Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk

Alessio Naccarati, Barbara Pardini, Stefano Landi, Debora Landi, Jana Slyskova, Jan Novotny, Miroslav Levy, Veronika Polakova, Ludmila Lipska, and Pavel Vodicka

1Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Prague, Czech Republic; 2Human Genetics Foundation, Torino, Italy; 3Department of Biology, University of Pisa, Italy; 4Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic; 5Department of Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic, and 6Department of Surgery, First Faculty of Medicine, Charles University and Thomayer University Hospital, Prague, Czech Republic

Reduced DNA repair capacity and DNA damage accumulation may lead to cancer development. Regulation of and coordination between genes involved in DNA repair pathways is fundamental for maintaining genome stability, and post-transcriptional regulation by miRNAs (miRNAs) may therefore be of particular relevance. In this context, the presence of single nucleotide polymorphisms (SNPs) within the 3′ untranslated regions of target DNA repair genes could alter the binding with specific miRNAs, modulating gene expression and ultimately affecting cancer susceptibility. In this study, we investigated the role of genetic variations in miRNA-binding sites of nucleotide excision repair (NER) genes in association with colorectal cancer risk. From 28 NER genes, we screened among SNPs residing in their 3′ untranslated regions and simultaneously located in miRNA-binding sites, with an in silico approach. Through the calculation of different binding free energy according to both alleles of identified SNPs, and with global binding free energies median providing a threshold, we selected nine NER gene variants. We tested those SNPs in 1098 colorectal cancer cases and 1469 healthy controls from the Czech Republic. Rs7356 in RAD2 and rs4596 in GTF2H1 were associated with colorectal cancer risk. After stratification for tumor location, the association of both SNPs was significant only for rectal cancer (rs7356: OR 1.52, 95% CI 1.02–2.26, P = 0.04 and rs4596: OR 0.69, 95% CI 0.50–0.94, P = 0.02; results not adjusted for multiple testing). Variation in miRNA target binding sites in the 3′ untranslated region of NER genes may be important for modulating colorectal cancer risk, with a different relevance according to tumor location.

Introduction

Individuals with severe defects in DNA repair are at greatly increased risk of cancer and other diseases and many examples of rare, but highly penetrant, germline mutations in the major DNA repair pathways are known (1). However, the involvement of subtle alterations in different DNA repair pathways in association with the vast cancer typology and individual susceptibility is still not exhaustively investigated. Colorectal cancer (CRC, MIM: 114 500) can be caused by variations of the individual DNA repair capacity (DRC) may significantly modulate the susceptibility to this cancer, especially in the context of gene–environment interactions. In particular, single nucleotide polymorphisms (SNPs) in genes directly involved in the DNA repair mechanisms have attracted a massive and enthusiastic research in order to determine whether different genotypes are associated with CRC risk (2). Surprisingly, variation in DNA repair genes is absent in the list of common CRC predisposition loci from genome-wide association study. Recently, to further investigate this topic, a set of SNPs within 100kb of 157 DNA repair genes from three genome-wide association studies were specifically assessed with CRC risk. The set of 5609 SNPs as a whole was associated with this cancer risk, but no individual SNP showed evidence of association (3). However, the etiological role of common genetic polymorphisms in DNA repair genes in modulating the individual DRC and subsequently the sporadic CRC risk have not been comprehensively studied either in the context of epidemiological studies or phenotypically. In this respect, we have recently performed a case–control study to characterize newly diagnosed sporadic CRC patients for nucleotide excision repair (NER)–DRC and to investigate their possible relations with endogenous DNA damage, genetic polymorphisms and expression levels of genes specific for this pathway. Interestingly, patients had a lower NER–DRC and simultaneously exhibited higher DNA damage than age-matched healthy controls. A reduced NER capacity may enhance the CRC risk due to a diminished protection of intestinal epithelium against genotoxic compounds, present in the lumen or transported by the blood. Expression levels of 6 out of 9 analyzed NER genes also differed among cases and controls, but none of them was directly related to measured DRC. In patients, XPC Ala499Val (rs2228000) modulated expression levels of XPC, XPD and XPD genes, whereas in controls XPC Lys939Gln (rs2228001) was associated with XPA expression level. This study highlights the role of NER in sporadic CRC and additionally shows that individual DRC is a complex marker resulting from different genes in interplay, with a potential role of genetic variation (4).

