Manuscript I

Naccarati A, Rosa F, Vymetalkova V, Barone E, **Jiraskova K**, Di Gaetano C, Novotny J, Levy M, Vodickova L, Gemignani F, Buchler T, Landi S, Vodicka P, Pardini B.

Double-strand break repair and colorectal cancer: gene variants within 3′ UTRs and microRNAs binding as modulators of cancer risk and clinical outcome

Oncotarget (2016) 7(17):23156-69

Double-strand break repair and colorectal cancer: gene variants within 3′ UTRs and microRNAs binding as modulators of cancer risk and clinical outcome

Alessio Naccarati[1,2], Fabio Rosa[3], Veronika Vymetalkova[2,4], Elisa Barone[5], Katerina Jiraskova[2,4], Cornelia Di Gaetano[3,6], Jan Novotny[7], Miroslav Levy[8], Ludmila Vodickova[2,4], Federica Gemignani[5], Tomas Buchler[9], Stefano Landi[5], Pavel Vodicka[2,4,*], Barbara Pardini[3,6,*]

[1]Molecular and Genetic Epidemiology Research Unit, Human Genetics Foundation, Turin, Italy

[2]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Prague, Czech Republic

[3]Genomic Variation in Human Populations and Complex Diseases Research Unit, Human Genetics Foundation, Turin, Italy

[4]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic

[5]Department of Biology, University of Pisa, Pisa, Italy

[6]Department of Medical Sciences, University of Turin, Turin, Italy

[7]Department of Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic

[8]Department of Surgery, First Faculty of Medicine, Charles University and Thomayer University Hospital, Prague, Czech Republic

[9]Department of Oncology, Thomayer Hospital and First Faculty of Medicine, Charles University, Prague, Czech Republic

[*]These authors contributed equally to this work

Correspondence to: Alessio Naccarati, **e-mail:** alessio.naccarati@hugef-torino.org

Keywords: *double-strand break repair (DSBR) genes, colorectal cancer risk and clinical outcomes, miRNA binding sites, 3'UTR polymorphisms, MRE11A*

Received: *July 07, 2015* ***Accepted:*** *December 09, 2015* ***Published:*** *December 31, 2015*

ABSTRACT

Genetic variations in 3′ untranslated regions of target genes may affect microRNA binding, resulting in differential protein expression. microRNAs regulate DNA repair, and single-nucleotide polymorphisms in miRNA binding sites (miRSNPs) may account for interindividual differences in the DNA repair capacity. Our hypothesis is that miRSNPs in relevant DNA repair genes may ultimately affect cancer susceptibility and impact prognosis.

In the present study, we analysed the association of polymorphisms in predicted microRNA target sites of double-strand breaks (DSBs) repair genes with colorectal cancer (CRC) risk and clinical outcome. Twenty-one miRSNPs in non-homologous end-joining and homologous recombination pathways were assessed in 1111 cases and 1469 controls. The variant CC genotype of rs2155209 in *MRE11A* was strongly associated with decreased cancer risk when compared with the other genotypes (OR 0.54, 95% CI 0.38–0.76, p = 0.0004). A reduced expression of the reporter gene was observed for the C allele of this polymorphism by *in vitro* assay, suggesting a more efficient interaction with potentially binding miRNAs. In colon cancer patients, the rs2155209 CC genotype was associated with shorter survival while the TT genotype of *RAD52* rs11226 with longer survival when both compared with their respective more frequent genotypes (HR 1.63, 95% CI 1.06-2.51, p = 0.03 HR 0.60, 95% CI 0.41–0.89, p = 0.01, respectively).

miRSNPs in DSB repair genes involved in the maintenance of genomic stability may have a role on CRC susceptibility and clinical outcome.

INTRODUCTION

Colorectal cancer (CRC) is among the most frequent malignancies worldwide and is the third highest cause of cancer mortality among men and women [1]. Though CRC detected at an early stage can be successfully removed, cancers undetected until an advanced stage with metastases remain incurable [2]. The growing incidence of CRC (2001–2011 growth index 6.0%) was accompanied by a relatively low rate of early detection of the disease [3]. Therefore, there is an urgent need to find biomarkers to aid prevention, treatment and prognosis in CRC.

The molecular etiology of CRC has been explored extensively, revealing that this cancer develops from an accumulation of genomic mutations. Accumulating cellular DNA damage, if not correctly repaired, can lead to genomic instability, apoptosis or senescence and may ultimately predispose the organism to various disorders including cancers. The importance of DNA repair is highlighted by the fact that mutations in a number of DNA repair genes lead to human syndromes that include multiple cancers, immunodeficiency, and phenotypes with chromosomal anomalies [4]. There is a large body of evidence on the associations between DNA repair and the risk of cancer, including CRC [5].

The repair of double-strand breaks (DSBs), the most deleterious type of DNA damage, is a fundamental cellular mechanism to preserve genomic stability [6]. Two pathways are specifically dedicated to the repair of DSBs: homologous recombination (HRe) and non-homologous end joining (NHEJ) [7–9]. The repression of these efficient repair systems permits an accumulation of damage in rapidly dividing cells (such as cancer cells) that can induce apoptosis. Such an effect may also be exerted by radiation therapy (an inducer of DSBs) in cancer patients [7, 9, 10].

DNA repair capacity varies markedly among individuals, and there is evidence that its decrease is associated with increased cancer risk [11, 12]. In this respect, DNA repair genes present numerous single nucleotide polymorphisms (SNPs) with different allelic distributions in the general population. Some of these SNPs have been reported to be associated with cancer susceptibility in a number of malignancies that include CRC [13]. We have previously investigated associations between functional SNPs in DNA repair genes (including DSB repair genes *XRCC3* and *NBS1*) and CRC susceptibility in cases and controls from the Czech Republic [14, 15]. Our findings have suggested that variations in DNA repair genes may be associated with cancer susceptibility through an altered repair function that can also explain some of the phenotypic differences observed in CRC [11, 16, 17].

In recent years, there has been a growing interest in the role of post-transcriptional regulation of gene expression modulated by microRNAs (miRNAs). In concomitance, the importance of the SNPs located within miRNA-binding target sites (miRSNPs) on cancer risk has been highlighted [18, 19]. Regulation and coordination between genes involved in the DNA repair pathways are fundamental for maintaining genome stability, and post-transcriptional gene regulation by miRNAs is one of the critical players in these processes [20]. Thus, subtle effects displayed by SNPs in DNA repair signaling genes may account for some of these variations. In this sense, specific polymorphisms in regulatory regions such as miRNA target sites may also modulate survival and response to therapy in cancer patients [18, 21].

We recently reported associations between miRSNPs in genes of 3 DNA repair pathways (Nucleotide Excision Repair, Base Excision Repair and Mismatch Repair) and CRC risk or clinical outcome [21–23]. SNPs in miRNA target regions of important genes for DSBs repair may also affect the efficiency of translation of corresponding proteins. Thus, in the present study, we hypothesized that variations in DSB genes may modulate signaling response and the maintenance of genomic stability ultimately affecting cancer susceptibility, cancer survival and efficacy of chemotherapy. We investigated the role of 21 polymorphisms in miRNA predicted target sites of NHEJ and HRe genes in association with CRC risk and its clinical outcome in cases and controls from the Czech Republic.

RESULTS

miRSNP selection

Out of the 21 genes involved in the HRe pathway, only 11 had polymorphisms predicted to bind miRNAs in their 3′UTRs. After further selection based on MAF and LD study criteria (see Materials and Methods section), 15 miRSNPs within the 3′UTRs of seven genes (*RAD51*, *RAD52*, *BRCA1*, *MRE11A*, *NBN*, *GEN1* and *XRCC2*) were identified. For NHEJ, from the initial seven genes involved in the pathway, a total of 39 miRSNPs in the 3′UTRs were found. Since the majority of the SNPs are not represented in the Caucasian population, only six polymorphisms in four genes (*XRCC4*, *XRCC5*, *LIG4*, and *NHEJ1)* passing the selection criteria were finally included in the study.

Case-control study

The characteristics of the study participants are presented in Table 1 [21].

None of the 21 SNPs deviated from Hardy-Weinberg equilibrium in control subjects. The strongest association with CRC susceptibility was observed for rs2155209 in *MRE11A,* a gene involved in HRe. The variant genotype CC of this SNP was associated with a decreased risk of cancer (odds ratios (OR) 0.54, 95% confidence intervals (CI) 0.38–0.76, $p = 0.0004$). This association remained significant also

after correction for multiple testing. Moreover, a similar significant association was observed when stratifying the case group according to tumor site (for rectal cancer: OR 0.32, 95% CI 0.18–0.59, p = 0.0002; for colon cancer: OR 0.66, 95% CI 0.45–0.96, p = 0.03) (Table 2 and Supplementary Table 1). Conversely, the variant genotype AA of *RAD52* rs1051669 was associated with increased risk of cancer (OR 1.68, 95% CI 1.11–2.54, p = 0.01).

After stratification for tumor site, two polymorphisms in *RAD52* gene (rs1051669 and rs11571475) were associated with colon cancer risk while one SNP in *NBN* (rs14448) was associated with rectal cancer risk. In particular, carriers of the AA genotype or the variant A allele in rs1051669 were at increased risk to develop cancer in the colon (OR 1.78, 95% CI 1.13–2.80, p = 0.01 and OR 1.72, 95% CI 1.10–2.692, p = 0.02, respectively); whereas carriers of the heterozygous TC genotype of rs11571475 were at decreased risk to develop colon cancer (OR 0.76, 95% CI 0.58–1.00, p = 0.05). This last observed association should be cautiously considered: in the dominant model the presence of the variant C allele was associated with a decreased risk of colon cancer (OR 0.74, 95% CI 0.57–0.97, p = 0.03). However, due to the low frequency of the CC genotype in our study group it was not possible to observe the same effect in the co-dominant model (Table 2). A decreased risk of rectal cancer was observed for carriers of the heterozygous genotype in rs14448 (OR 0.41, 95% CI 0.21–0.80, p = 0.01).

Globally, no significant associations with the risk of CRC were found for any of the studied polymorphisms in the NHEJ pathway. The only observed exception was for *XRCC5* rs1051677 when comparing only rectal cancer patients with controls (codominant model: OR 3.84, 95% CI 1.11–13.31, p = 0.03; recessive model: OR 3.75, 95% CI1.08–12.95, p = 0.04) (Supplementary Table 2).

Contingency tables for SNP interaction analyses

As the variants under investigation are part of two DNA repair pathways where genes work functionally coupled, the polymorphisms emerging from the case-control study were also explored for their potential SNP-SNP interaction in modulating CRC susceptibility. In general, the results revealed a tendency for the under-representation of cases in comparison with controls among carriers of the variant rs2155209 genotype CC in *MRE11A* in combinations with other SNPs in genes of HRe pathway (Supplementary Table 3). Among the most interesting results, the observed protective effect of rs2155209 was increased by the concomitant presence of AA genotype of *XRCC2* rs3218547, whose protective effect was not reaching the significance when analysed alone. Conversely, there was an under-representation of *RAD52* rs1051669 AA genotype (alone associated with an increased risk to develop cancer) in carriers of the variant C allele of rs2155209 (Supplementary Table 3).

Survival analyses

The mean (median) overall survival (OS) and event-free survival (EFS) for patients were 86.5 (80.5) and 72.6 (62.4) months, respectively. Age, gender, T, N, M status, chemotherapy treatment and CRC stage were associated with OS and EFS in the preliminary univariate assessment of covariates (Table 3). Advanced age, male gender and current smoking status were related to a shorter OS. Likewise, men were also at higher risk of relapse or metastasis (OS: Hazard ratio (HR) 1.54; 95% CI 1.23–1.92; p = 0.0001; EFS: HR 1.35; 95% CI 1.09–1.68; p = 0.006). Four established prognostic factors (T, N, M status and stage) were associated with decreased patients' survival and increased risk of recurrence. Moreover, adjuvant chemotherapy was also associated with survival (Table 3).

