Variation within 3’-UTRs of Base Excision Repair Genes and Response to Therapy in Colorectal Cancer Patients: A Potential Modulation of microRNAs Binding

Barbara Pardini1, Fabio Rosa1, Elisa Barone3, Cornelia Di Gaetano1,2, Jana Slyskova5,8, Jan Novotny9, Miroslav Levy6, Sonia Garritano4, Ludmila Vodickova5,8, Tomas Buchler7, Federica Gemignani3, Stefano Landi3, Pavel Vodicka5,8, and Alessio Naccarati1,5

Abstract

Purpose: Colorectal cancer is routinely treated with a 5-fluorouracil (5-FU)-based chemotherapy. 5-FU incorporates into DNA, and the base excision repair (BER) pathway specifically recognizes such damage. We investigated the association of single-nucleotide polymorphisms (SNP) in the 3’-untranslated regions (UTR) of BER genes, and potentially affecting the microRNA (miRNA) binding, on the risk of colorectal cancer, its progression, and prognosis. SNPs in miRNA-binding sites may modulate the posttranscriptional regulation of gene expression operated by miRNAs and explain interindividual variability in BER capacity and response to 5-FU.

Experimental Design: We tested 12 SNPs in the 3’-UTRs of five BER genes for colorectal cancer susceptibility in a case–control study (1,098 cases and 1,459 healthy controls). Subsequently, we analyzed the role of these SNPs on clinical outcomes of patients (866 in the Training set and 232 in the Replication set).

Results: SNPs in the SMUG1 and NEIL2 genes were associated with overall survival. In particular, SMUG1 rs2233921 TT carriers showed increased survival compared with those with GT/GG genotypes [HR, 0.54; 95% confidence interval (CI), 0.36–0.81; P = 0.003] in the Training set and after pooling results from the Replication set. The association was more significant following stratification for 5-FU–based chemotherapy (P = 5.6 × 10–5). A reduced expression of the reporter gene for the T allele of rs2233921 was observed when compared with the common G allele by in vitro assay. None of the genotyped BER polymorphisms were associated with colorectal cancer risk.

Conclusions: We provide the first evidence that variations in miRNA-binding sites in BER genes 3’-UTR may modulate colorectal cancer prognosis and therapy response. Clin Cancer Res; 19(21); 6044–56. ©2013 AACR.

Introduction

Colorectal cancer is the second most common cancer in Western countries (1). Although it is a preventable and potentially curable malignancy, it still represents the second leading cause of cancer-related death among men and women (2).

Impaired DNA repair capacity and DNA damage accumulation may lead to cancer development. In particular, reduced DNA repair activity due to differential genetic background in DNA repair genes may modulate individual cancer risk (3). Base excision repair (BER) is the main pathway involved in resolving spontaneous, alkylating, and oxidative DNA lesions (4). Defects in BER have been linked to several diseases and the disruption of BER coordination or efficacy can potentially foster tumorigenesis. Subtle variations in BER genes may be related to disease proneness or susceptibility, and this is particularly important when considering relevant exposures or during the cumulative life span of an individual (5). In our previous study, subjects homozygous for the variant alleles of two exon-localized polymorphisms of BER genes, APE1 and hOGG1, were at an increased risk of colorectal cancer (6). Such variant alleles were also associated with significantly lower BER rates in healthy subjects, suggesting the importance of DNA repair in cancer susceptibility, as well as gene–gene and gene–environment interactions (7).
SNPs in microRNA-Binding Sites and Colorectal Cancer Clinical Outcome

Translational Relevance

Base excision repair (BER) is the main pathway involved in resolving DNA lesions, including misincorporated uracil (or its analogs), as a result of chemotherapy with 5-fluorouracil (5-FU). A fine modulation of BER genes expression, such as the posttranscriptional regulation exerted by microRNAs (miRNA), could affect the efficiency of this repair system. The presence of single-nucleotide polymorphisms (SNP) within 3’-untranslated regions (UTR) of target genes could alter the binding with specific miRNAs, modulating gene expression and ultimately affecting cancer susceptibility, prognosis, and therapy outcomes. This study comprehensively evaluated the associations between SNPs in miRNA target-binding sites of BER genes and clinical outcomes in patients with colorectal cancer. Variations in the SMUG1 gene provided biologically plausible predictors of survival, with an additional effect of 5-FU chemotherapy. The validation of those SNPs in interaction with the clinical variables may allow the identification of patients with colorectal cancer with a specific prognosis, which can aid the design of personalized cancer therapy.

BER also removes uracil (or its analogs) misincorporated into the DNA as a result of chemotherapy with 5-fluorouracil (5-FU; ref. 8). 5-FU is widely used in the treatment of a variety of solid tumors, including colorectal cancer (9). Intracellular metabolites of 5-FU inhibit thymidylate synthase exerting a cytotoxic effect or they may incorporate into RNA and DNA and ultimately activate apoptosis. In this sense, DNA repair pathways, particularly BER, deserve more attention as they represent important factors modulating the cell response to 5-FU and its resistance (8).