In general, the maintenance of DNA integrity is a critical and controlled process that requires regulation of gene expression levels and coordination of specific processes, so far studied mainly at the transcriptional and post-transcriptional levels (5). In the last decade, post-transcriptional gene regulation by microRNAs (miRNAs) has emerged as a novel important mechanism of gene expression control. MiRNAs are single-stranded non-coding RNA molecules of ~18–24 nucleotides and their function is to interact with the 3′ untranslated region (3′UTR) of target miRNAs and either induce their degradation/destabilization, or inhibit their translation and consequently silence gene expression (7). Approximately one-third of the protein-coding genes are controlled by miRNAs, thus almost all cellular pathways are directly or indirectly influenced by these molecules (8–10). Additionally, recent studies have shown that miRNAs participate in human carcinogenesis as tumor suppressors or oncogenes (11–13) and aberrant miRNA expression and/or function is frequently observed in many cancers, including CRC (14,15). The sequence complementarity and thermodynamics of the binding between miRNAs and their targets play an essential role in this interaction. Sequence variations, such as SNPs, in the seed region (2–7 nucleotides of the mature sequence) or in a target gene site could alter the miRNA–mRNA interaction and affect the expression of miRNA targets, thus, would probably lead to a corresponding decrease or increase in protein translation (16). A SNP residing within the 3′UTRs of genes involved in pathways such as DNA repair, DNA signaling or apoptosis may abolish/create, weaken/strengthen a miRNA target and could indirectly contribute in
affecting the individual risk to develop cancer (17,18). Polymorphic target sites for miRNAs have been proven to be important in modulating the individual risk of CRC, as shown by our group. Two SNPs (rs17281995 in CD86 and rs1051690 in INS), screened from a list of 129 genes in pathways implicated in CRC, were associated with cancer risk in two independent case–control populations from the Czech Republic and Spain (19,20).

Based on the above considerations, the aim of this study was to investigate the role of polymorphisms residing in miRNA-binding sites within 3′UTR of genes specifically involved in NER pathway in modulating the risk of CRC. From an initial list of NER genes, we have selected SNPs affecting the binding with miRNAs by predicted in silico target sequences for miRNAs within their 3′UTRs. Using a series of stringent a priori hypotheses, the final selection comprised nine SNPs. These SNPs were checked for their association with the risk of CRC in a case–control study on 1098 cases and 1459 controls from the Czech Republic. In this country, CRC constitutes a serious health problem as it has among the highest rate of incidence and mortality worldwide (21).

Material and methods

Study population

Cases were collected among patients with histologically confirmed CRC, recruited between September 2004 and October 2010 from several oncological departments in the Czech Republic: Prague (three), Benesov, Brno, Liberec, Pils, Pribram, Usti nad Labem and Zlin. This study included 1098 subjects who could be interviewed, provided biological samples and who were genotyped appropriately.

Two control groups were included in the study. The first group was selected among individuals admitted to five large gastroenterological departments (Prague, Brno, Jihlava, Liberec and Pribram) in the Czech Republic, at the same time period as the recruitment of cases took place. This group consisted of 688 hospital-based volunteers with negative colonoscopy results for malignancy or idiopathic bowel diseases (CFCC, cancer-free colonoscopy inspected controls). The second group of controls consisted of 72 voluntary blood donors who underwent the colonoscopy test, (ii) hemorrhoids, (iii) abdominal pain of unknown origin and (iv) macroscopic bleeding. Cases and CFCC had the same inclusion and exclusion criteria. Due to the high incidence of CRC in the Czech Republic, colonoscopy is widely recommended and practiced. Subjects with negative colonoscopy results for malignancy or idiopathic bowel diseases were included in the control group. To reduce selection bias, only those subjects with no previous diagnosis of any chronic disease were included into the study. This criterion was used to avoid inclusion of individuals with chronic diseases who might have been repeatedly admitted to the hospital and modified their habits because of the disease. The second group of controls consisted of 781 healthy blood donor volunteers (HBDV) collected from a blood donor center in Prague. All individuals were subjected to standard examinations to verify the health status for blood donation (detailed blood count, urinary examination, blood pressure and general examination). They were cancer-free at the time of the sampling. The sample collection was performed at the same time as that of the other two study groups.