After adjusting for above significant covariates, CRC patients, particularly those with colon cancer carrying the TT genotype of *RAD52* rs11226, displayed a longer survival in a recessive model (HR 0.70; 95% CI 0.52–0.93; p = 0.02 and HR 0.60; 95% CI 0.41–0.89; p = 0.01, respectively; Supplementary Table 4). Overall, patients also showed a similar significant trend across genotypes in the Kaplan–Meier curves (log-rank test p = 0.004; Median survival time (MST) for CT carriers was 136 months; MST not reached for the other genotypes; Figure 1). Likewise, a similar trend was also found for colon cancer patients (log-rank test p = 0.005; MST for CT carriers was 162 months; MST not reached for the other genotypes; data not shown). Colon cancer patients with the variant CC genotype of *MRE11A* rs2155209 showed a shorter survival when compared with the most frequent TT genotype (HR 1.63; 95% CI 1.06–2.51; p = 0.03) or with T-allele carriers (HR 1.54; 95% CI 1.03–2.31; p = 0.04) (Supplementary Table 4). A similar trend was observed in the univariate Cox model and in the relative Kaplan–Meier curves (log-rank test p = 0.005; MST for CC carriers being 99 months; MST not reached for the other genotypes; Figure 2). No significant association with recurrence risk was observed for any of the HRe genes (Supplementary Table 5).

Overall, no strong associations with survival and risk of recurrence were observed for all analysed miRSNPs in NHEJ genes (Supplementary Table 6 and Supplementary Table 7). Among CRC cases, carriers of the GG genotype of *XRCC5* rs1051685 showed a decreased survival (OS: HR 2.12; 95% CI 1.04–4.32; p = 0.04). A similar trend was observed in the univariate Kaplan–Meier curves, although not being statistically significant (log-rank test p = 0.07; MST for AA and AG carriers = 176 and 178 months, respectively; MST for GG carriers = 65 months).

Luciferase assay

The role of rs2155209 in modulating *MRE11A* expression was investigated by a dual 3′UTR luciferase reporter assay. A statistically significant difference

between the two constructs carrying the different alleles of the SNP was observed ($p = 0.007$, MANOVA). Figure 3 shows the luciferase activity following transfection with the pmirGLO vector without the 3′UTR (used as reference and set as 100%) and with the vectors carrying the alternative alleles in HCT-116 cell line. The average luciferase activity of the vector carrying the C-allele showed a reduction by 14% in comparison with the values obtained for the construct with the T-allele.

Validation on TCGA database

RNA sequencing (RNAseq) data of CRC patients in The Cancer Genome Atlas (TCGA) database were downloaded. The results from RNAseq from 327 tumor tissues and 13 normal-appearing, adjacent mucosa were available [24]. A general overexpression of all 20 transcripts of *MRE11A* was observed in the tumor tissues when compared with healthy tissues (for all $p < 10^{-7}$).

DISCUSSION

In the present study, we investigated the role of 21 miRSNPs in DSB repair genes in modulating CRC susceptibility and clinical outcome. The major finding was the association of the variant CC genotype of *MRE11A* rs2155209 with a decreased risk of CRC. This association was observed independently of the stratification of the cases according to tumor site recorded at diagnosis. The C-allele of the SNP was also related to a reduced activity of the reporter gene in a dual luciferase assay.

MRE11A encodes for a protein that is a component of the MRE11-RAD50-NBS1 (MRN) complex involved in DSB repair by both HRe and NHEJ, in the maintenance of telomere integrity, in DNA recombination during meiosis, and in the signaling of DSB damage [25]. Mutations in *NBS1*, *MRE11A,* and *RAD50* disrupting the functionality of MRN complex may lead to genome instability and carcinogenesis. For instance, these mutations have been

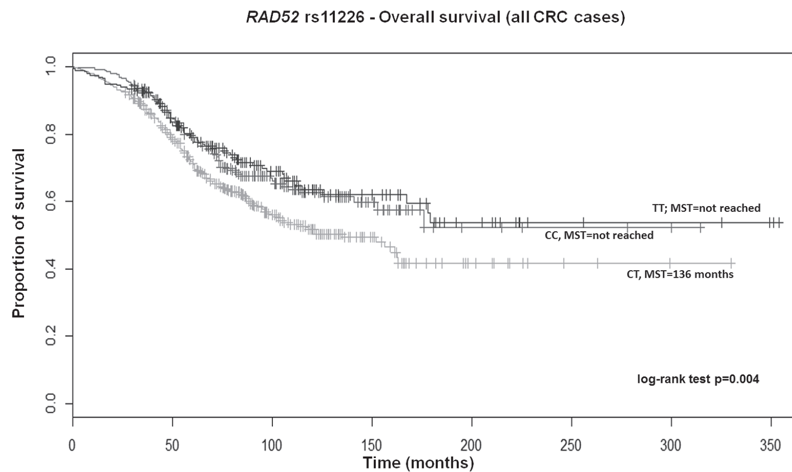


Figure 1: Kaplan-Meier OS curves for *RAD52* rs11226 in all CRC patients. MST = median survival time.

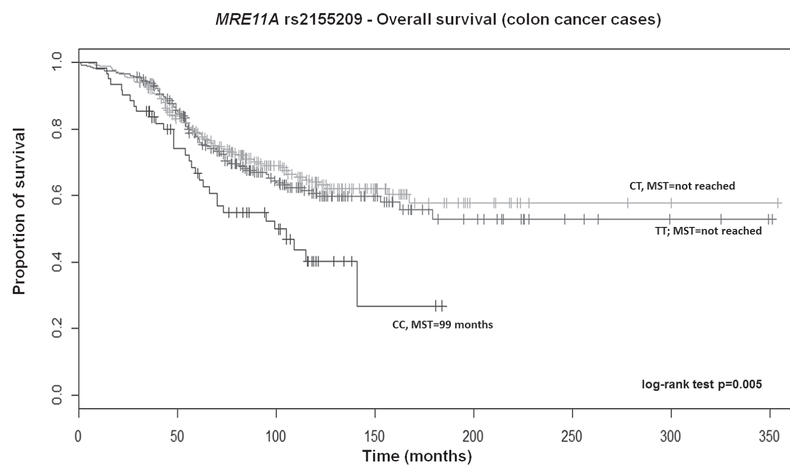


Figure 2: Kaplan-Meier OS curves for *MRE11A* rs2155209 in colon cancer patients. MST = median survival time.

Table 1: Characteristics of the study population

		Cases	Controls	OR	95% CI	*P*
Age (years)	[18,47]	90	591	Ref		
	(47, 55]	208	422	3.24	2.45–4.27	< 0.00001
	(55, 65]	375	286	8.61	6.57–11.28	< 0.00001
	(65,91]	438	170	16.92	12.74–22.47	< 0.00001
Sex	Females	433	660	Ref		
	Males	678	809	1.28	1.09–1.50	0.003
BMI	[0, 23.7]	187	367	Ref		
	(23.7, 26.2]	195	362	1.06	0.82–1.35	0.70
	(26.2, 28.9]	229	323	1.39	1.09–1.78	0.01
	(28.9, 53.1]	224	329	1.34	1.05–1.71	0.02
Smoking	Non smokers	541	815	Ref		
	Smokers	161	328	0.74	0.59–0.92	0.006
	Ex-smokers	341	253	2.01	1.65–2.45	< 0.001
Family History CRC	No	736	1204	Ref		
	Yes	146	142	1.68	1.31–2.16	< 0.0001
Living Area	Town	520	952	Ref		
	Town and country	128	171	1.37	1.06–1.76	0.02
	Country	244	269	1.66	1.35–2.04	< 0.00001
Education	Primary	271	224	Ref		
	Secondary	473	819	0.48	0.39–0.59	< 0.00001
	University or higher	141	345	0.34	0.26–0.44	< 0.00001

Abbreviations: BMI: Body Mass Index OR, odds ratio, 95% CI, confidence interval

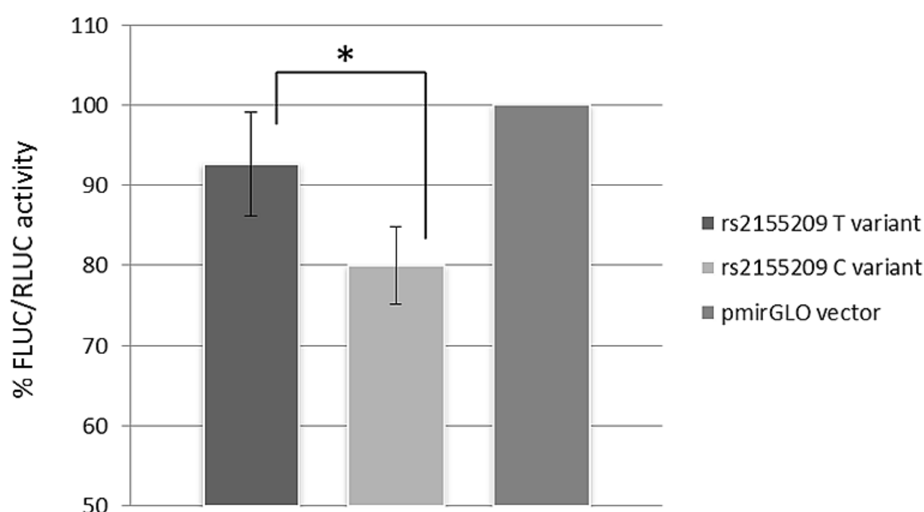


Figure 3: Data show mean values of luminescence activity, normalized to Renilla luciferase levels, (FLUC/RLUC) from four independent experiments. MRE11A expression show a statistical significant ($p = 0.007$) decrease of about 14% in presence of the rs2155209 C-variant, compared to the expression obtained with the T-variant.

Table 2: Significant associations of SNPs in HRe genes with CRC risk (stratification for colon and rectal cancer is also reported)

