *et al.* (2004) Human microRNA targets. *PLoS Biol.*, **2**, e363.

32. Krek,A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.

33. Rusinov,V. *et al.* (2005) MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res.*, **33**, W696–W700.

34. Kiriakidou,M. *et al.* (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, **18**, 1165–1178.

35. Lewis,B.P. *et al.* (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.

36. Breslow,N.E. *et al.* (1987) Statistical methods in cancer research. Volume II—the design and analysis of cohort studies. *IARC Sci. Publ.*, **82**, 1–406.

37. Rothman,K.J. (1990) No adjustments are needed for multiple comparisons. *Epidemiology*, **1**, 43–46.

38. Cummins,J.M. *et al.* (2006) The colorectal microRNAome. *Proc. Natl Acad. Sci. USA.*, **103**, 3687–3692.

39. Orabona,C. *et al.* (2004) CD28 induces immunostimulatory signals in dendritic cells via CD80 and CD86. *Nat. Immunol.*, **5**, 1134–1142.

40. Bhatia,S. *et al.* (2006) B7-1 and B7-2: similar costimulatory ligands with different biochemical, oligomeric and signaling properties. *Immunol. Lett.*, **104**, 70–75.

41. Linsley,P.S. *et al.* (1991) Binding of the B cell activation antigen B7 to CD28 costimulates T cell proliferation and interleukin 2 mRNA accumulation. *J. Exp. Med.*, **173**, 721–730.

42. Tivol,E.A. *et al.* (1995) Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity*, **3**, 541–547.

43. Suvas,S. *et al.* (2002) Distinct role of CD80 and CD86 in the regulation of the activation of B cell and B cell lymphoma. *J. Biol. Chem.*, **277**, 7766–7775.

44. Vuckovic,S. *et al.* (2001) CD40 and CD86 upregulation with divergent CMRF44 expression on blood dendritic cells in inflammatory bowel diseases. *Am. J. Gastroenterol.*, **96**, 2946–2956.

45. Kelly,D. *et al.* (2005) Bacterial modulation of mucosal innate immunity. *Mol. Immunol.*, **42**, 895–901.

46. Taniguchi,C.M. *et al.* (2006) Critical nodes in signalling pathways: insights into insulin action. *Nat. Rev. Mol. Cell Biol.*, **7**, 85–96.

47. Pirola,L. *et al.* (2004) Modulation of insulin action. *Diabetologia*, **47**, 170–184.

48. Keku,T.O. *et al.* (2005) Insulin resistance, apoptosis, and colorectal adenoma risk. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 2076–2081.

49. Youngren,J.F. (2007) Regulation of insulin receptor function. *Cell Mol. Life Sci.*, **64**, 873–891.

50. Puig,O. *et al.* (2005) Transcriptional feedback control of insulin receptor by dFOXO/FOXO1. *Genes Dev.*, **19**, 2435–2446.

51. Okabayashi,Y. *et al.* (1989) Mechanisms of insulin-induced insulin-receptor downregulation. Decrease of receptor biosynthesis and mRNA levels. *Diabetes*, **38**, 182–187.

52. Rome,S. *et al.* (2004) The ubiquitin-proteasome pathway is a new partner for the control of insulin signaling. *Curr. Opin. Clin. Nutr. Metab. Care*, **7**, 249–254.

53. Pardini,B. *et al.* (2007) DNA repair genetic polymorphisms and risk of colorectal cancer in the Czech Republic. *Mutat. Res.*, **638**, 146–153.

54. Vodicka,P. *et al.* (2007) Association of DNA repair polymorphisms with DNA repair functional outcomes in healthy human subjects. *Carcinogenesis*, **28**, 657–664.

55. Yu,Z. *et al.* (2007) Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res.*, **35**, 4535–4541.

56. Sethupathy,P. *et al.* (2007) Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3′ untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. *Am. J. Hum. Genet.*, **81**, 405–413.