The choice of two different control populations was done for two main reasons. (i) The inclusion of ‘colonoscopy negative’ individuals ensured cancer-free control individuals because a negative colonoscopy is the best available proof of the absence of CRC (34). (ii) Since this group of individuals may not necessarily represent the general population, we included also healthy, disease-free individuals recruited among volunteers from blood centers.

All subjects were informed and provided written consent to participate in the study and to approve the use of their biological samples for genetic analyses, according to the Helsinki declaration. The design of the study was approved by the local Ethics Committees. Cases and controls were personally interviewed by trained personnel using a structured questionnaire to determine biological, lifestyle and demographic characteristics and potential risk factors for CRC, such as body mass index (BMI), diabetes and family/personal history of cancer. Part of cases and controls presented here were also analyzed in previous association studies (22,23).

Selection of candidate genes

The list of 28 NER genes was extracted from the complete list of all DNA repair genes organized by pathways available online (http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html#NER*).

Selection of the SNPs in miRNA target binding sites

For each gene, SNPs in target binding sites for miRNAs were investigated by using freely available software (results from MicroSNiPer, Patrocles and PolymiTRs were compared). Four genes did not present any SNP in miRNA-binding sites. All 131 detected SNPs were tested for minor allele frequency (MAF) (>5% in Caucasian populations) in the SNP database (dbSNP; http://www.ncbi.nlm.nih.gov/SNP/) in order to have an appropriate statistical power. The selection was primarily done on the basis of HAPMAP CEU population. Whenever this was not possible, other populations were used as a reference. From this selection, ~40 SNPs with the required MAF were found in the 3′UTR of 12 NER genes. For each of these SNPs, miRNAs binding to the site were screened by MicroSNiPer (http://cdbib.nimh.nih.gov/microsnipper, [24]). Other bioinformatics tools were tested before adopting MicroSNipers (Patrocles and PolymiTRs). However, this tool resulted among the most informative about number of genes/SNPs and providing an easy way to select miRNA target sites according to the binding to different alleles and a straightforward possibility to test binding free energy. From the complete list of reported miR- NAs (updated at March 2011), those binding to both alleles were selected and compared for binding free energy (see below for details).

Bioinformatics

For the selected SNPs, the algorithm RNAcofold ([http://rna.tbi.univie.ac.at/cgi-bin/RNAcofold.cgi](http://rna.tbi.univie.ac.at/cgi-bin/RNAcofold.cgi)) was run to assess the Gibbs binding free energy (ΔG, expressed in KJ/mol), both for the common and the variant alleles. The genomic RNAcofolding score (RNAcofold score) computes the hybridization energy and base-pairing pattern of two RNA sequences (25). The difference of the free energies between the two alleles was computed as ‘variation of ΔG’ (i.e. |ΔΔG|) (Table I). Since the neighbor sequence of each SNP can be a target for different miRNAs, we calculated the sum of the absolute values of ΔGs for each SNP (i.e. |ΔAG| = Σ |ΔΔG|) (19). The |ΔAG| should be considered as a sort of ‘disturbance index’ predicting the likelihood for a given SNP to affect the function of the 3′UTR and it allows a ranking of SNPs for their relevance, as illustrated in previous studies (19,26). Out of 30 miRNAs, 16 |ΔAG| were calculated corresponding to 16 SNPs. The nine first SNPs that presented |ΔAG|, higher or equal than the median (i.e. ≥ 5.6 KJ/mol) were included in the study.

SNP genotyping

The genomic DNA was isolated from peripheral blood lymphocytes using standard procedures. The DNA samples from cases and controls were randomly placed on plates where an equal number of cases and controls could be run simultaneously. Genotyping of the nine selected SNPs was carried out by using the KASP chemistry of KBioscience (Hoddesdon, Herts, UK; [http://www.kbioscience.co.uk/reagents/KASP_manual.pdf](http://www.kbioscience.co.uk/reagents/KASP_manual.pdf)), which is a competitive allele-specific PCR SNP genotyping system that uses FRET quencher cassette oligos. The reaction employed the KASP 2× Reaction Mix, KASPar primers and probes, water and 10 μg of DNA for 10 μl of reaction and a standard PCR protocol available from KBioscience. Duplicate samples (5%), no template controls in each plate, and Hardy–Weinberg equilibrium test were used as quality control tests. One SNP in XPC (rs2229090) was in strong linkage (D 1, r2 0.995, LOD 19.86 in CEU population) with another XPC polymorphism, rs2228000, which was previously analyzed by us in a subset of the CRC cases and controls (27). The genotyping for rs2228000 was thus extended to the whole current study population.