Gene SNP	Genotype	[a]Controls (n = 1442)	All cancer patients				Rectal cancer patients				Colon cancer patients			
			[a]Cases (n = 1090)	[b]OR	95% CI	*P*	[a]Cases (n = 369)	[b]OR	95% CI	*P*	Cases (n = 710)	[b]OR	95% CI	*P*
RAD52	GG	839	610	Ref			213	Ref			395	Ref		
rs1051669	GA	505	381	1.05	0.85–1.29	0.65	121	0.96	0.70–1.30	0.78	260	1.09	0.86–1.37	0.49
	AA	70	81	1.68	1.11–2.54	**0.01**	28	1.39	0.74–2.59	0.31	52	1.78	1.13–2.80	**0.01**
	GA+AA	575	462	1.12	0.92–1.37	0.25	149	1.01	0.75–1.35	0.97	312	1.17	0.93–1.46	0.18
	GG+GA	1344	991	Ref			334	Ref			655	Ref		
	AA	70	81	1.65	1.10–2.50	**0.02**	28	1.41	0.76–2.61	0.28	52	1.72	1.10–2.69	**0.02**
RAD52	TT	1045	820	Ref			266	Ref			551	Ref		
rs11571475	TC	353	237	0.88	0.70–1.11	0.29	88	1.14	0.82–1.59	0.45	149	0.76	0.58–1.00	**0.05**
	CC	24	20	0.90	0.41–1.96	0.79	12	2.36	0.97–5.73	0.06	8	0.41	0.14–1.26	0.12
	TC+CC	377	257	0.88	0.70–1.11	0.29	100	1.21	0.88–1.67	0.24	157	0.74	0.57–0.97	**0.03**
	TT+TC	1398	1057	Ref			354	Ref			700	Ref		
	CC	24	20	0.92	0.42–2.01	0.84	12	2.28	0.95–5.52	0.07	8	0.44	0.14–1.34	0.15
RAD52	TT	1024	808	Ref			280	Ref			526	Ref		
rs7963551	TG	375	246	0.83	0.66–1.04	0.11	80	0.88	0.64–1.23	0.47	165	0.81	0.63–1.05	0.12
	GG	35	17	0.53	0.26-1.11	0.09	5	0.29	0.08–1.13	0.07	12	0.64	0.29–1.38	0.25
	TG+GG	410	263	0.80	0.64–1.00	**0.05**	85	0.83	0.60–1.15	0.26	177	0.80	0.62–1.02	0.08
	TT+TG	1400	1054	Ref			360	Ref			691	Ref		
	GG	35	17	0.56	0.27–1.16	0.12	5	0.30	0.08–1.16	0.08	12	0.67	0.31–1.46	0.31
MRE11A	TT	610	499	Ref			181	Ref			316	Ref		
rs2155209	TC	638	485	0.86	0.70–1.06	0.16	162	0.75	0.56–1.01	0.06	322	0.94	0.75–1.19	0.61
	CC	180	92	0.54	0.38–0.76	**0.0004**	22	0.32	0.18–0.59	**0.0002**	70	0.66	0.45–0.96	**0.03**
	TC+CC	818	577	0.79	0.65–0.96	**0.02**	184	0.66	0.49–0.87	**0.004**	392	0.88	0.70–1.09	0.24
	TT+TC	1248	984	Ref			343	Ref			638	Ref		
	CC	180	92	0.58	0.42–0.81	**0.001**	22	0.37	0.21–0.67	**0.0009**	70	0.68	0.47–0.97	**0.03**
NBN	TT	1308	1010	Ref			343	Ref			664	Ref		
rs14448	TC	107	74	0.67	0.45–1.00	**0.05**	22	0.41	0.21–0.80	**0.01**	52	0.78	0.51–1.19	0.24
	CC	0	0	-	–	–	0	–	–	–	0	–	–	–
	TC+CC	107	74	0.67	0.45–1.00	**0.05**	22	0.41	0.21-0.80	**0.01**	52	0.78	0.51-1.19	0.24
	TT+TC	1415	1084	Ref			365	Ref			716	Ref		
	CC	0	0	–	–	–	0	–	–	–	0	–	–	–

Abbreviations: OR, odds ratio; 95% CI, confidence interval. Significant results in bold.
[a]Numbers may not add up to 100% of available subjects because of genotyping failure. 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.
[b]Adjusted for sex, age and smoking.
[c]χ^2 and *P*-values for the deviation of observed and the numbers expected from the Hardy–Weinberg equilibrium (HWE) considering all controls.

reported for acute lymphoblastic leukemia [26], head and neck [27], prostate, breast and colorectal [28] cancers. *MRE11A* has been identified as a possible candidate for breast cancer susceptibility by Bartkova and colleagues [29]. Interestingly, *MRE11* overexpression, commonly observed among cancer patients, has been postulated as a mechanism responsible for increasing cancer risk [24]. To support this hypothesis, RNAseq data available in TCGA database also show an overexpression of all available *MRE11A* transcripts in CRC tumor tissues when compared with their healthy tissue counterparts. In this sense, we may hypothesize that a miRNA post-transcriptional regulation of *MRE11A* may be finely modulated by the presence of the identified miRSNP, with the CC genotype contributing to a reduced risk of developing CRC. The low-risk allele (C) is in fact associated with a lower expression of *MRE11A* most probably due to the C-allele stronger interactions with the putative binding miRNAs, as suggested by the results of the functional study.

To correctly interpret these results, we should not exclude the possibility that the observed association may be due to the *linkage disequilibrium* (LD) structure of the investigated locus. Rs2155209 in the Caucasian population is in a LD block spanning over 125Kbp and containing 31 SNPs. Among those SNPs, it is the only one in the 3′UTR and is indicated as one of the variants in the block describing one of the most common haplotypes (with a frequency of approx 25.4% of haplotypes harbouring the C allele) by Haploview software (HG19). The *MRE11A* 3`UTR hosts several binding sites for transcription factors (for instance SMC3, CTCF, RAD21). However, the region surrounding rs2155209, and including the seed of

Table 3: Clinical and anamnestic characteristics significantly affecting Overall Survival (OS) and Event Free Survival (EFS) of the CRC patients with complete follow up (Cox regression)

		OS			EFS	
		N[a]	HR (95% CI)	*P*	HR (95% CI)	*P*
Sex	Females	427	Ref		Ref	
	Males	656	1.54 (1.23–1.92)	**0.0001**	1.35 (1.09–1.68)	**0.006**
Age (years)	55 ≤	293	Ref		Ref	
	56–62	248	1.43 (1.05–1.95)	**0.02**	1.41 (1.06–1.87)	**0.02**
	63–70	294	1.39 (1.04–1.88)	**0.03**	1.19 (0.90–1.58)	0.22
	> 70	248	2.02 (1.50–2.72)	**< 0.0001**	1.04 (0.76–1.42)	0.80
Smoking habit*	No	533	Ref		Ref	
	Yes	493	1.26 (1.02–1.56)	**0.03**	1.14 (0.93–1.41)	0.19
pT	1	50	Ref		Ref	
	2	166	2.64 (0.94–7.40)	0.06	2.18 (0.85–5.55)	0.10
	3	535	5.84 (2.17–15.71)	**0.0005**	5.58 (2.30–13.53)	**0.0001**
	4	136	9.21 (3.36–25.26)	**< 0.0001**	6.96 (2.80–17.27)	**< 0.0001**
pN	0	498	Ref		Ref	
	1	260	2.17 (1.69–2.79)	**< 0.0001**	1.87 (1.46–2.41)	**< 0.0001**
	2	68	3.40 (2.35–4.91)	**< 0.0001**	3.43 (2.45–4.81)	**< 0.0001**
pM	0	725	Ref		Ref	
	1	177	4.80 (3.83–6.02)	**< 0.0001**	4.56 (3.68–5.65)	**< 0.0001**
5FU-based chemotherapy	Yes	411	Ref		Ref	
	No	440	1.42 (1.13–1.790)	**0.003**	0.85 (0.68–1.06)	0.14
Histologic Grade	1	125	Ref		Ref	
	2	464	1.84 (1.26–2.69)	**0.002**	1.42(1.00–2.02)	**0.05**
	3-4	199	2.35 (1.57–3.53)	**< 0.0001**	1.88 (1.29–2.76)	**0.001**
Stage	1	149	Ref		Ref	
	2	293	2.14 (1.32–3.48)	**0.002**	2.47 (1.51–4.05)	**0.0003**
	3	244	3.75 (2.33–6.03)	**< 0.0001**	3.87 (2.38–6.31)	**< 0.0001**
	4	177	11.87 (7.44–18.95)	**< 0.0001**	11.86 (7.42–18.98)	**< 0.0001**

*Ex-smokers included in non-smokers

Abbreviations: HR, hazard ratio; 95% CI, confidence interval. Significant results in bold.

[a] Numbers may not add up to 100% of available subjects because of missing information

the miRNAs predicted to bind on the SNP of interest, is not a site for RNA-binding proteins (RBPs), supporting the hypothesis of a miRNA related post-transcriptional mechanism. Although evidence highlighted a potential miRNA-dependent regulation of the gene, *MRE11A* expression could not be affected solely by rs2155209; other SNPs could be causally linked to the risk of CRC by different mechanisms.

A significant role for genetic factors in CRC has been confirmed by genome-wide association studies (GWAS) and large-scale replication studies, which have identified so far 124 SNPs associated with this cancer (the GWAS catalog: https://www.ebi.ac.uk/gwas/search?query = colorectal% 20cancer). However, the loci identified were estimated to account collectively for approximately 6% of the excess familial risk of CRC [30], suggesting that additional SNPs remain to be identified. The rs2155209 polymorphism has also been previously associated with an increased risk of myocardial infarction, breast and bladder cancer [31–33]. For the latter, the rare allele was found at an increased risk, but genotype distribution in controls was not found in HW equilibrium. To the best of our knowledge, no reports have documented any association with CRC risk. Interestingly, other genetic variants in *MRE11* (not in linkage with rs2155209) have been associated with various cancers including breast, bladder and ovarium [28, 29, 34, 35].

In the last years, the interest on miRNAs has increased since they have been recognized as pivotal players in diverse biologic processes, including DNA repair and DNA damage response [36, 37]. An increasing body of evidence indicates the possibility to use miRNAs as diagnostic, prognostic and predictive clinical biomarkers [20]. In this context, the presence of SNPs within the 3′UTRs of target DNA repair genes could alter the binding with specific miRNAs, modulating gene expression and ultimately affecting, besides cancer susceptibility [18, 22, 38], also therapy outcomes [39] and survival [21]. As an example, an association between a miRNA binding site SNP within the DNA repair gene *RAD51* with bladder cancer risk and radiotherapy outcomes has been reported [39].

miRNAs typically mediate fine regulation of gene expression, tuning rather than altering protein levels [37]. There is evidence that miRNAs can control DNA damage response by interacting with DNA repair genes. Most of the studies have been conducted on cancer cell lines, and it is not clear whether miRNAs mediate DNA repair in healthy cells [37]. Most recently it has been hypothesized that high expression levels of DNA repair proteins are detrimental to DSB repair as the stoichiometry of factors in specific pathways is important. miRNAs could then facilitate DNA repair by maintaining the optimal levels of repair proteins [37], and there could be a further modulation mediated by SNPs in miRNA seeds or in target regions. In the context of CRC and DNA repair, our group has provided the first evidence that variations in miRNA-binding sites in Base Excision Repair genes 3′UTRs may modulate prognosis and therapy response [21]. In the present study, among CRC patients, and specifically those with colon cancer, carriers of the TT genotype of *RAD52* rs11226 displayed a better survival while carriers of the *MRE11A* rs2155209 variant CC genotype showed a shorter survival. Notably, MRE11 protein deficiency has been recently observed to be associated with improved survival of stage III colon cancer patients, independently of treatment [40]. This study supports our finding where CC genotype of *MRE11A* rs2155209 is associated with shorter survival. We can theorize that the modulatory role by the observed SNP on the expression of MRE11 protein may also influence the prognosis of cancer. RAD52 is a key protein in the homologous recombination pathway. In humans, it is known to exist in an oligomeric form in order to bind single-stranded DNA (ssDNA), to promote ssDNA annealing, to interact with RPA, and under certain specialized conditions, to simulate Rad51-mediated homologous DNA pairing [41]. There is an established interplay between *MRE11A* and *RAD52* genes since the binding of MRN complex to a DSB permits a following recruitment of RAD52 to start the resolution of the damage [42]. Both genes have numerous predicted binding miRNAs in their 3′UTRs, although only very few of them have been validated so far. We have investigated miRNAs predicted to bind to *RAD52* and *MRE11A* where the SNPs found in association lie (reported in Supplementary Table 8). From the available data (source http://www.genecards.org/), many of these miRNAs are expressed in colon tissue. Interestingly, among them, two miRNAs (miR-1296 and miR-296–5p) are predicted to bind both genes. These miRNAs have been described deregulated in cancer and other diseases and, in particular, miR-296–5p has frequently been associated with cancer prognosis [43, 44].

Protein expression levels of NBS1, MRE11, and RAD50 in malignant tissues have also been measured in previous studies. For instance, it was observed that a lower MRE11 expression in tumor cells in bladder and breast tissues was also associated with worse cancer-specific survival compared with high expression [45, 46], and the underlying control mechanism determining these lower expression levels was essentially post-transcriptional and regulated by miR-153 [47]. Additionally, *RAD50* and *NBS1* mRNA levels correlated with expression of all three proteins, implying that transcription of these two genes determines the amount of MRN complex formed. In this sense, MRE11 protein levels seem to adapt in line with the complex formation, with the following degradation of protein molecules that are not required for complex formation [47]. This strong interconnection may explain the other observed associations, in particular for the variants related to patient clinical outcome.