Notably, a fine regulation of BER genes, such as posttranscriptional and posttranslational regulation could affect the efficiency of this repair system. In the last decade, there has been increasing interest in the role of posttranscriptional regulation of gene expression by microRNAs (miRNA) and their possible effect on cancer risk and clinical outcomes. miRNAs are a class of short noncoding RNAs of approximately 20 to 25 nucleotides that modulate gene expression through complementary base pairing between the seed sequence and its complementary seed match sequence, mainly located in the 3′-untranslated region (UTR) of target mRNAs (10). 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 (11, 12). Aberrant miRNA expression and/or function is frequently observed in many malignancies, including colorectal cancer (13). Furthermore, sequence variations, such as single-nucleotide polymorphisms (SNP), in the seed region or in a target gene site could alter the expression of miRNA targets with consequences on protein translation (14). When SNPs occur in 3’-UTRs, they may interfere with mRNA stability and translation by altering polyadenylation, protein::mRNA and miRNA::mRNA regulatory interactions, which may lead to altered cancer susceptibility (15–17). Polymorphic target sites for miRNAs may play a role in modulating important pathways such as DNA repair (18), thereby potentially affecting the individual risk of colorectal cancer (19, 20).

The aim of the present study was to investigate the role of polymorphisms in potential miRNA-binding sites within the 3′-UTRs of BER genes in modulating the risk of colorectal cancer, its progression, and prognosis. From all identified polymorphisms within the 3′-UTRs, using a series of a priori hypotheses, 12 SNPs in five BER genes were selected and analyzed in a case–control study on 1,098 colorectal cancer cases and 1,459 controls from the Czech Republic. In this country, colorectal cancer constitutes a serious health problem as is among the countries with the highest rate of incidence and mortality worldwide (21). The same set of BER SNPs were additionally investigated for their relationship to the patients’ prognosis and for the effect of 5-FU–based therapy on 1,098 patients (866 in the Training set and 232 in the Replication set) with a detailed follow-up. In addition, the effect of an identified relevant SNP on the regulation of gene expression was further tested by an in vitro functional assay.

Materials and Methods

Study population and data collection

Blood samples from 1,098 subjects were collected among patients with histologically confirmed colorectal cancer, recruited between September 2003 and October 2010 from several oncological departments in the Czech Republic [Prague (three), Benesov, Brno, Liberec, Pils, Pribram, Usti nad Labem, and Zlin]. Two control groups, whose samples were collected at the same time of cases recruitment, were included in the study. The first group consisted of 688 hospital-based individuals admitted to five of the above-mentioned gastroenterologic departments that had negative colonoscopy results for malignancy or idiopathic bowel diseases (CFCC, cancer-free colonoscopy inspected controls). 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. 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 and were cancer-free at the time of the sampling. A detailed description of colorectal cancer cases and controls is reported elsewhere (22, 23).

All subjects were informed and provided written consent to participate in the study and to approve the use of their biologic samples for genetic analyses, according to the Helsinki declaration. The design of the study was approved by the local Ethics Committee. Study subjects provided information on their lifestyle habits, body mass index (BMI), diabetes, and family/personal history of cancer,
using a structured questionnaire to determine demographic characteristics and potential risk factors for colorectal cancer.

Follow-up of the patients

Eight hundred sixty-six of the colorectal cancer cases were monitored with follow-ups until August 31, 2011. They represented the initial Training set recruited in nine oncological departments in Prague and throughout the Czech Republic. For these subjects, clinical data at the time of diagnosis, including location of the tumor, UICC (Union for International Cancer Control) tumor–node–metastasis (TNM) stage system, grade, and adjuvant chemotherapy treatment were collected, as well as information about distant metastasis, relapse, and date of death. Because 223 patients had incomplete clinical information, they were excluded from the analyses. Three hundred and nineteen colorectal cancer cases received a 5-FU–based adjuvant regimen as first-line postoperative therapy. The therapy consisted of either a Mayo regimen, delivered as a bolus infusion of 5-FU (425 mg/m²) and leucovorin (10 mg/m²) for 5 days every 4 weeks six times or a simplified DeGra-

Selection of the SNPs in miRNA target–binding sites

For each BER gene, SNPs within target-binding sites for miRNAs were identified by using the freely available software MicroSNiper [http://cbdb.nimh.nih.gov/microsniper (24); updated at March 2012]. The 50 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/) to reach an appropriate represen-

SNP genotyping

Genomic DNA was isolated from peripheral blood lymphocytes using standard procedures. The DNA from cases and controls was randomly placed on plates where an equal number of samples could be run simultaneously. The 12 selected SNPs were genotyped using the KASP genotyping assay, a competitive allele-specific PCR SNP genotyping system (LGc Genomics). For quality control purposes, duplicate samples (5% of the total numbers of samples) were repeated for each SNP, non-template controls were included in each plate, and Hardy–Weinberg equilibrium test was performed.