Statistical analyses

To verify whether the genotypes were in Hardy–Weinberg equilibrium in controls, we used the Chi-square test (1 degree of freedom), with a type-I alpha error of 0.05. The multivariate logistic regression analysis was used to test the association between genotypes and risk of CRC. The covariates included into the model were sex, age, smoking habit (non-smokers, smokers and ex-smokers), BMI, any positive family history of CRC, education level (high, intermediate and low) and living area (country, town neighborhood and town). The association between SNPs and CRC risk was calculated by estimating the odds ratio (OR) and its 95% confidence intervals (CI), adjusted for both confounding factors. Results Table 1 shows the final list of selected SNPs within miRNA-binding sites in the 3′UTR of NER genes and the calculation of their |ΔAG| T. The successful genotyping rate was 98.0% and the quality control of
genotypes was assured (>99% of concordance). The SNP in CETN2, whose MAF was based on genotyping of the EGP_CEPH panel, was found to be monoallelic in both controls and cases in the study population. The distribution of genotypes within all the remaining selected genes in the controls was in agreement with Hardy–Weinberg equilibrium.

Table II illustrates the characteristics of the study population. The definitive number of CRC patients was 1098 among which approximately 2/3 patients were diagnosed with colon cancer and the rest with rectal cancer. Among the 1469 controls, 688 were from CFCC (colonoscopy negative controls) and 781 from HBDV (blood donor volunteers). Compared with subjects of both control groups, CRC cases were more probably to be older, have a slightly higher BMI, and, compared with the HBDV group, they were more probably to have a positive family history of CRC, and to be less educated.

Supplementary Table I, available at Carcinogenesis Online, reports the genotype and allele frequencies for the investigated SNPs and presents the best model for association with CRC risk after adjusting for all listed confounders. The Hardy–Weinberg equilibrium is also reported. Table III summarizes the most significant results. The ORs, adjusted for the covariates, showed that the polymorphism rs4596 within GTF2H1 was associated with CRC risk, when a dominant model is applied. In fact, the carriers of at least one G allele were at a decreased risk of CRC, with a statistically significant OR of 0.79 (95% CI 0.64–0.99, \( P = 0.03 \)).

Moreover, the rs7356 within RPA2 was associated with CRC risk when a recessive model was applied and the individuals carrying the GG genotype were at higher cancer risk, as compared with the group of carriers of the AA+AG genotypes (OR 1.33, 95% CI 1.01–1.75, \( P = 0.04 \)). Interestingly, when the analyses were carried out by selecting specific subgroups, the stratification of the cases by tumor location showed that the associations between the risk of CRC and the polymorphisms within GTF2H1 and RPA2 were statistically significant only in rectal cancer. Table III shows the best models for both rs7356 and rs4596 (\( P = 0.04 \) and \( P = 0.02 \), respectively).

Discussion

MiRNAs, involved in a wide range of biological processes and diseases, have recently disclosed also their potentialities to indirectly study the impact of human genetic variation on cancer risk, through the
modulation of the expression of relevant genes in different pathways (16). In this study, we have selected nine SNPs within the 3′UTRs of 7 NER genes (CETN2, ERCC3, ERCC4, GTF2H1, RPA1, RPA2 and XPC) from the whole pathway, potentially altering the binding with specific miRNAs and we have tested their association with CRC risk in a case–control study. The rationale for selecting the genes/poly-morphisms has been provided by MicroSNiPer, a web-based tool that predicts the impact of a SNP on putative miRNA targets. MicroSNiPer computes these sites and examines the effects of SNPs, simultaneously giving information as to which miRNA binds on each SNP and showing the sequence complementarity for each SNP allele (24). On the basis of these results, we have investigated the Gibbs binding free energy by means of RNAcofold software and created a priority list of SNPs of interest to be tested in the present association study based on a population of CRC cases and controls from the Czech Republic. As a result, two SNPs (rs4596 C/G and rs7356 A/G) have provided a significant association with CRC risk. The G allele in rs4596 in GTF2H1 was associated with a decreased cancer risk. The GTF2H1 gene (general transcription factor IIH, polypeptide 1, 62kDa) maps at 11p15.1–p14 and encodes for a component of the core-TFIIH basal transcription factor involved in NER and, when in complex with CAK, in RNA transcription by RNA polymerase II. To the best of our knowledge, no previous study has highlighted a role of this SNP in CRC risk. However, a recent study on colorectal adenomas found that a SNP (rs4150628) in this gene (and in strong linkage disequilibrium with the one investigated by us) was associated with a moderate adenoma risk (28). From our analysis, rs4596 is located in a haplotype block that constitutes a cluster transcribed in c-myc, a very large (>300 kb) locus located downstream of c-myc on chromosome 8q24. The mature