To our knowledge, this is the first study comprehensively investigating the role of SNPs residing in miRNA target sites of DSB repair genes in association with CRC risk and clinical outcome. The study population included in the present work is genetically homogeneous (all Caucasian from the Czech Republic), and clinically well-defined (cases and controls recruited in the same centers with follow-up data collected by the same physicians), thus excluding any possible population stratifications and bias. In addition, the inclusion of 'colonoscopically negative' individuals ensured disease-free control individuals because a negative colonoscopy result is the best available proof of the CRC absence [48]. Since this group of individuals may not necessarily represent the general population, we also included healthy cancer-free individuals recruited among volunteers from blood centers.

We are aware of certain limitations of the present investigation. In the case-control study, controls differed from cases in age and gender distribution, as well as other parameters such as BMI. However, we attempted to control tentative age effect by matching cases and controls by age quartiles through bootstrap sampling, and no changes were observed in the ten different resamplings.

The main and novel finding of the present study was that *MRE11A* rs2155209 resulted strongly associated with a decreased risk of CRC, taking into account also multiple comparisons (by considering a 5% Bonferroni-corrected significance threshold). Moreover, the presence

of one or the other allele of rs2155209 was associated with a different luciferase activity. The present results support the emerging idea of a "miRNA network" that may contribute to CRC [49]. Other miRSNPs, both in the same gene and in other DSB repair genes, were also associated with clinical outcomes highlighting the importance of this repair pathway in survival, most probably as a consequence of an impaired DNA repair system.

It is generally accepted that all DNA repair pathways act in an integrative and collaborative way. Numerous factors affect the decision to repair a DSB via NHEJ or HRe, and accumulating evidence suggests these major repair pathways both cooperate and compete with each other at DSB sites to facilitate efficient repair and promote genomic integrity [50, 51]. We have observed in particular that both *MRE11A* and *RAD52* share miRNAs predicted to bind to regions where SNPs were associated with survival while a SNP interaction analyses revealed an under-representation of certain genotypes among concomitant genotypes of SNPs in both genes in association with CRC risk. However, a larger population is necessary to test the interaction/cooperation of different genes/SNPs in various pathways.

In conclusion, we identified plausible candidate miRSNPs potentially affecting miRNA binding in DSB repair genes that were related either to CRC susceptibility or to patients' survival. Further studies are needed to replicate our findings and assess these miRSNPs as predictive biomarkers in independent populations, to functionally characterize the significant genetic variants and to find the biologic mechanisms underlying the associations.

MATERIALS AND METHODS

Study population and data collection

Blood samples were collected from 1126 patients with histologically confirmed CRC attending between September 2003 and October 2010 several oncological departments in the Czech Republic (three in Prague, one in Benesov, Brno, Liberec, Ples, 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 gastroenterological departments that had negative colonoscopy results for malignancy or idiopathic bowel diseases (Control Group 1). 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 (Control Group 2) collected from a blood donor centre 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. Among the CRC cases, 397 patients were diagnosed with a tumor in the colon, 334 in the sigmoideum and 377 with rectal cancer (3 cases were lacking the information about the site of the tumor; however, since they had complete survival data, they remained in the survival analysis). Out of the 1469 controls, 688 were cancer-free colonoscopy inspected controls (Control Group 1) and 781 were healthy blood donor volunteers (Control Group 2). Details of CRC cases and controls have been reported previously [21].

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 Committee. Study subjects provided information on their lifestyle habits, BMI, diabetes, and family/personal history of cancer, using a structured questionnaire to determine demographic characteristics and potential risk factors for CRC.

Follow-up of patients

Eight hundred sixty-six CRC cases were monitored with follow-ups until August 31st, 2011. A second group consisting of 232 CRC patients was recruited later on and followed up until March 31st, 2013. For all subjects, clinical data at the time of diagnosis, including location of the tumor, UICC (International Union Against Cancer) tumor-node-metastasis (TNM) stage system, grade and adjuvant chemotherapy treatment were collected, along with information about distant metastasis, relapse and date of death [52].

Four hundred and eleven CRC 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/m^2) and leucovorin (10 mg/m^2) for five days every four weeks six times or a simplified De Gramont regimen which consisted of a 2 h intravenous (i.v.) infusion of leucovorin (200 mg/m^2), then a 5-FU i.v. bolus (400 mg/m^2) followed by a 46h 5-FU continuous i.v. infusion (2400–3000 mg/m^2). Four hundred forty subjects did not receive any adjuvant chemotherapy after surgery. In this study, the outcome variables measured were 5-FU-based chemotherapy, OS (time from diagnosis until death or censorship), and EFS (time of surgery or end of chemotherapy until date of relapse, death or censorship).

Selection of candidate genes and SNPs in miRNA target binding sites

From the complete list of DNA repair genes available online (http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html March 2014 version), seven genes were retrieved in the NHEJ pathway and 21 genes in the HRe pathway.

The approach used to select the candidate miRSNPs was similar to that described in [21]. Briefly, for each gene, SNPs within target binding sites for miRNAs were identified by using the freely available software: *MicroSNiper* (http://cbdb.nimh.nih.gov/microsniper [53], *MiRSNP* (http://202.38.126.151/hmdd/mirsnp/search/ [54]), *Mirnsnpscore* (http://www.bigr.medisin.ntnu.no/mirsnpscore/ [55]), and *Polymirt* (http://compbio.uthsc.edu/miRSNP/ [56]). The 50 detected SNPs were then filtered for their minor allele frequency (MAF > 5%) in Caucasian populations in the SNP database to reach an appropriate representation of all genotypes in our set of cases and controls. The information was primarily derived from 1000genomes project database, phase 1, CEU population; whenever this was not possible, other reference populations were considered (i.e. HAPMAP CEU population) (dbSNP; http://www.ncbi.nlm.nih.gov/SNP/). SNPs with the required MAF were further tested for the possibility to be in LD using HaploView (v. 4.2) with the data from HapMap v. 3, release R2 in the CEU population.

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 selected SNPs were genotyped using the KASP™ genotyping assay, a competitive allele-specific PCR SNP genotyping system (LGC Genomics, Hoddesdon, Herts, UK). For quality control purposes, duplicate samples (5% of the total numbers of samples) were repeated for each SNP, no template controls were included in each plate (NTCs).

DNA cloning and *in vitro* assay

A Dual-Luciferase reporter assay was used to investigate whether the *MRE11A* rs2155209 alleles were associated with a differential gene expression. Initially, a 1031 bp fragment of the 3′UTR region of *MRE11A* containing the T-allele of the SNP was PCR-amplified. The PCR primers were specifically designed to allow the cloning reaction with ClonEZ enzyme. The 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 *XhoI* restriction site within the multiple cloning sites of the pmirGLO vector (Promega, Madison, USA). Each primer was also designed to include a *XhoI* restriction site sequence (c^tcgag) between the two sequences. The complete sequences were: sense primer = AACGAGCTCGCTAGCCTCGAGGGGTG ATAAATCTCTCCAGCTAATTC; and anti-sense primer = CAGGTCGACTCTAGACTCGAGAGCCC ATTGAGATACTTTTTTACTCAG. The vector was linearized with *XhoI* (NEB Inc, Ipswich, USA) and the PCR product was cloned downstream from the firefly luciferase (*Photinus pyralis*) reporter gene, using the Clone EZ PCR Cloning Kit (Genscript, Piscataway, USA). Competent cells NZY5α (NZYTech, Lisbon, Portugal) were used for transformation after the cloning reaction, as suggested by manufacturers. To obtain a vector with an *MRE11A* 3′UTR bearing the C-allele of rs2155209, the construct underwent site-specific mutagenesis using the Quick Change Lightning Site Direct Mutagenesis kit (Agilent, Milano, Italy). The sequences of the mutagenic primers were: sense = attgttttctcctttctgggtaaca cgccctaacttctg; and anti-sense = cagaagttagggcgtgttacccaga aaggagaaaacaat. Following the digestion of the parental (methylated) supercoiled double-stranded DNA with *Dpn I*, *XL10-Gold* ultra-competent cells (Agilent, Milano, Italy) were used for transformation.

For the functional assay, HCT-116 cells were plated at a density of approximately 7×10^4 cells per well in 24-well plates and incubated overnight at 5% CO_2, 37°C in a humidified incubator. Cells were transiently transfected at about 80% confluence using 3 µl of Polyfect transfection reagent (Qiagen, Milano, Italy) and 0.4 µg of the chimeric construct carrying the T or the C allele.

The assays were carried out using the dual-luciferase reporter assay kit (Promega, Milano, Italy). A pmirGLO vector without 3′UTR insert was used as a reference. PmirGLO vectors contain the luciferase gene from *Renilla reniformis* (hRluc-neo), acting as a control reporter to normalize transfection efficiency. Twenty-four hours after transfection, cells were washed with a phosphate-buffered saline solution and lysed with 100 µl of Passive Lysis Buffer (PLB Promega, Milano, Italy) for an optimal stability of the firefly and *Renilla* luciferase reporter enzymes. The culture vessel was shaken for 8 minutes at room temperature. The lysates were used for measuring the activity of firefly (FLUC) and *Renilla* (RLUC) luciferases. Three replicates of all experimental points were performed in each experiment. For each transfection, luminescence intensity was evaluated by a luminometer (Optima FluoStar, BMG, Ortenberg, Germany), and luciferase activities were averaged from four measurements. The luminescence intensities of firefly and *Renilla* luciferase of the non-transfected cells (background) were subtracted from the values obtained for the transfected cells with the pmiRGLO vector containing the 3′UTR. The luminescence of the *Renilla* luciferase was used as the control reporter to calculate the normalized firefly luciferase activity (FLUC/RLUC activity).

Statistical analyses

Pearson`s 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. SNPs were excluded from further analyses if the call rate was < 95%, deviated from Hardy-Weinberg equilibrium in controls at $p < 10^2$, or if genotypes were discrepant in more than 2% of duplicate samples. The multivariate logistic regression

(MLR) analysis was used to test the association between genotypes and risk of CRC. The covariates analysed in the multivariate model were: sex, age, smoking habit (non-smokers vs. smokers and ex-smokers), BMI, familial history of CRC, education level (high, intermediate and low) and living area (country, suburbs, and town). The associations between SNPs and CRC risk were calculated by estimating the ORs and their 95% CI, adjusted for both continuous and discrete covariates. For all the genotypes, regression coefficients for additive models were estimated. For each SNP, we evaluated its association with cancer risk using two different genetic models—dominant, and recessive—to define the best fitting model with the most significant *p*-value. The Bonferroni-corrected significance threshold is 0.002 (for 21 SNPs and $\alpha = 0.05$).

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 CRC patients was estimated using the date of death or the date of follow-up termination as the end point. 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 were used as the end point of follow- up. EFS was defined as the time from surgery/ end 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). Multivariate survival analyses were adjusted for age, gender, smoking and stage.

For the *in vitro* assays, the ratios (FLUC/RLUC) of the measurements of luminescence, each subtracted of its respective background, were compared between genotypes using the multifactor analysis of variance with interactions (MANOVA), where "experiment" and "genotype" were entered as independent factors in the model. The statistical tests were 2-tailed and carried out using Statgraphics Centurion software (StatPoint Technologies, Warrenton, Va).