DNA cloning and in vitro assay

A dual-luciferase reporter assay was used to investigate whether for SMUG1 rs2233921 alleles were associated with differential gene expression. Initially, a 640 bp fragment of the 3′-UTR region of SMUG1 containing the G-allele of rs2233921 was PCR-amplified. The PCR primers were both 41 bp long. The twenty bases at the primers’ 3′ ends were specific to the region to be amplified, whereas the 15 bases at the 5′ ends were homologous to either side of the Xhol restriction site within the multiple cloning site of the pmirGLO vector (Promega). Between the two sequences, each primer was also designed to include a Xhol restriction site sequence (C^TCGAG). The complete sequences were: sense primer, AAGCGACCTGCTAGCCCTGAGGTTGCCCCGTGGGGCCCTTCAT; and antisense primer, CAGGTGCCCTTGGGGCCTTGAGTACAGATGAGGAGCCTGAGG. The vector was linearized with Xhol (NEB Inc.) and the PCR product was cloned downstream from the firefly luciferase (Photinus pyralis) reporter gene, using the Clone EZ PCR Cloning Kit (GenScript). As suggested by the manufacturers, competent cells DH5α (NZYTech) were used for transformation after the cloning reaction. To obtain a vector with a SMUG1 3′-UTR bearing the T-allele of rs2233921, the construct underwent site-specific mutagenesis using the QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent). The sequences of the mutagenic primers were: sense, TGAACACACATCCTCGTGACTGAGCAGCAAAGTTTC; and antisense, GGACCTTTTGCTCCAGTGAGATGAGAGGTCTTGAG. Following the digestion of the parental (methylated) supercoiled dsDNA with Dpn I, XL10-Gold ultra-competent cells (Agilent) were used for transformation.

For the functional assay, HCT-116 cells were plated at a density of approximately 5 × 10⁴ cells per well in 24-well plates and incubated overnight at 5% CO₂, 37°C in a humidified incubator. Cells were transiently transfected at
about 80% confluence using 3 μL of Polyfect transfection reagent (Qiagen), and 0.4 μg of the chimeric construct carrying the T or the G allele. The pmirGLO vector contains the luciferase gene from *Renilla reniformis* (hRLuc-neo), acting as a control reporter to normalize transfection efficiency. Three independent experiments were carried out. Six replicates of each experimental point were performed in each experiment. Twenty-four hours after transfection, cells were washed with a PBS solution and lysed with 100 μL of Passive Lysis Buffer (PLB; Promega) for optimal stability of the firefly and *Renilla* luciferase reporter enzymes. The culture vessel was shaken for 15 minutes at room temperature. The lysates were then transferred to 1.5-mL microcentrifuge tubes and centrifuged at 13,000 rpm for 1 minute to pellet the debris. Supernatants were transferred to clean tubes and used for measuring the activity of firefly and *Renilla* luciferases. The assays were carried out using the Dual-Luciferase Reporter Assay Kit (Promega). Because the same vector carries both the luciferase and the *Renilla* genes, the experimental variability was greatly reduced allowing a more precise evaluation of small biologic differences such as those reflecting the effect of the different tested alleles.

For each transfection, luminescence intensity was evaluated by a luminometer (Optima Fluostar, BMG), and luciferase activities were averaged from four measurements. The luminescence intensities of firefly and *Renilla* luciferase of the nontransfected cells (background) were subtracted from the values obtained for the transfected cells with the pmirRGLO vector containing the 3'-UTR. The luminescence of the *Renilla* luciferase was used as the reference value to calculate the value of firefly luciferase (Luc/Ren ratio of luminescence).

**Statistical analyses**

Chi-square test (1 degree of freedom), with a type-I error threshold set at \( \alpha = 0.05 \), was used to verify whether the genotypes were in Hardy–Weinberg equilibrium in controls. The multivariate logistic regression (MLR) analysis was used to test the association between genotypes and risk of colorectal cancer. The covariates analyzed for inclusion in the multivariate model were sex, age, smoking habit (non-smokers vs. smokers and ex-smokers), BMI, familial history of colorectal cancer, education level (high, intermediate, and low), and living area (country, suburbs, and town). The associations between SNPs and colorectal cancer risk were calculated by estimating the ORs and their 95% confidence intervals (CI), adjusted for both continuous and discrete covariates. For all the genotypes, regression coefficients for additive model were estimated. For all SNPs with significant \( P \) values per genotype, the dominant or recessive models were also calculated. The model with the highest likelihood was additionally checked for the significance of possible interaction terms in the MLR analysis. Statistical analyses were performed using R (http://www.rproject.org).

OS in patients with colorectal cancer was evaluated using the date of death or the date of follow-up termination as the endpoint. For the EFS, in patients who did not have distant metastasis at the time of diagnosis, date of relapse, death, or...
end of the study was used as the endpoint of follow-up. EFS was defined as the time from surgery/endpoint of therapy to the occurrence of distant metastasis, recurrence, or death, whichever came first. The survival curves for OS and EFS were derived by the Kaplan–Meier method (R version 2.14-2, Survival package). The relative risk of death was estimated as HR using Cox regression (R version 2.14-2, Survival package). The genotype calls had 99% concordance with duplicate analyses. Multivariate survival analyses were adjusted for age, gender, TNM, and chemotherapy.

For the in vitro assays, the ratios (Luc/Ren) of the measurements of luminescence, each subtracted of its respective background, were compared between genotypes using the multifactor ANOVA with interactions (MANOVA), where "experiment" and "genotype" were entered as independent factors in the model. The statistical tests were two-tailed and carried out using Statgraphics Centurion software (StatPoint Technologies).