---

### Table II. Characteristics of the study population

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CRC cases</th>
<th>Control group I (colonoscopy negative)</th>
<th>P-value</th>
<th>Control group II (blood center)</th>
<th>P-value</th>
<th>All controls</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon cancer (%)</td>
<td>725 (66.0)</td>
<td>—</td>
<td>—</td>
<td></td>
<td>—</td>
<td>781</td>
<td>—</td>
</tr>
<tr>
<td>Rectal cancer (%)</td>
<td>370 (33.7)</td>
<td>—</td>
<td>—</td>
<td></td>
<td>—</td>
<td>1469</td>
<td>—</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>61.7 (10.8)</td>
<td>56.1 (13.4)</td>
<td>&lt;10^−3</td>
<td>45.6 (8.3)</td>
<td>&lt;10^−3</td>
<td>50.5 (12.2)</td>
<td>&lt;10^−3</td>
</tr>
<tr>
<td>Median</td>
<td>62</td>
<td>57</td>
<td>—</td>
<td>46</td>
<td>—</td>
<td>51</td>
<td>—</td>
</tr>
<tr>
<td>Range</td>
<td>26–89</td>
<td>24–91</td>
<td>18–63</td>
<td>18–91</td>
<td>18–91</td>
<td>18–91</td>
<td>18–91</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males (%)</td>
<td>662 (60.3)</td>
<td>371 (53.9)</td>
<td>0.007</td>
<td>438 (56.1)</td>
<td>0.062</td>
<td>809 (55.1)</td>
<td>0.007</td>
</tr>
<tr>
<td>Females (%)</td>
<td>436 (39.7)</td>
<td>317 (46.1)</td>
<td>—</td>
<td>343 (43.9)</td>
<td>—</td>
<td>660 (44.9)</td>
<td>—</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>26.8 (4.2)</td>
<td>26.8 (4.54)</td>
<td>0.93</td>
<td>26.1 (3.8)</td>
<td>&lt;10^−3</td>
<td>26.4 (4.2)</td>
<td>0.026</td>
</tr>
<tr>
<td>BMI ≤ 26.2</td>
<td>386</td>
<td>306</td>
<td>—</td>
<td>445</td>
<td>—</td>
<td>751</td>
<td>—</td>
</tr>
<tr>
<td>BMI ≥ 26.2</td>
<td>439</td>
<td>306</td>
<td>—</td>
<td>321</td>
<td>—</td>
<td>627</td>
<td>—</td>
</tr>
<tr>
<td>Family history of CRC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (%)</td>
<td>144 (16.5)</td>
<td>90 (15.6)</td>
<td>0.65</td>
<td>52 (6.8)</td>
<td>&lt;10^−3</td>
<td>142 (10.5)</td>
<td>&lt;10^−3</td>
</tr>
<tr>
<td>No (%)</td>
<td>727 (83.5)</td>
<td>486 (84.4)</td>
<td>—</td>
<td>718 (93.2)</td>
<td>—</td>
<td>1204 (89.5)</td>
<td>—</td>
</tr>
<tr>
<td>Smoking habit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (%)</td>
<td>501 (48.3)</td>
<td>364 (58.9)</td>
<td>0.005</td>
<td>452 (58.1)</td>
<td>0.007</td>
<td>816 (58.5)</td>
<td>&lt;10^−3</td>
</tr>
<tr>
<td>No (%)</td>
<td>537 (51.7)</td>
<td>254 (41.1)</td>
<td>—</td>
<td>326 (41.9)</td>
<td>—</td>
<td>580 (41.5)</td>
<td>—</td>
</tr>
<tr>
<td>Living area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>511</td>
<td>338</td>
<td>0.051</td>
<td>614</td>
<td>&lt;10^−3</td>
<td>952</td>
<td>&lt;10^−3</td>
</tr>
<tr>
<td>Suburbs</td>
<td>128</td>
<td>118</td>
<td>0.99</td>
<td>53</td>
<td>112</td>
<td>171</td>
<td>171</td>
</tr>
<tr>
<td>Countryside</td>
<td>243</td>
<td>157</td>
<td>—</td>
<td>112</td>
<td>—</td>
<td>269</td>
<td>—</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>266</td>
<td>171</td>
<td>0.090</td>
<td>53</td>
<td>&lt;10^−3</td>
<td>224</td>
<td>&lt;10^−3</td>
</tr>
<tr>
<td>Intermediate</td>
<td>470</td>
<td>327</td>
<td>—</td>
<td>492</td>
<td>—</td>
<td>819</td>
<td>—</td>
</tr>
<tr>
<td>High</td>
<td>138</td>
<td>114</td>
<td>—</td>
<td>231</td>
<td>—</td>
<td>345</td>
<td>—</td>
</tr>
</tbody>
</table>