Authors' contributions

Conception and design: Alessio Naccarati, Barbara Pardini, Pavel Vodicka, Stefano Landi

Development of methodology: Elisa Barone, Katerina Jiraskova, Stefano Landi, Federica Gemignani

Experimental analyses: Veronika Vymetalkova, Barbara Pardini,

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Fabio Rosa, Cornelia Di Gaetano, Stefano Landi, Barbara Pardini

Acquisition of data (acquired and managed patients, etc.): Jan Novotny, Miroslav Levy, Ludmila Vodickova, Tomas Buchler

Writing, review, and/or revision of the manuscript: Alessio Naccarati, Barbara Pardini, Veronika Vymetalkova, Ludmila Vodickova, Elisa Barone, Katerina Jiraskova, Federica Gemignani, Stefano Landi, Pavel Vodicka, Jan Novotny, Miroslav Levy, Tomas Buchler

ACKNOWLEDGMENTS

The authors are very thankful to all volunteers that participated with enthusiasm to the present study and a special thanks to Eric Van Emburgh for his technical support.

GRANT SUPPORT

This work was supported by the Internal Grant Agency of the Ministry of Health of the Czech Republic (grant n.IGA MZ: NT 15–26535A) and the Grant Agency of the Czech Republic (grant n.: CZ GACR:GAP304/12/1585 and 15–14789S) and from the Istituto Toscano Tumori (Italy) (grant n. I56D15000010002).

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer Journal international du cancer. 2015; 136:E359–386.
2. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. Journal of the National Cancer Institute. 2004; 96:1420–1425.
3. Dusek L, Muzik J, Maluskova D, Majek O, Pavlik T, Koptikova J, Melichar B, Buchler T, Finek J, Cibula D, Babjuk M, Svoboda M, Vyzula R, et al. Cancer Incidence and Mortality in the Czech Republic. Klin Onkol. 2014; 27:406–423.
4. Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA. DNA repair pathways as targets for cancer therapy. Nat Rev Cancer. 2008; 8:193–204.
5. Vineis P, Manuguerra M, Kavvoura FK, Guarrera S, Allione A, Rosa F, Di Gregorio A, Polidoro S, Saletta F, Ioannidis JP, Matullo G. A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility. Journal of the National Cancer Institute. 2009; 101:24–36.
6. Symington LS. DNA repair: Making the cut. Nature. 2014; 514:39–40.
7. Jekimovs C, Bolderson E, Suraweera A, Adams M, O'Byrne KJ, Richard DJ. Chemotherapeutic compounds

targeting the DNA double-strand break repair pathways: the good, the bad, and the promising. Front Oncol. 2014; 4:86.

8. Heyer WD, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. Annu Rev Genet. 2010; 44:113–139.
9. Wyman C, Kanaar R. DNA double-strand break repair: all's well that ends well. Annu Rev Genet. 2006; 40:363–383.
10. Helleday T. Homologous recombination in cancer development, treatment and development of drug resistance. Carcinogenesis. 2010; 31:955–960.
11. Slyskova J, Naccarati A, Pardini B, Polakova V, Vodickova L, Smerhovsky Z, Levy M, Lipska L, Liska V, Vodicka P. Differences in nucleotide excision repair capacity between newly diagnosed colorectal cancer patients and healthy controls. Mutagenesis. 2012; 27:519–522.
12. Slyskova J, Cordero F, Pardini B, Korenkova V, Vymetalkova V, Bielik L, Vodickova L, Pitule P, Liska V, Matejka VM, Levy M, Buchler T, Kubista M, et al. Post-treatment recovery of suboptimal DNA repair capacity and gene expression levels in colorectal cancer patients. Molecular carcinogenesis. 2015; 54:769–778.
13. Naccarati A, Pardini B, Hemminki K, Vodicka P. Sporadic colorectal cancer and individual susceptibility: a review of the association studies investigating the role of DNA repair genetic polymorphisms. Mutation research. 2007; 635:118–145.
14. Pardini B, Naccarati A, Novotny J, Smerhovsky Z, Vodickova L, Polakova V, Hanova M, Slyskova J, Tulupova E, Kumar R, Bortlik M, Barale R, Hemminki K, et al. DNA repair genetic polymorphisms and risk of colorectal cancer in the Czech Republic. Mutation research. 2008; 638:146–153.
15. Pardini B, Naccarati A, Polakova V, Smerhovsky Z, Hlavata I, Soucek P, Novotny J, Vodickova L, Tomanova V, Landi S, Vodicka P. NBN 657del5 heterozygous mutations and colorectal cancer risk in the Czech Republic. Mutation research. 2009; 666:64–67.
16. Slyskova J, Cordero F, Pardini B, Korenkova V, Vymetalkova V, Bielik L, Vodickova L, Pitule P, Liska V, Matejka VM, Levy M, Buchler T, et al. Post-treatment recovery of suboptimal DNA repair capacity and gene expression levels in colorectal cancer patients. Molecular carcinogenesis. 2014.
17. Slyskova J, Korenkova V, Collins AR, Prochazka P, Vodickova L, Svec J, Lipska L, Levy M, Schneiderova M, Liska V, Holubec L, Kumar R, Soucek P, et al. Functional, genetic, and epigenetic aspects of base and nucleotide excision repair in colorectal carcinomas. Clinical cancer research. 2012; 18:5878–5887.
18. Cipollini M, Landi S, Gemignani F. MicroRNA binding site polymorphisms as biomarkers in cancer management and research. Pharmgenomics Pers Med. 2014; 7:173–191.
19. Stegeman S, Amankwah E, Klein K, O'Mara TA, Kim D, Lin HY, Permuth-Wey J, Sellers TA, Srinivasan S, Eeles R, Easton D, Kote-Jarai Z, Amin Al Olama A, et al. A Large-Scale Analysis of Genetic Variants within Putative miRNA Binding Sites in Prostate Cancer. Cancer discovery. 2015; 5:368–379.
20. Iorio MV, Croce CM. Causes and consequences of microRNA dysregulation. Cancer journal. 2012; 18:215–222.
21. Pardini B, Rosa F, Barone E, Di Gaetano C, Slyskova J, Novotny J, Levy M, Garritano S, Vodickova L, Buchler T, Gemignani F, Landi S, Vodicka P, et al. Variation within 3′-UTRs of base excision repair genes and response to therapy in colorectal cancer patients: A potential modulation of microRNAs binding. Clinical cancer research. 2013; 19:6044–6056.
22. Naccarati A, Pardini B, Stefano L, Landi D, Slyskova J, Novotny J, Levy M, Polakova V, Lipska L, Vodicka P. Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. Carcinogenesis. 2012; 33:1346–1351.
23. Vymetalkova V, Pardini B, Rosa F, Di Gaetano C, Novotny J, Levy M, Buchler T, Slyskova J, Vodickova L, Naccarati A, Vodicka P. Variations in mismatch repair genes and colorectal cancer risk and clinical outcome. Mutagenesis. 2014; 29:259–265.
24. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–337.
25. Williams GJ, Lees-Miller SP, Tainer JA. Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. DNA Repair. 2010; 9:1299–1306.
26. Mosor M, Ziolkowska I, Pernak-Schwarz M, Januszkiewicz-Lewandowska D, Nowak J. Association of the heterozygous germline I171V mutation of the NBS1 gene with childhood acute lymphoblastic leukemia. Leukemia. 2006; 20:1454–1456.
27. Ziolkowska-Suchanek I, Mosor M, Wierzbicka M, Rydzanicz M, Baranowska M, Nowak J. The MRN protein complex genes: MRE11 and RAD50 and susceptibility to head and neck cancers. Molecular cancer. 2013; 12:113.
28. Dzikiewicz-Krawczyk A. The importance of making ends meet: mutations in genes and altered expression of proteins of the MRN complex and cancer. Mutation research. 2008; 659:262–273.
29. Bartkova J, Tommiska J, Oplustilova L, Aaltonen K, Tamminen A, Heikkinen T, Mistrik M, Aittomaki K, Blomqvist C, Heikkila P, Lukas J, Nevanlinna H, Bartek J. Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. Molecular oncology. 2008; 2:296–316.
30. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, Jaeger E, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nature genetics. 2008; 40:1426–1435.

31. Choudhury A, Elliott F, Iles MM, Churchman M, Bristow RG, Bishop DT, Kiltie AE. Analysis of variants in DNA damage signalling genes in bladder cancer. BMC medical genetics. 2008; 9:69.

32. Verschuren JJ, Trompet S, Deelen J, Stott DJ, Sattar N, Buckley BM, Ford I, Heijmans BT, Guchelaar HJ, Houwing-Duistermaat JJ, Slagboom PE, Jukema JW. Non-homologous end-joining pathway associated with occurrence of myocardial infarction: gene set analysis of genome-wide association study data. PLoS One. 2013; 8:e56262.

33. Wu Z, Wang P, Song C, Wang K, Yan R, Li J, Dai L. Evaluation of miRNA-binding-site SNPs of MRE11A, NBS1, RAD51 and RAD52 involved in HRR pathway genes and risk of breast cancer in China. Mol Genet Genomics. 2015; 290:1141–53.

34. Teo MT, Dyrskjot L, Nsengimana J, Buchwald C, Snowden H, Morgan J, Jensen JB, Knowles MA, Taylor G, Barrett JH, Borre M, Orntoft TF, Bishop DT, Kiltie AE. Next-generation sequencing identifies germline MRE11A variants as markers of radiotherapy outcomes in muscle-invasive bladder cancer. Annals of oncology. 2014; 25:877–883.

35. Park YB, Chae J, Kim YC, Cho Y. Crystal structure of human Mre11: understanding tumorigenic mutations. Structure. 2011; 19:1591–1602.

36. Sharma V, Misteli T. Non-coding RNAs in DNA damage and repair. FEBS letters. 2013; 587:1832–1839.

37. Chowdhury D, Choi YE, Brault ME. Charity begins at home: non-coding RNA functions in DNA repair. Nature reviews Molecular cell biology. 2013; 14:181–189.

38. Slaby O, Bienertova-Vasku J, Svoboda M, Vyzula R. Genetic polymorphisms and microRNAs: new direction in molecular epidemiology of solid cancer. Journal of cellular and molecular medicine. 2012; 16:8–21.

39. Teo MT, Landi D, Taylor CF, Elliott F, Vaslin L, Cox DG, Hall J, Landi S, Bishop DT, Kiltie AE. The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes. Carcinogenesis. 2012; 33:581–586.

40. Pavelitz T, Renfro L, Foster NR, Caracol A, Welsch P, Lao VV, Grady WB, Niedzwiecki D, Saltz LB, Bertagnolli MM, Goldberg RM, Rabinovitch PS, Emond M, et al. MRE11-deficiency associated with improved long-term disease free survival and overall survival in a subset of stage III colon cancer patients in randomized CALGB 89803 trial. PLoS One. 2014; 9:e108483.

41. Lok BH, Powell SN. Molecular pathways: understanding the role of Rad52 in homologous recombination for therapeutic advancement. Clinical cancer research. 2012; 18:6400–6406.

42. Ferrari M, Dibitetto D, De Gregorio G, Eapen VV, Rawal CC, Lazzaro F, Tsabar M, Marini F, Haber JE, Pellicioli A. Functional interplay between the 53BP1-ortholog Rad9 and the Mre11 complex regulates resection, end-tethering and repair of a double-strand break. PLoS genetics. 2015; 11:e1004928.

43. Shivapurkar N, Mikhail S, Navarro R, Bai W, Marshall J, Hwang J, Pishvaian M, Wellstein A, He AR. Decrease in blood miR-296 predicts chemotherapy resistance and poor clinical outcome in patients receiving systemic chemotherapy for metastatic colon cancer. International journal of colorectal disease. 2013; 28:887.

44. van Jaarsveld MT, Wouters MD, Boersma AW, Smid M, van Ijcken WF, Mathijssen RH, Hoeijmakers JH, Martens JW, van Laere S, Wiemer EA, Pothof J. DNA damage responsive microRNAs misexpressed in human cancer modulate therapy sensitivity. Molecular oncology. 2014; 8:458–468.