Results

SNP selection and prediction of their effects on target sites
Out of 16 genes involved in the BER pathway, two did not present any SNPs within their 3′-UTRs (MUYTH and MPC), whereas 50 SNPs were identified in the remaining 14 genes. A SNP in OGG1 (rs7373786) was not located within a known miRNA-binding site. All 49 remaining SNPs were tested for MAF (>5% in Caucasian populations) in dbSNP. Thirty-three SNPs did not reach the required MAF, and therefore were excluded from the study. An additional four SNPs were found to be in the same haplotype block (D′ = 1; r² = 1) of other identified SNPs (NEIL2 rs804292 with rs1534862; NEIL3 rs1055677 and rs34652756 with rs1055678; and rs2740438 with rs2740439). Therefore, only one SNP for each group was considered (rs1534862, rs1055678, and rs2740439, respectively). The final set of 12 selected SNPs within putative miRNA-binding sites in the 3′-UTR of five BER genes is reported in Table 1.

Case–control study
The total number of patients with colorectal cancer included in the study was 1,095, among which approximately two third were diagnosed with colon and the rest with rectal cancer (3 of them were excluded because of missing tumor site information). Out of the 1,469 controls, 688 were CFCC and 781 were HBDV. Compared with subjects of both control groups, colorectal cancer cases were more likely to be older and have a slightly higher BMI. Compared with the HBDV group, they were more likely to have a positive family history of colorectal cancer and lower formal education (Table 2).

The genotype and allele frequencies for the investigated SNPs and their association with colorectal cancer risk after adjustment for all listed confounders are reported in Supplementary Table S1, together with the calculation of the Hardy–Weinberg equilibrium in controls. For all analyzed SNPs, the genotype calls had 99% concordance with duplicate quality controls. Only one SNP, rs2307294, was not in Hardy–Weinberg equilibrium in the control group (χ² = 9.01 and P = 0.01), therefore it was excluded from further analyses.

None of the polymorphisms genotyped in this study were significantly associated with colorectal cancer risk. To overcome issues due to the age effect (mean age in cases = 61.7 years; mean age in controls = 50.6 years; t test = 24.4; p value = 0.0001) we used age as a covariate in multivariate analysis. As an example, Table 2 shows the impact of rs2307294 on colorectal cancer risk after adjustment for age, gender, TNM, chemotherapy, and a positive family history of colorectal cancer. The relative risk of death was estimated as HR using Cox regression (R version 2.14-2, Survival package). The genotype and allele frequencies for the investigated SNPs and their association with colorectal cancer risk after adjustment for all listed confounders are reported in Supplementary Table S1, together with the calculation of the Hardy–Weinberg equilibrium in controls. For all analyzed SNPs, the genotype calls had 99% concordance with duplicate quality controls. Only one SNP, rs2307294, was not in Hardy–Weinberg equilibrium in the control group (χ² = 9.01 and P = 0.01), therefore it was excluded from further analyses.

None of the polymorphisms genotyped in this study were significantly associated with colorectal cancer risk. To overcome issues due to the age effect (mean age in cases = 61.7 years; mean age in controls = 50.6 years; t test = 24.4; p value = 0.0001) we used age as a covariate in multivariate analysis. As an example, Table 2 shows the impact of rs2307294 on colorectal cancer risk after adjustment for age, gender, TNM, chemotherapy, and a positive family history of colorectal cancer.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>dbSNP ID</th>
<th>Position</th>
<th>Variation</th>
<th>MAF (population studied)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMUG1</td>
<td>rs2233921</td>
<td>80</td>
<td>G/T</td>
<td>T = 0.482 (HAPMAPCEU)</td>
</tr>
<tr>
<td></td>
<td>rs971</td>
<td>422</td>
<td>G/A</td>
<td>A = 0.320 (HAPMAPCEU)</td>
</tr>
<tr>
<td>MBD4</td>
<td>rs2307285</td>
<td>383</td>
<td>A/G</td>
<td>G = 0.197 (PDR90)</td>
</tr>
<tr>
<td></td>
<td>rs2307294</td>
<td>447</td>
<td>G/C</td>
<td>C = 0.100 (NH-PDR90)</td>
</tr>
<tr>
<td>NEIL2</td>
<td>rs1534862</td>
<td>21</td>
<td>C/T</td>
<td>T = 0.219 (HAPMAPCEU)</td>
</tr>
<tr>
<td></td>
<td>rs804292a</td>
<td>133</td>
<td>A/G</td>
<td>G = 0.292 (HAPMAPCEU)</td>
</tr>
<tr>
<td></td>
<td>rs6997097</td>
<td>190</td>
<td>T/C</td>
<td>C = 0.055 (PDR90)</td>
</tr>
<tr>
<td></td>
<td>rs8191670</td>
<td>356</td>
<td>T/C</td>
<td>C = 0.25 (PDR90)</td>
</tr>
<tr>
<td></td>
<td>rs2740439</td>
<td>866</td>
<td>G/C</td>
<td>C = 0.08 (HAPMAPCEU)</td>
</tr>
<tr>
<td></td>
<td>rs4839</td>
<td>969</td>
<td>A/G</td>
<td>G = 0.401 (HAPMAPCEU)</td>
</tr>
<tr>
<td></td>
<td>rs1043180</td>
<td>1,039</td>
<td>C/T</td>
<td>T = 0.111 (HAPMAPCEU)</td>
</tr>
<tr>
<td>NEIL3</td>
<td>rs1055677b</td>
<td>173</td>
<td>C/G</td>
<td>G = 0.068 (HAPMAPCEU)</td>
</tr>
<tr>
<td></td>
<td>rs34652756b</td>
<td>132</td>
<td>T/C</td>
<td>C = 0.087 (HAPMAPCEU)</td>
</tr>
<tr>
<td></td>
<td>rs1055678</td>
<td>342</td>
<td>C/T</td>
<td>T = 0.068 (HAPMAPCEU)</td>
</tr>
<tr>
<td>LIG3</td>
<td>rs1052536</td>
<td>50</td>
<td>C/T</td>
<td>T = 0.500 (HAPMAPCEU)</td>
</tr>
</tbody>
</table>