*Ex-smokers are included in this group.*

### Table III. Association of two SNPs in selected NER genes with CRC risk. A further stratification for rectal cancer only is also presented.

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP ID</th>
<th>Model</th>
<th>Genotype</th>
<th>All controls* n (%)</th>
<th>All cases* n (%)</th>
<th>OR(95% CI)</th>
<th>P-value</th>
<th>Cases with rectal cancer* n (%)</th>
<th>OR(95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPA2</td>
<td>rs7356</td>
<td>Rec</td>
<td>AA+GA</td>
<td>1244 (85.9)</td>
<td>908 (83.8)</td>
<td>Ref</td>
<td></td>
<td>305 (83.3)</td>
<td>Ref</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>205 (14.1)</td>
<td>175 (16.2)</td>
<td>1.33 (1.01-1.75)</td>
<td>0.04</td>
<td>61 (16.7)</td>
<td>1.52 (1.05-2.26)</td>
<td>0.04</td>
</tr>
<tr>
<td>GTF2H1</td>
<td>rs4596</td>
<td>Dom</td>
<td>CC</td>
<td>396 (27.5)</td>
<td>321 (29.6)</td>
<td>Ref</td>
<td></td>
<td>120 (32.7)</td>
<td>Ref</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG +GG</td>
<td>1043 (72.5)</td>
<td>764 (70.4)</td>
<td>0.79 (0.64-0.99)</td>
<td>0.03</td>
<td>247 (67.3)</td>
<td>0.69 (0.50-0.94)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

For each SNP the best model is presented (multivariate logistic regression analysis).

OR: odds ratio; CI: 95% confidence interval. OR adjusted for sex, age, smoking habit, BMI, familial history of CRC, education level and living area. Significant results in bold.

*All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to two additional rounds of genotyping. Data points that were still not filled after this procedure had been left blank.*
forms of these miRNAs have been found to be expressed, though differently, in several cancer cell lines, including Burkitt lymphoma and breast and colon cancer cell lines (29,30). Despite this interesting findings, miR-1205 does not seem to have a stronger impact on energy binding all specificity. On the other hand, miR-518a and miR-527 were recently found to predict the progesterone receptor status of breast cancer patients, providing insight into the regulation of breast cancer phenotypes and progression (31).

Variant A allele of rs7356 in RPA2 (replication protein A 32kDa subunit) was associated with increased rectal cancer risk. From our analysis, miR-3149 binding to this site is not known to be involved in CRC or other diseases, whereas mir-1183 seems to be a novel candidate for sarcoma (32) and a relevant element for differentiation of astrocytes (33). Both identified miRNAs are more prone to bind to variant G allele, because of the less energy binding needed, eventually resulting in a stronger negative regulation on target gene expression. With an observed increased risk of cancer for the G allele carriers, this finding supports once again our initial hypothesis of different allele specificity on miRNA target binding sites—different effect on miRNA regulation, which may impact DNA repair activity and ultimately cancer susceptibility. Interestingly, RPA2 gene maps at 1q35 and encodes for a protein that is a part of three subunit RPA protein required for DNA recombination, repair and replication. RPA binds to single-stranded DNA and interacts with other proteins such as XPA. Givalos et al. (34) have suggested that these proteins are implicated in colon cancer growth, due to the widespread expression of RPA1 and RPA2 proteins in colon carcinomas. Additionally, the association of these multifunctional molecules with advanced stage, lymph node metastasis as well as their increased expression in metastatic sites supports their role in cancer progression.