45. Laurberg JR, Brems-Eskildsen AS, Nordentoft I, Fristrup N, Schepeler T, Ulhoi BP, Agerbaek M, Hartmann A, Bertz S, Wittlinger M, Fietkau R, Rodel C, Borre M, et al. Expression of TIP60 (tat-interactive protein) and MRE11 (meiotic recombination 11 homolog) predict treatment-specific outcome of localised invasive bladder cancer. BJU international. 2012; 110:E1228–1236.

46. Choudhury A, Nelson LD, Teo MT, Chilka S, Bhattarai S, Johnston CF, Elliott F, Lowery J, Taylor CF, Churchman M, Bentley J, Knowles MA, Harnden P, et al. MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer. Cancer research. 2010; 70:7017–7026.

47. Martin RM, Kerr M, Teo MT, Jevons SJ, Koritzinsky M, Wouters BG, Bhattarai S, Kiltie AE. Post-transcriptional regulation of MRE11 expression in muscle-invasive bladder tumours. Oncotarget. 2014; 5:993–1003. doi: 10.18632/oncotarget.1627.

48. Slattery ML, Curtin K, Wolff RK, Boucher KM, Sweeney C, Edwards S, Caan BJ, Samowitz W. A comparison of colon and rectal somatic DNA alterations. Diseases of the colon and rectum. 2009; 52:1304–1311.

49. Vishnubalaji R, Hamam R, Abdulla MH, Mohammed MA, Kassem M, Al-Obeed O, Aldahmash A, Alajez NM. Genome-wide mRNA and miRNA expression profiling reveal multiple regulatory networks in colorectal cancer. Cell death & disease. 2015; 6:e1614.

50. Kass EM, Jasin M. Collaboration and competition between DNA double-strand break repair pathways. FEBS letters. 2010; 584:3703–3708.

51. Khanna A. DNA Damage in Cancer Therapeutics: A Boon or a Curse? Cancer research. 2015; 75:2133–2138.

52. Pardini B, Rosa F, Naccarati A, Vymetalkova V, Ye Y, Wu X, di Gaetano C, Buchler T, Novotny J, Matullo G, Vodicka P. Polymorphisms in microRNA genes as predictors of clinical outcomes in colorectal cancer patients. Carcinogenesis. 2015; 36:82–86.

53. Barenboim M, Zoltick BJ, Guo Y, Weinberger DR. MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. Hum Mutat. 2010; 31:1223–1232.

54. Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, Zhang D. MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. Bmc Genomics. 2012; 13:661.

55. Thomas LF, Saito T, Saetrom P. Inferring causative variants in microRNA target sites. Nucleic acids research. 2011; 39:e109.

56. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic acids research. 2014; 42:D86–91.

Manuscript II

Schneiderova M, Naccarati A, Pardini B, Rosa F, Gaetano CD, **Jiraskova K**, Opattova A, Levy M, Veskrna K, Veskrnova V, Buchler T, Landi S, Vodicka P, Vymetalkova V.

MicroRNA-binding site polymorphisms in genes involved in colorectal cancer etiopathogenesis and their impact on disease prognosis

Mutagenesis (2017) 32(5):533-542

Mutagenesis, 2017, 32, 533–542
doi:10.1093/mutage/gex026
Original Manuscript

OXFORD

Original Manuscript

MicroRNA-binding site polymorphisms in genes involved in colorectal cancer etiopathogenesis and their impact on disease prognosis

Michaela Schneiderova[1], Alessio Naccarati[2,3], Barbara Pardini[3], Fabio Rosa[3], Cornelia Di Gaetano[3,4], Katerina Jiraskova[2,8], Alena Opattova[2,8], Miroslav Levy[5], Karel Veskrna[5], Veronika Veskrnova[6], Tomas Buchler[6], Stefano Landi[7], Pavel Vodicka[2,8,†] and Veronika Vymetalkova[2,8,*,†]

[1]Department of Surgery, General University Hospital in Prague, Prague 12800, Czech Republic, [2]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Videnska 1083, 14200 Prague, Czech Republic, [3]Molecular and Genetic Epidemiology; Genomic Variation in Human Populations and Complex Diseases, IIGM Italian Institute for Genomic Medicine, Via Nizza 52, 10126 Turin, Italy, [4]Department of Medical Sciences, University of Turin, Via Verdi 8, 10124 Turin, Italy, [5]Department of Surgery, First Faculty of Medicine, Charles University and Thomayer Hospital, Videnska 800, 140 59 Prague, Czech Republic, [6]Department of Oncology, First Faculty of Medicine, Charles University and Thomayer Hospital, Videnska 800, 140 59 Prague, Czech Republic, [7]Department of Biology, University of Pisa, Via Derna 1, 56126 Pisa, Italy and [8]Institute of Biology Medicine Genet., First Faculty of Medicine, Charles University, Katerinska 1660/32, 121 08 Prague, Czech Republic

*To whom correspondence should be addressed. Tel: +420 2296 2251; Fax: +420 2 41062782; Email: vpolakova@biomed.cas.cz

[†]These authors contributed equally.

Received 23 November 2016; Editorial decision 29 August 2017; Accepted 30 August 2017.

Abstract

According to the Vogelstein's model of colorectal carcinogenesis, genetic variations in highly penetrant genes may be involved in the colorectal cancer (CRC) pathogenesis. Similarly, aberrant function and/or altered expression of microRNAs (miRNAs) often occur in CRC. In this context, polymorphisms in miRNA-binding sites (miRSNPs) may affect miRNA/target gene interaction, resulting in differential mRNA/protein expression and increased susceptibility to common diseases. To explore this phenomenon, we have mined the 3′ untranslated regions (3′UTRs) of genes known to be frequently mutated in CRC to search for miRSNPs and tested their association with CRC risk and clinical outcome. Eight miRSNPs (rs1804191, rs397768, rs41116 in *APC*; rs1137918, s227091, rs4585 in *ATM*; rs712, rs1137282, rs61764370 in *KRAS*; rs8674 in *PARP1* and rs16950113 in *SMAD7*) were tested for their association with CRC risk in a case-control study (1111 cases and 1469 healthy controls). The role of these miRSNPs was also investigated in relation to clinical outcome on a subset of patients with complete follow-up. rs8679 within *PARP1* was associated with CRC risk and patients' survival. In the dominant model, carriers of at least one C allele were at a decreased risk of cancer ($P = 0.05$). The CC genotype in rs8679 was also associated with an increased risk of recurrence/progression in patients that received 5-FU-based chemotherapy (log-rank test $P = 0.03$). Carriers of the homozygous variant genotype TT for rs712 in *KRAS* gene were associated with a decreased risk of rectal cancer (odds ratio (OR) = 0.65, 95% confidence intervals (CI) 0.43–1.00, $P = 0.05$) while individuals with colon cancer carrying the heterozygous GT genotype showed a longer overall survival (OS) ($P = 0.04$).

© The Author 2017. Published by Oxford University Press on behalf of the UK Environmental Mutagen Society. All rights reserved.
For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

We provide the first evidence that variations in potential miRNA-binding target sites in the 3′ UTR of *PARP1* gene may modulate CRC risk and prognosis after therapy. Further studies are needed to replicate our finding and assess miRSNPs as predictive biomarkers in independent populations.

Introduction

Colorectal cancer (CRC) ranks among the leading causes of cancer-related deaths worldwide (1). In the Czech Republic, the incidence of this kind of cancer ranks the seventh highest worldwide and represents a serious health problem (2 and www.svod.cz). Survival of CRC patients largely depends on the disease stage at diagnosis: the 5-year survival for stage I is 93.6%, whereas it drops dramatically to 8.1% for stage IV patients (3). 5-Fluorouracil (5-FU) is the most frequently used chemotherapeutic agent for treating patients with CRCs. A better understanding of the regulatory factors and signalling pathways that are altered in CRC could provide new insights into disease etiology and individual response to chemotherapeutics.

Decades of studies have revealed that certain genes and pathways, such as *KRAS, BRAF, PIK3CA, SMAD7, PARP1, TP53* and mismatch repair genes, are important in initiation and progression of CRC (4). The *KRAS* mutational status is a predictive marker for response to established epidermal growth factor receptor (EGFR) inhibitors used for the treatment of advanced CRC. In fact, mutations in *KRAS* are associated with resistance to anti-EGFR monoclonal antibody (mAb) therapy with cetuximab or panitumumab. The adverse prognostic impact of *BRAF* mutations has also been documented (5, 6), while that of *PIK3CA* mutations on survival has been restricted to patients with a *BRAF* wild-type tumour (7).

Genetic variability is an important factor modulating disease onset as well as response to therapy and drug toxicity. For example, inherited changes in DNA repair genes have been indicated as useful tools to identify patients at risk for aberrant pharmacokinetic or pharmacodynamic effects (8). Genome-wide association studies have identified an increasing number of single nucleotide polymorphisms (SNPs) showing solid associations with CRC risk that have also been replicated in independent cohorts (9, 10). However, meta-analyses suggest that altogether those SNPs account for only a small proportion of CRC risk so far (11).

In recent years, the interest on the mechanisms of post-transcriptional regulation of gene expression mediated by microRNAs (miRNAs) has increased. In concomitance, the role of SNPs located within miRNA-binding target sites (miRSNPs) on cancer susceptibility has been highlighted. Thus, subtle effects exerted by SNPs in genes relevant for CRC may contribute to its etiopathogenesis. In this sense, specific genetic variations in regulatory regions such as miRNA target sites may also modulate survival and response to therapy in CRC patients (12). We recently reported associations between miRSNPs in genes of four relevant DNA repair pathways (nucleotide excision repair, base excision repair (BER), double-strand breaks repair and mismatch repair) and CRC risk or clinical outcome (13–16). SNPs in miRNA target regions of important genes for CRC etiology (*APC, ATM, BRAF, KRAS, PARP1, PIK3CA, SMAD7* and *TP53*) may also affect the efficiency of translation of the corresponding proteins. Thus, in the present study, we hypothesised that variations in above genes might modify CRC susceptibility, survival and efficacy of chemotherapy via a modulation of the signalling response and the maintenance of genomic stability.

Materials and Methods

Study population

Blood samples were collected from patients with histologically confirmed CRC, recruited between September 2003 and May 2012 at several oncological departments in the Czech Republic. The study included 1111 CRC patients and 1469 controls that provided biological samples and could be genotyped appropriately. All CRC patients suffered from adenocarcinomas. Cases and controls were previously described in details elsewhere (14). All subjects were informed and provided written consent to participate in the study and to use their biological samples for genetic analyses, according to the Helsinki declaration. The design of the study was approved by the local Ethics Committees. Lifestyle and demographic characteristics and potential risk factors for CRC, such as body mass index (BMI), diabetes and family/personal history of cancer were collected in structured questionnaires.

Follow-up of the patients

Eight hundred and sixty-six CRC cases were monitored with follow-up until March 31, 2013. A second group consisting of 232 CRC patients was recruited later on and followed up until March 31, 2013. For all subjects, clinical data at the time of diagnosis, including the location of the tumour, the International Union Against Cancer (UICC) tumour-node-metastasis (TNM) stage system, grade and adjuvant or first line FU-based chemotherapy treatment were assembled, along with information about distant metastasis, relapse and date of death.

Four hundred and eleven CRC cases received as first-line 5-FU-based post-operative therapy. The therapy consisted of either a Mayo regimen, delivered as a bolus infusion of 5-FU (425 mg/m^2) and leucovorin (10 mg/m^2) for 5 days every 4 weeks 6 times or a simplified De Gramont regimen which consisted of a 2 h intravenous (i.v.) infusion of leucovorin (200 mg/m^2), then a 5-FU i.v. bolus (400 mg/m^2) followed by a 46 h 5-FU continuous i.v. infusion (2400–3000 mg/m^2). Metastatic CRC (mCRC) patients were administrated with FOLFOX4 regimen: oxaliplatin (85 mg/m^2) and leucovorin (200 mg/m^2) infusions both given over 120 min at the same time, followed by 5-FU (400 mg/m^2) bolus given over 2–4 min, followed by 5-FU (600 mg/m^2) on the first day. On the second day, leucovorin (200 mg/m^2) infusion both given over 120 min at the same time, followed by 5-FU (400 mg/m^2) bolus given over 2–4 min, followed by 5-FU (600 mg/m^2). The recommended dose schedule is given every 2 weeks. Four hundred and forty subjects did not receive any adjuvant chemotherapy after surgery. In this study, the outcome variables measured were 5-FU-based chemotherapy, overall survival (OS) (time from diagnosis until death or censorship), and event-free survival (EFS) (time of surgery or end of chemotherapy until the date of relapse, death or censorship). mCRC patients were not included in the 5-FU-based survival analysis.