*bIn linkage disequilibrium (LD) with rs1534862 (Haploview 4.2). Not genotyped.
*aIn LD with rs1055678 (Haploview 4.2). Not genotyped.
cases and controls were matched by age quartiles through bootstrap sampling (10 repetitions). For each subset, the association analyses were repeated and the results averaged. With this approach, we could ascertain that no changes were observed in the 10 different resamplings. In addition, the analysis was repeated in a subgroup of 800 cases and 800 controls matched for sex and age (after matching, mean age in cases \(= 59.0 \) years, mean age in controls \(= 57.7 \) years) and similar results were observed as in the whole group of study; none of the SNPs were associated with cancer risk (data not shown).

Survival analysis

**Training set.** The average (median) OS and EFS for patients was 87.8 (72.0) and 83.0 (68.1) months, respectively. In the preliminary univariate assessment of covariates, gender, age, smoking habit, TNM status, and chemotherapy treatment were associated with OS (Table 3). Advanced age, male gender, and current smoking status were related to a shortened OS. Men in particular also showed a higher risk of relapse or metastasis (OS: HR, 1.54; 95% CI, 1.22–1.94; \(P < 0.01\); EFS: HR, 1.43; 95% CI, 1.14–1.80; \(P < 0.01\)). Four established prognostic factors (TNM status and chemotherapy treatment) were associated with patient survival. Moreover, TNM status was also associated with an increased risk of recurrence.

The data analysis showed that rs2233921 in SMUG1 gene was significantly associated with OS (HR, 0.62; 95% CI, 0.39–1.00; \(P = 0.05\)) but not with EFS (Table 4; Supplementary Table S2). In the dominant model, the association of the T allele with a better survival was even more pronounced (HR, 0.54; 95% CI, 0.36–0.89; \(P = 0.003\)). The univariate Kaplan–Meier plots for OS also reflected these results (log-rank test for the dominant model \(P = 0.008\); Fig. 2A). In contrast, carriers of the GA and TC genotypes of SMUG1 rs971 and NEIL2 rs6997097 polymorphisms were associated with a shorter survival (HR, 1.38; 95% CI, 1.02–1.86; \(P = 0.04\); HR, 1.77; 95% CI, 1.23–2.53; \(P = 0.002\), respectively). However, in the dominant model, only C allele carriers of rs6997097 in NEIL2 showed a shorter survival (HR, 1.73; 95% CI, 1.21–2.47; \(P = 0.003\); Table 4; Supplementary Table S2).

After stratification of patients according to treatment (treated with 5-FU or not), the previously observed
**Table 3.** Clinical and anamnestic characteristics significantly affecting OS and EFS of the colorectal cancer patients with complete follow-up (Cox regression) in the Training set and the Replication set

<table>
<thead>
<tr>
<th></th>
<th>Training set</th>
<th></th>
<th>Replication set</th>
<th></th>
<th>Pooled</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N*</td>
<td>HR (95% CI)</td>
<td>P</td>
<td>N*</td>
<td>HR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>357</td>
<td>Ref.</td>
<td>&lt;0.01</td>
<td>78</td>
<td>Ref.</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Males</td>
<td>509</td>
<td>1.54 (1.22–1.94)</td>
<td>&lt;0.01</td>
<td>154</td>
<td>2.18 (1.05–4.52)</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Age, y</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤55</td>
<td>236</td>
<td>Ref.</td>
<td>&lt;0.01</td>
<td>61</td>
<td>Ref.</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>56–62</td>
<td>203</td>
<td>1.40 (1.00–1.94)</td>
<td>0.05</td>
<td>59</td>
<td>1.12 (0.47–2.70)</td>
<td>0.80</td>
</tr>
<tr>
<td>63–69</td>
<td>221</td>
<td>1.33 (0.98–1.84)</td>
<td>0.08</td>
<td>54</td>
<td>1.40 (0.60–3.31)</td>
<td>0.44</td>
</tr>
<tr>
<td>&gt;69</td>
<td>209</td>
<td>1.87 (1.37–2.57)</td>
<td>&lt;0.01</td>
<td>58</td>
<td>2.43 (1.08–5.50)</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Smoking habit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yesa</td>
<td>116</td>
<td>1.54 (1.13–2.09)</td>
<td>0.01</td>
<td>42</td>
<td>2.60 (1.33–5.08)</td>
<td>0.005</td>
</tr>
<tr>
<td>No</td>
<td>428</td>
<td>Ref.</td>
<td>&lt;0.01</td>
<td>180</td>
<td>Ref.</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><strong>pT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>Ref.</td>
<td>&lt;0.01</td>
<td>19</td>
<td>Ref.</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>2</td>
<td>130</td>
<td>2.66 (0.82–8.67)</td>
<td>0.10</td>
<td>39</td>
<td>2.16 (0.25–18.53)</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>420</td>
<td>5.84 (1.87–18.28)</td>
<td>&lt;0.01</td>
<td>122</td>
<td>4.76 (0.65–35.90)</td>
<td>0.13</td>
</tr>
<tr>
<td>4</td>
<td>90</td>
<td>12.30 (3.85–39.23)</td>
<td>&lt;0.01</td>
<td>46</td>
<td>4.63 (0.57–34.58)</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>pN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>359</td>
<td>Ref.</td>
<td>&lt;0.01</td>
<td>149</td>
<td>Ref.</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>1</td>
<td>213</td>
<td>2.04 (1.56–2.67)</td>
<td>&lt;0.01</td>
<td>48</td>
<td>2.04 (1.03–4.07)</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>4.69 (0.37–6.94)</td>
<td>&lt;0.01</td>
<td>17</td>
<td>1.47 (0.44–4.95)</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>pM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>545</td>
<td>Ref.</td>
<td>&lt;0.01</td>
<td>189</td>
<td>Ref.</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>1</td>
<td>150</td>
<td>4.67 (3.68–5.93)</td>
<td>&lt;0.01</td>
<td>29</td>
<td>6.07 (3.06–12.04)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><strong>5-FU-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>319</td>
<td>Ref.</td>
<td>&lt;0.01</td>
<td>102</td>
<td>Ref.</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>No</td>
<td>324</td>
<td>1.66 (1.29–2.12)</td>
<td>&lt;0.01</td>
<td>121</td>
<td>0.71 (0.39–1.30)</td>
<td>0.27</td>
</tr>
</tbody>
</table>