In this study, the association of the two SNPs was stronger for rectal cancer risk. Previous studies have observed different tissue specificity of miRNAs (35) and Slattery et al. (15) have reported that over 200 miRNAs are differentially expressed in healthy colon and rectal tissue. These differences in miRNA expression levels support the hypothesis that the two diseases arising in different tissues may be represented by distinct markers and those associations may be masked when studying them as one disease. Despite the smaller number of cases affected by rectal cancer in our study, this fact confirms a different genetic/environmental and also epigenetic situation, according to tumor location for CRC (36).

It should be stressed that this is the first report on associations between SNPs in miRNA-binding site specifically of NER genes. The interest to focus on genes of this repair pathway in this study was derived from previous findings on significantly altered NER-DRC in newly diagnosed CRC patients as compared with healthy controls (5). The role of miRNAs in the DNA repair is a relatively new field and may help to understand the complex interplay between genes, their expression in the DNA damage response (together with cell cycle regulation and apoptosis processes) and in the individual susceptibility to cancers, including CRC (37). Several miRNAs involved in DNA repair processes have been identified (38). MiR-210 and miR-373 regulated the expression of RAD52 and RAD23B, key proteins involved in the homology-dependent repair and in NER (39). Recently, the inhibition of NER in HepG2.2.15 cells was found in concomitance with the upregulation of miR-192 that targets ERCC3 and ERCC4 (40).

Despite the great interest and expectations on the role of DNA repair variation in cancer susceptibility, the majority of studies did not provide convincing evidence of associations between SNPs and CRC risk (3). Besides, also genome-wide association study on cancer risks have failed to find interesting associations (4), with minor exception for rs999737 near RAD51L1 and breast cancer (41). This has several potential causes as follows: (i) a strong negative selection on small changes in DNA repair and genome integrity because of their extreme importance for the cell and organism; (ii) conversely, in a multistep and multigenic process such as carcinogenesis, single polymorphisms in single genes may not alter the expression or function of specific proteins to the extent of producing a pathological phenotype (42); (iii) gene–gene interactions as well as gene interactions with environmental factors may explain different relevance of variants in cancer susceptibility (43). In this sense, pathway approaches, network analyses and complex statistical methodologies may allow a more comprehensive studies of multiple SNPs in multiple genes (44).

We are aware that some limitations could hamper our findings. In particular, we did not apply any correction for multiple testing analyses. It should be considered that the SNPs were selected for their high prior probability of functional significance, based on differential binding of miRNAs to their predicted polymorphic target sites. Additionally, in this study, with 1098 cases and 1469 controls, we had greater than 80% power to detect an association for a recessive/dominant model with OR > 1.30 at α = 0.05 (with MAF = 0.30).

Our results encourage further investigations on the role of polymorphisms within miRNA-binding sites and miRNA-dependent gene regulation as a possible functional significance of variation in humans. This study and a very recent one (45) provide evidences of this modulation in DNA repair pathways in association with altered cancer risk. Functional analyses are warranted to demonstrate the biological effect of the SNPs on miRNA regulation and in vitro assays may become more routinely employed in order to evaluate the importance of SNPs within miRNA target sites for the individual risk of cancer.

Supplementary material
Supplementary Table I can be found at http://carcin.oxfordjournals.org/.

Funding
Grant Agency of the Czech Republic (CZ GA ČR: GAP304/10/1286, CZ GA ČR: GA305/09/P194) and the AIRC (Associazione Italiana Ricerca Cancro) Investigator Grant 2008.

Acknowledgements
We acknowledge the study participants for their time and for donating biological material that makes this study possible.

Conflict of Interest Statement: None declared.

References

Received January 12 2012; revised April 27 2012; accepted May 8 2012