Selection of candidate genes and SNPs in miRNA-binding target sites

Eight genes (*APC, ATM, BRAF, KRAS, PARP1, PIK3CA, SMAD7* and *TP53*) frequently mutated in CRC and important for its etiology were included into this analysis. SNPs in miRNA target regions of above genes may also affect the efficiency of translation of the corresponding proteins. Thus, in the present study, we hypothesised that variations in above genes may modify CRC susceptibility, survival and efficacy of chemotherapy via a modulation of the signalling response and the maintenance of genomic stability.

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

For each of them, we mined their 3′UTR (untranslated region) for miRSNPs by using the freely available software *MicroSNiPer* [http://cbdb.nimh.nih.gov/microsniper (17), *miRSNP* http://202.38.126.151/hmdd/mirsnp/search/ (18), *Mirnsnpscore* http://www.bigr.medisin.ntnu.no/mirsnpscore/ (19) and PolymiRTS http://compbio.uthsc.edu/miRSNP/ (20)] which interrogate the 3′UTR and predicts if an SNP within the target site will disrupt/eliminate or enhance/create a miRNA-binding site.

All 160 detected SNPs were tested for minor allele frequency (MAF, >5% in Caucasian populations) in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) to reach an appropriate statistical power. The selection was primarily done by HapMap CEU population. Whenever this was not possible, other populations were checked as well (i.e. 1000 genomes: phase 1, CEU population). SNPs with the required MAF were further tested for the possibility to be in *linkage disequilibrium* using HaploView (v. 4.2) with the data from HapMap v. 3 (release R2) in the CEU population.

After this selection, 11 SNPs (rs1804191, rs397768, rs41116 in *APC*; rs1137918, s227091, rs4585 in *ATM*; rs712, rs1137282, rs61764370 in *KRAS*; rs8674 in *PARP1* and rs16950113 in *SMAD7*) in the 3′UTRs of five candidate genes complied with the required selection criteria. The workflow for the selection of the miRSNPs is depicted in Figure 1.

SNP genotyping

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 selected SNPs was carried out by using the KASPar chemistry of LGC Genomics (Hoddesdon, Herts, UK: http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/), as previously described (21). For quality control purposes, duplicate samples (5% of the total numbers of samples) were repeated for each SNP, and no template controls were included in each plate. The genotype screening was performed simultaneously for cases and controls. The results were regularly confirmed by random re-genotyping of more than 5% of the samples for each polymorphism, which yielded concordant results. The genotypes with unclear results were excluded from the study.

Bioinformatics

For the selected SNPs, the algorithm RNAcofold (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 algorithm RNAcofold computes the hybridization energy and base-pairing pattern of two RNA sequences (13, 22).

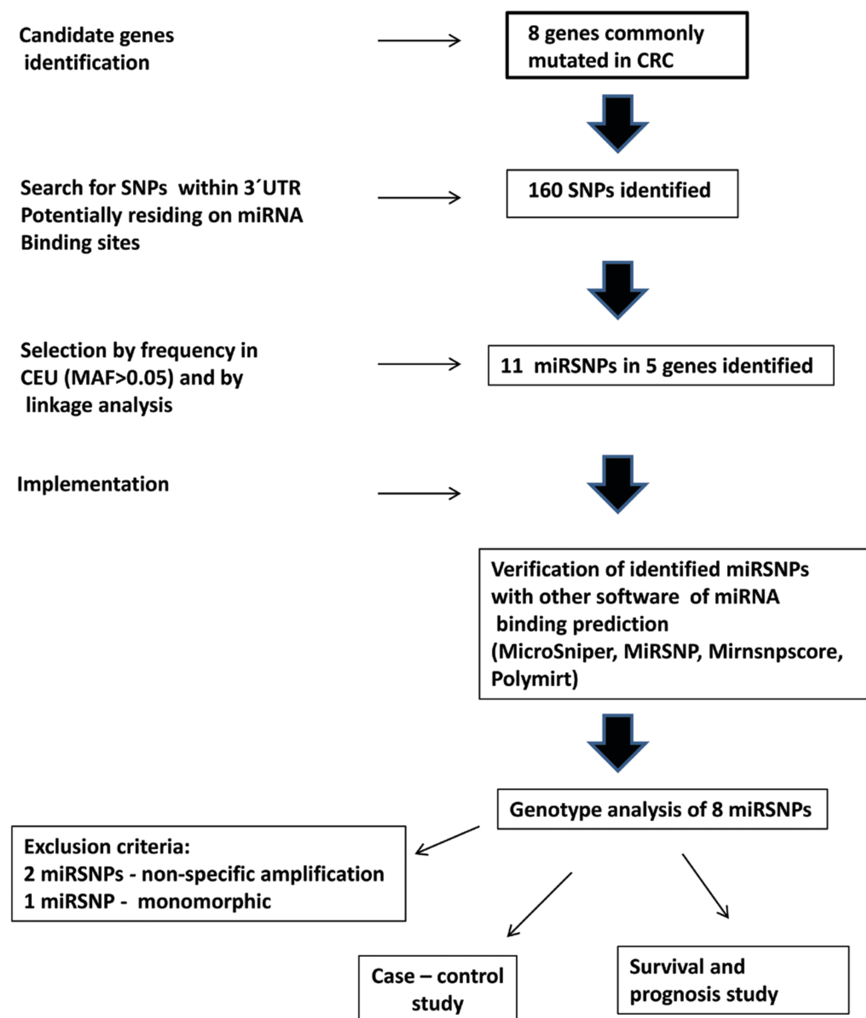


Figure 1. Workflow strategy for selection and analysis of SNPs residing in potential miRNA-binding sites in 3′UTR of genes involved in colorectal cancer etiopathogenesis.

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

The difference of the free energies between the two alleles was computed as 'variation of ΔG' (i.e. |ΔΔG|). Since the neighbour sequence of each SNP can be a target for different miRNAs, we calculated the sum of the absolute values of ΔΔGs for each SNP (i.e. |ΔΔG|tot = Σ |ΔΔG|) (22). The |ΔΔG| tot 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 (13, 22).

We have searched on miRanda (http://www.microrna.org/microrna/getDownloads.do) for the miRSV scores of those miRNAs targeting rs8679 on *PARP1* and rs31764370 on *KRAS*. Similarly, we mined TargetScan (http://www.targetscan.org/vert_71/) for evolutionary conservation of miRNA-binding sites.

Identification of candidate miRSNPs through eQTL analysis

The association between the miSNPs analysed in the study and gene expression levels was obtained from the Genotype-Tissue Expression project (GTEx; https://gtexportal.org/home/, version V6p) and SCAN database (http://www.scandb.org/newinterface/about.html). The GTEx project allows viewing and downloading computed eQTL results and provides a controlled access system for de-identified individual-level genotype, expression and clinical data. The SCAN database provides the number of differentially transcribed genes for each SNP in lymphoblastoid cell lines (LCL) from individuals of Caucasian origin.

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 CRC. The covariates analysed in the multivariate model were sex, age, smoking habit (non-smokers vs. smokers and ex-smokers), BMI, any positive familial history of CRC, education level (high, intermediate and low) and living area (country, town neighbourhood and town). The association between SNPs and CRC risk was calculated by estimating the odds ratio (ORs) and their 95% confidence intervals (CI), adjusted for both continuous and discrete covariates. For all the genotypes, regression coefficients for the additive model were estimated. For all SNPs with significant p values per genotype, the best model (dominant or recessive) was calculated. The Bonferroni-corrected significance threshold is 0.006 (for 9 SNPs and $\alpha = 0.05$).

To overcome a potential age effect due to the difference between cases and controls, they were matched by age quartiles through bootstrap sampling (10 repetitions). For each subset, the association between investigated SNPs and CRC risk was recalculated by estimating the ORs and their 95% CI adjusted for both continuous and discrete covariates (including age).

OS in CRC patients was evaluated using the date of death or the date of the end of the study (August 31, 2011) as the end point of follow-up. 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 end point of follow-up. EFS was defined as the time from surgery/end of therapy to the occurrence of distant metastasis, recurrence or death, whichever came first. The relative risk of death and recurrence was estimated as hazard ratio (HR) using Cox regression. The survival curves for overall and EFS were derived by the Kaplan–Meier method. Multivariate survival analyses were adjusted for age, gender, *T*, *N*, *M* and chemotherapy. Statistical analyses were performed using *R* (http://www.rproject.org, R version 2.14-2).

Results

SNP selection

Out of the eight genes selected (*APC, ATM, BRAF, KRAS, PARP1, PIK3CA, SMAD7* and *TP53*), only five had miRSNPs in their 3′UTRs (160 miRSNPs retrieved). All miRSNPs were filtered according to the study criteria previously described (i.e. MAF and linkage disequilibrium (LD)). In total, 149 SNPs were excluded. Finally, 11 SNPs in 5 genes (rs1804197, rs397768 and rs41116 in *APC;* rs1137918, rs227091 and rs4585 in *ATM;* rs712, rs1137282 and rs61764370 in *KRAS;* rs8674 in *PARP1* and rs16950113 in *SMAD7*) passed the selection and were analysed in the present study. The workflow strategy used in the study for the selection and the analysis of miRSNPs residing in the 3′UTR of genes relevant for CRC is depicted in Figure 1.

After the genotyping run, the assays for SNPs rs1137282 in *KRAS* and rs227091 in *ATM* failed to pass their validation. The data from these assays showed only non-specific amplification and so could not be used to generate any genotyping results. All the other SNPs were genotyped successfully. However, rs1137918 in *ATM* resulted monomorphic in all cases and controls (only AA genotype). Therefore, these three SNPs were not included in the following analyses. Finally, in the present study, we reported the genotype results of eight miRSNPs (Supplementary Table 1, available at *Mutagenesis* online).

Case–control study

The characteristics of the study participants are shown in Supplementary Table 2, available at *Mutagenesis* online. Among the 1111 CRC cases, 731 patients were diagnosed with a tumour in the colon and 377 with rectal cancer (despite for 3 cases the information about the site of the tumour was missing, they remained in the survival analysis since they had complete survival data). Out of the 1469 controls, 688 were cancer-free colonoscopy inspected controls (control group 1) and 781 were healthy blood donor volunteers (control group 2). Compared with subjects of both control groups, CRC cases were more likely to be older and had a slightly higher BMI and were more likely to have a positive family history of CRC and lower formal education. Control group had a higher number of male individuals and current smokers and non-smokers (Supplementary Table 2, available at *Mutagenesis* online).

Thirty-one CRC cases and 28 control samples were eliminated due to genotyping faulty. The distribution of genotypes within the studied genes in the controls was in agreement with Hardy–Weinberg equilibrium (Table 1). None of the investigated miRSNPs showed an association with CRC risk in a co-dominant model. rs8679 within *PARP1* was associated with CRC risk when a dominant model was applied. In particular, carriers of at least one C allele were at a decreased risk of CRC, with a statistically significant OR of 0.82 (95% CI 0.67–1.00, $P = 0.05$). When outcomes were analysed stratifying for cancer site, carriers of the homozygous variant genotype TT for rs712 in *KRAS* gene were associated with a decreased rectal cancer risk (OR = 0.65, 95% CI 0.43–1.00, $P = 0.05$; Table 1). SNPs within *APC*, *ATM* and *SMAD7* were not associated with CRC risk.