**NOTE:** Significant results are given in bold.

*aNumbers may not add up to 100% of available subjects because of missing information.

bEx-smokers excluded.
Table 4. SNPs associated with OS of patients in each set and in the pooled population (Cox regression for adjusted estimates)

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP ID</th>
<th>Genotype</th>
<th>Training set</th>
<th>Replication set</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>N&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Events</td>
<td>Expected</td>
</tr>
<tr>
<td>SMUG1</td>
<td>rs2233921</td>
<td>GG</td>
<td>205</td>
<td>94</td>
<td>85.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT</td>
<td>302</td>
<td>181</td>
<td>171.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TT</td>
<td>121</td>
<td>38</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT+TT</td>
<td>257</td>
<td>35</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT+TT</td>
<td>257</td>
<td>35</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>92</td>
<td>35</td>
<td>39.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA+TT</td>
<td>157</td>
<td>42</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>5</td>
<td>1</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC+TT</td>
<td>63</td>
<td>16</td>
<td>26.2</td>
</tr>
</tbody>
</table>

NOTE: Significant results are given in bold.

<sup>a</sup>Numbers may not add up to 100% of available subjects because of missing information.

<sup>b</sup>Adjusted for sex, age, TNM, and chemotherapy.
associations for OS analysis were confirmed only for SMUG1 rs2233921 in both groups (data not shown). The univariate Kaplan–Meier plot including the stratification for 5-FU–based chemotherapy provided a strong association with OS (log-rank test \( P = 5.6 \times 10^{-5} \); Fig. 2B). In particular, a better survival was observed for patients carrying the TT genotype and undergoing 5-FU–based chemotherapy. On the other hand, SMUG1 rs971 and NEIL2 associated with a shorter survival only in those subjects who did not undergo 5-FU chemotherapy.

For EFS analysis, a SNP in NEIL2 (rs1534862) was associated with a lower risk of relapse or metastasis in both codominant and dominant models (for CT genotype HR, 0.73; 95% CI, 0.54–0.98; \( P = 0.04 \); for T allele HR, 0.73; 95% CI, 0.55–0.98; \( P = 0.03 \)). Regarding rs1055678 in NEIL3, the TT genotype was not only associated with an increased risk of recurrence in the codominant model (HR, 7.98; 95% CI, 1.04–61.34; \( P = 0.05 \)), but also in the dominant model (HR, 7.76; 95% CI, 1.01–59.52; \( P = 0.05 \); Supplementary Tables S2 and S3). Given the low numbers of cases with variant genotypes, the latter result should be considered with caution. Considering EFS, no association was observed after stratification according to chemotherapy treatment. The only exception was for rs1055678 in NEIL3; however, again, in this case only a reduced number of observations were available (data not shown).

Replication set and pooled analysis. The average (median) OS and EFS for the Replication set were 82.5 (69.3) and 74.2 (58.7) months, respectively. In the univariate assessment of covariates, gender, age, smoking habit, TNM status, and chemotherapy treatment were found to be associated with either OS or EFS, similarly as in the Training set (see Table 3). In this second group, although not reaching statistical significance, all the SNPs previously found associated with either OS and EFS (and also after stratification for 5-FU–based chemotherapy) showed a similar trend of association with the exception of NEIL2 rs6997097 and rs1534862 (Table 4; Supplementary Tables S2 and S3). The pooled analysis of the two sets confirmed the previously observed associations with OS for SMUG1 rs2233921 (for TT genotype HR, 0.61; 95% CI, 0.39–0.95; \( P = 0.03 \); for T allele HR, 0.55; 95% CI, 0.37–0.80; \( P = 0.002 \)) and for SMUG1 rs971 (for GA genotype HR, 1.41; 95% CI, 1.06–1.87; \( P = 0.02 \); for A allele HR, 1.35; 95% CI, 1.03–1.77; \( P = 0.03 \)), as well as for NEIL2 rs6997097 (for TC genotype HR,
Figure 3. *In vitro* assay to test the luciferase activity according to SMUG1 rs2233921 different alleles (G or T). Six independent replicates in three independent experiments are summarized as mean values of relative intensity of luciferase/Renilla luminescence ratio. When considering the luciferase expression with the G allele as the reference (100%), the luciferase expression with the T allele of SMUG1 was 71.2%, 52.6%, and 65.2% respectively in the three experiments.

1.68; 95% CI, 1.19–2.38; \( P = 0.003 \); for C allele HR, 1.66; 95% CI, 1.18–2.34; \( P = 0.003 \). For rs2233921, these results reflected also the univariate Kaplan–Meier plots for OS (log-rank test for the dominant model \( P = 0.01 \); Fig. 2C), even after stratification for 5-FU–based chemotherapy (log-rank test \( P = 0.003 \); Fig. 2D). For EFS analysis, none of the previously associated SNPs were confirmed in the pooled analysis.

**In vitro assay**

For SMUG1 rs2233921, an *in vitro* assay was used to investigate whether its alleles were associated with differential gene expression. The results in the HCT-116 cell line derived from colorectal cancer are reported in Fig. 3. The G-to-T point mutation of rs2233921 (SMUG1) resulted in a reduction of the expression of luciferase in all the experiments performed. Combining results from all the experiments, when considering the luciferase expression with the G allele as the reference (100%), the average luciferase expression with the T allele of SMUG1 was reduced to 63.1% and this difference was statistically significant \( P < 10^{-4} \).

**Discussion**

miRNAs are increasingly recognized as central players in diverse biologic processes, including DNA repair and DNA damage response (25). An increasing body of evidence has indicated miRNAs as cancer diagnostic, prognostic, and predictive clinical biomarkers (26, 27). The presence of SNPs within the 3'-UTRs of target DNA repair genes that alter the binding with specific miRNAs and modulate gene expression, may ultimately affect cancer susceptibility (18). Less is known about the effect of variations in such genes on patient survival and on the effect of therapy via miRNA modulation (28).

In the present study, we selected 12 SNPs within the 3'-UTRs of five genes from the BER pathway, which potentially alter miRNA:mRNA–binding affinity, and we investigated their role in colorectal cancer susceptibility in a case–control study. Subsequently, we tested those SNPs as prognostic factors in patients with colorectal cancer with detailed follow-up information.

The novel finding was that a SNP in the BER gene SMUG1 (rs2233921) was associated with patient survival in a set of 718 patients with colorectal cancer. In particular, carriers of the TT genotype showed a significantly increased OS compared with those with the CG genotype, and the results were even more pronounced in relation to the 5-FU–based chemotherapy. When another set of 229 patients was added, the above results were replicated, though reaching a more robust statistical significance only in the pooled analysis.

The covariate analysis indicated that the two populations (Training and Replication sets), although of considerable different size, displayed similar clinical and anamnestic characteristics to justify a pooled analysis. However, we cannot rule out certain differences, for example due to a different time interval of follow-up and the collection of clinical information. Importantly, the observed associations remained significant even if we would apply the stringent Bonferroni’s correction for multiple comparisons \( (P_{\text{threshold}} = 0.0045, \text{considering } \alpha = 0.05 \text{ and } 11 \text{ independent tests}) \).

The association of SMUG1 rs2233921 with increased OS is of particular interest, especially in relation to the widely used antimetabolite 5-FU–based chemotherapy. Variations in the DNA repair capacity may play a key role in the response to 5-FU exposure (8, 29). In fact, BER glycosylases are the principal enzymes involved in the cellular response to 5-FU–induced damage (8, 30, 31), with SMUG1 recognized as a main player (32). The overall picture suggests that SMUG1 excision activity is protective against toxicity caused by 5-FU (8). Other findings also imply that SMUG1 upregulation might mediate the development of tumor resistance to 5-FU treatment (33).

Because no genetic variants have been found responsible for a differential functionality of SMUG1 enzyme (31), variations in posttranslational regulation, such as the one exerted by miRNAs, could be responsible for a different response to 5-FU treatment and survival. An *in vitro* assay, carried out to test functional differences between the different alleles of SMUG1 rs2233921, confirmed that the T allele was associated with a reduced expression of the reporter gene when compared with the common G allele. This finding supports the hypothesis that the presence of an “alternative” 3'-UTR placed in a “normal” cellular context (and exploiting the “normal” cellular machineries) may modulate the levels of expression of SMUG1. The assay was performed on colorectal cell lines, and it is conceivable that these cells express a miRNA set typical of colorectal cells. At present, we cannot state whether miRNAs, mRNA stability, or other mechanisms are involved in the observed genotype-dependent differential expression of SMUG1. However, this represents an interesting issue for further investigation. In fact, our outcomes support the relevant role of regulation of BER genes in response to DNA-damaging chemotherapy in the...
treatment of colorectal malignancy, which may be affected by interindividual variability.

Other interesting associations were found for SMUG1 rs971 or NEIL2 rs6997097 polymorphisms with a shorter survival in both the Training set and the pooled analysis. For an additional two SNPs, one in NEIL2 (rs1534862) and another in NEIL3 (rs1055678), the associations with risk of recurrence were found only in the Training set. Further investigations on a larger scale are needed to confirm the role of these SNPs. The modulation of NEIL2 and NEIL3 could also be of interest as both genes belong to a family of glycosylases specialized in the recognition of oxidized pyrimidines (31). NEIL2, in particular, recognizes lesions on single-stranded DNA and was also associated with the transcription process (34). Missense polymorphic variants in NEIL2 were associated with colorectal cancer risk, whereas variants in NEIL3 were associated with colorectal adenoma (reviewed in ref. 31). However, to the best of our knowledge, none of their polymorphic variants were evaluated for their functional effects or for their possible modulation in relationship with 5-FU therapy.

Unlike our previous studies (18, 19, 35), we did not observe any association of these 3′-UTR-localized SNPs with risk of colorectal cancer. With a study group of 1,095 cases and 1,469 controls, we had more than 80% power to detect an association for a recessive/dominant with risk of colorectal cancer. With a study group of rs2233921 had the highest DeltaS score among the SNPs having an effect on miRNA binding. In particular, a DeltaS score that summarizes the predicted effect of a was also queried to verify the regulatory effect of the predictions of SNP effects on miRNA-based gene regulation, ware available online (Mirnsnpscore; http://www.bigr

Several miRNAs predicted to bind to the above-mentioned BER genes have been reported in the colorectal miRNAome (36). Interestingly, new findings suggest that miRNAs also affect drug resistance in colorectal cancer (37). For instance, Mir455-3p, predicted to bind in SMUG1 3′-UTR where rs2233921 resides, may target molecular pathways involved in cell survival after 5-FU or cisplatin chemotherapy (38). miRNA pharmacogenomics is a novel and promising field that investigates the role of miRNAs and polymorphisms affecting miRNAs and their target functions for the prediction of drug response or the improvement of drug efficiency (39). Recently, polymorphisms in miRNA-related genes were described as predictors of clinical outcomes in patients with stage III colorectal cancer treated with 5-FU–based chemotherapy (40). A proposed hypothesis was that the generation of a target gene expression control modification due to the SNPs in the miRNA precursor and in the mature molecules could have an effect on the response to 5-FU. We may also extend this hypothesis to the SNPs in miRNA-binding sites of our target BER genes. Accordingly, SNPs residing in the target sites of AP-2a gene were found to alter the miRNA–mRNA interaction resulting in differential target gene expression, which associated with cisplatin chemosensitivity (41). A different posttranslation-regulation of an enzyme due to different miRNA sets binding according to specific alleles has been reported and elucidated in ref. (42). Thus, characterization of the polymorphisms in miRNA-related genes/target sites and the relative functional impact may provide a solid basis for miRNA-based therapeutic approaches.

In conclusion, we have identified SNPs potentially affecting miRNA binding to the 3′-UTR of BER genes among which SMUG1 rs2233921 was strongly associated with clinical outcome of patients with colorectal cancer treated with 5-FU–based chemotherapy. By in vitro assay, we have demonstrated that this polymorphism shows a differential expression of the reporter gene according to its different alleles, supporting a possible regulatory role on the expression of SMUG1. Further studies are needed to replicate these SNPs as predictive biomarkers in independent and larger populations, to more specifically characterize functional relevance of the identified geneticvariants and to find the biologic mechanisms underlying the observed associations.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: B. Pardini, E. Barone, J. Novotny, M. Levy, S. Garritano, L. Vodickova, T. Buchler, P. Vodicka, A. Naccarati Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): B. Pardini, E. Barone, J. Novotny, M. Levy, S. Garritano, L. Vodickova, T. Buchler, P. Vodicka, A. Naccarati Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): B. Pardini, F. Rosa, C. Di Gaetano, T. Buchler, S. Landi, A. Naccarati Writing, review, and/or revision of the manuscript: B. Pardini, F. Rosa, E. Barone, J. Slyskova, J. Novotny, S. Landi, P. Vodicka, A. Naccarati Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): B. Pardini, F. Rosa, J. Slyskova, M. Levy, S. Garritano, L. Vodickova, A. Naccarati Study supervision: B. Pardini, F. Gemignani, P. Vodicka, A. Naccarati
Acknowledgments
Authors would like to thank T. O’Hearn II, B. O’Brien, and E. Van Emburgh for their excellent technical support.

Grant Support
This work was supported by the Grant Agency of the Czech Republic (grants CZ.1.05/2.1.00/01.0024, CZ.1.05/2.1.00/02.0071).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 5, 2013; revised August 24, 2013; accepted September 3, 2013; published OnlineFirst September 13, 2013.

References


Variation within 3’-UTRs of Base Excision Repair Genes and Response to Therapy in Colorectal Cancer Patients: A Potential Modulation of microRNAs Binding

Barbara Pardini, Fabio Rosa, Elisa Barone, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-0314

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2013/09/13/1078-0432.CCR-13-0314.DC1

Cited articles
This article cites 42 articles, 10 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/19/21/6044.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.