However, by taking into account also multiple comparisons (by considering a 5% Bonferroni-corrected significance threshold), none of the above associations remained significant.

Survival analysis

The average (median) OS and EFS for the studied population were 86.5 (80.5) and 72.6 (62.4) months, respectively. In the preliminary univariate assessment of covariates known to affect prognosis, several parameters, such as gender, age, BMI, smoking habit,

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

Table 1. CRC risk according to genotype distribution of investigated SNPs. A further stratification for tumour site is also presented

	All patients						Rectal cancer patients				Colon cancer patients				
	Genotypes	Controls[a] (n = 1442)	Cases[a] (n = 1080)	OR[b]	95% CI	P	Cases[a] (n = 369)	OR[b]	95% CI	P	Cases[a] (n = 708)	OR[b]	95% CI	P	χ^2, P-value HWE[c]
APC	CC	1378	1035	Ref			341	Ref			691	Ref			0.34, 0.85
rs1804197	CA	49	35	0.87	0.51–1.50	0.63	18	1.51	0.76–3.01	0.24	17	0.56	0.28–1.12	0.10	
APC	CC	391	290	Ref			100	Ref			189	Ref			0.30, 0.86
rs41116	CT	700	547	1.06	0.84–1.34	0.62	177	0.90	0.63–1.26	0.53	368	1.10	0.84–1.43	0.50	
	TT	333	219	1.04	0.78–1.38	0.81	77	1.01	0.67–1.52	0.96	142	1.01	0.73–1.39	0.96	
	CT + TT	1033	766	1.13	0.84–1.32	0.65	254	0.93	0.67–1.29	0.66	510	1.07	0.83–1.38	0.60	
	CC + CT	1091	837	Ref			277	Ref			557	Ref			
	TT	333	219	1.00	0.79–1.26	0.97	77	1.09	0.77–1.53	0.64	142	0.95	0.72–1.24	0.69	
APC	AA	497	375	Ref			129	Ref			246	Ref			0.30, 0.86
rs397768	AG	669	501	0.92	0.74–1.14	0.42	172	0.87	0.64–1.20	0.41	327	0.95	0.74–1.21	0.69	
	GG	212	162	0.82	0.60–1.11	0.20	48	0.79	0.51–1.25	0.32	113	0.87	0.60–1.23	0.44	
	AG+ GG	881	663	0.89	0.73–1.09	0.27	220	0.85	0.63–1.16	0.31	440	0.93	0.74–1.18	0.55	
	AA + AG	1166	876	Ref			301	Ref			573	Ref			
	GG	212	162	0.86	0.65–1.14	0.29	48	0.86	0.57–1.23	0.46	113	0.90	0.65–1.23	0.50	
ATM	TT	446	333	Ref			109	Ref			222	Ref			0.45, 0.80
rs4585	TG	685	539	1.14	0.91–1.42	0.26	184	1.37	0.98–1.92	0.07	355	1.08	0.84–1.39	0.54	
	GG	282	203	1.00	0.75–1.33	0.99	73	1.18	0.78–1.79	0.44	129	0.94	0.68–1.29	0.70	
	TG + GG	967	742	1.10	0.87–1.35	0.40	257	1.31	0.95–1.80	0.09	484	1.04	0.82–1.32	0.75	
	TT + TG	1131	872	Ref			293	Ref			577	Ref			
	GG	282	203	0.92	0.72–1.18	0.52	73	0.97	0.68–1.39	0.88	129	0.90	0.68–1.19	0.44	
KRAS	GG	434	356	Ref			132	Ref			224	Ref			0.72, 0.70
rs712	GT	679	502	0.88	0.70–1.01	0.26	168	0.83	0.60–1.15	0.27	331	0.91	0.70–1.18	0.47	
	TT	292	199	0.91	0.68–1.20	0.49	60	0.65	0.43–1.00	**0.05**	139	1.02	0.75–1.40	0.88	
	GT + TT	970	701	0.89	0.72–1.10	0.26	228	0.78	0.57–1.05	0.11	470	0.94	0.74–1.20	0.64	
	GG + GT	1113	858	Ref			300	Ref			555	Ref			
	TT	291	199	0.98	0.77–1.25	0.87	60	0.73	0.50–1.06	0.10	139	1.08	0.83–1.42	0.56	
KRAS	TT	1200	916	Ref			309	Ref			604	Ref			0.01, 0.99
rs61764370	TG	215	167	1.03	0.78–1.35	0.83	62	1.04	0.71–1.54	0.84	105	1.04	0.77–1.41	0.81	
	GG	10	6	1.56	0.51–4.84	0.44	0	–	–	–	6	2.31	0.75–7.15	0.15	
	TG + GG	225	173	1.05	0.81–1.37	0.72	62	1.00	0.65–1.41	1.00	111	1.08	0.81–1.46	0.59	
	TT + TG	1415	1083	Ref			371	Ref			710	Ref			
	GG	10	6	1.56	0.53–4.81	0.44	0	–	–	–	6	2.30	0.74–7.10	0.15	
PARP1	TT	873	687	Ref			230	Ref			455	Ref			0.03, 1.00
rs8679	TC	482	335	0.82	0.66–1.02	0.07	111	0.83	0.61–1.14	0.26	223	0.83	0.65–1.05	0.12	
	CC	66	53	0.79	0.49–1.28	0.34	20	1.11	0.58–2.14	0.74	31	0.73	0.42–1.28	0.27	
	TC + CC	548	388	0.82	0.67–1.00	**0.05**	131	0.87	0.64–1.17	0.35	254	0.82	0.65–1.03	0.08	
	TT + TC	1355	1022	Ref			341	Ref			678	Ref			
	CC	66	53	0.84	0.52–1.36	0.49	20	1.19	0.62–2.25	0.60	31	0.78	0.45–1.35	0.37	
SMAD7	TT	1306	969	Ref			328	Ref			639	Ref			2.62, 0.27
rs16950113	TC	117	101	1.12	0.80–1.57	0.51	34	0.93	0.55–1.56	0.78	66	1.22	0.84–1.78	0.30	
	CC	0	2	–	–	–	0	–	–	–	2	–	–	–	
	TC + CC	117	103	1.16	0.83–1.62	0.39	34	0.93	0.55–1.56	0.78	68	1.27	0.88–1.85	0.20	

Significant results in bold.

[a]Numbers may not add up to 100% of available subjects because of genotyping failure. 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.

[b]Adjusted for sex, age and smoking.

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

Table 2. Clinical and anamnestic characteristics significantly affecting OS and EFS of the CRC patients with complete follow-up (cox regression)

		OS			EFS	
		N[a]	HR (95% CI)	*P*	HR (95% CI)	*P*
Sex	Females	427	Ref		Ref	
	Males	656	1.54 (1.23–1.92)	**0.0001**	1.35 (1.09–1.68)	**0.006**
Age (years)	55≤	293	Ref		Ref	
	56–62	248	1.43 (1.05–1.95)	**0.02**	1.41 (1.06–1.87)	**0.02**
	63–70	294	1.39 (1.04–1.88)	**0.03**	1.19 (0.90–1.58)	0.22
	>70	248	2.02 (1.50–2.72)	**<0.0001**	1.04 (0.76–1.42)	0.80
Smoking habit[b]	No	533	Ref		Ref	
	Yes	493	1.26 (1.02–1.56)	**0.03**	1.14 (0.93–1.41)	0.19
T	1	50	Ref		Ref	
	2	166	2.64 (0.94–7.40)	0.06	2.18 (0.85–5.55)	0.10
	3	535	5.84 (2.17–15.71)	**0.0005**	5.58 (2.30–13.53)	**0.0001**
	4	136	9.21 (3.36–25.26)	**<0.0001**	6.96 (2.80–17.27)	**<0.0001**
N	0	498	Ref		Ref	
	1	260	2.17 (1.69–2.79)	**<0.0001**	1.87 (1.46–2.41)	**<0.0001**
	2	68	3.40 (2.35–4.91)	**<0.0001**	3.43 (2.45–4.81)	**<0.0001**
M	0	725	Ref		Ref	
	1	177	4.80 (3.83–6.02)	**<0.0001**	4.56 (3.68–5.65)	**<0.0001**
5-FU-based	Yes	411	Ref		Ref	
chemotherapy	No	440	1.42 (1.13–1.790)	**0.003**	0.85 (0.68–1.06)	0.14
Stage	I	149	Ref		Ref	
	II	293	2.14 (1.32–3.48)	**0.002**	2.47 (1.51–4.05)	**0.0003**
	III	244	3.75 (2.33–6.03)	**<0.0001**	3.87 (2.38–6.31)	**<0.0001**
	IV	177	11.87 (7.44–18.95)	**<0.0001**	11.86 (7.42–18.98)	**<0.0001**

HR, hazard ratio; 95% CI, confidence interval.

[a]Numbers may not add up to 100% of available subjects because of missing information.

[b]Ex-smokers included in non-smokers.

T, *N*, *M* status and chemotherapy treatment, were associated with OS (Table 2). Advanced age, male gender and current smoking status were related to a shorter OS. Likewise, men also exhibited shorter OS and higher risk of relapse or metastasis (OS: HR 1.54; 95% CI 1.23–1.92; p = 0.0001; EFS: HR 1.35; 95% CI 1.09–1.68; p = 0.006). Four established prognostic factors (*T*, *N*, *M* status and stage) were associated with decreased patients' survival and increased risk of recurrence.

Overall, no association with survival was observed for all the analysed SNPs after adjustment for significant covariates (Supplementary Tables 3 and 4, available at *Mutagenesis* online). After stratification of patients according to tumour location, carriers of the GT genotype in *KRAS* rs712 with a malignancy in colon showed a longer OS (HR = 0.71; 95% CI 0.51–0.98, P = 0.04; Supplementary Table 3, available at *Mutagenesis* online).

In patients undergoing 5-FU-based chemotherapy, rs8679 in *PARP1* gene was associated with EFS (log-rank test P = 0.03; Figure 2). In the univariate analysis, the CC genotype in rs8679 was associated with an increased risk of recurrence or progression only in patients that received 5-FU-based chemotherapy. Additionally, we have also observed a longer OS in patients carrying the TT genotype of rs712 in *KRAS* and not undergoing 5-FU-based chemotherapy (HR = 0.60; 95% CI 0.37–0.97, P = 0.04) (Supplementary Table 5, available at *Mutagenesis* online).

Further stratification of patients according to the stage of disease did not show any significant associations with the OS or EFS (data not shown).

Identification of candidate miRSNPs through eQTL analysis

According to GTEx data, no significant eQTLs were found for SNP rs8679 and rs61764370 in any tissues. Significant cis-eQTL effect was observed for rs712 and *KRAS* in brain, lung, nerve, skin, thyroid and esophagus tissues. Strong eQTL effect was also observed for rs712 and *CASC1* in nerve and skin tissues. In the SCAN database, we observed several differentially expressed genes associated to rs8679 genotypes. In particular, *CDCA7, ULK2* and *C11orf76* genes resulted differentially expressed according to the different rs8679 genotypes. For rs712, only *SFI1* gene expression was altered but only in population from Nigeria. No significant eQTLs were found for rs61764370.

Searching on miRanda for the miRSV scores of those miRNAs targeting rs8679 on *PARP1* and rs31764370 on *KRAS*, we could not retrieve information for any of the miRNAs predicted to bind on *KRAS*. On the other hand, for *PARP1*, 3 out of the 6 miRNAs had a miRSV score (Supplementary Table 6, available at *Mutagenesis* online). In general, only miR-335-5p shows a quite high score, for the other miRNAs the values were close to 0.

Additionally, we searched on TargetScan for the target genes (*PARP1* and *KRAS*) and looked whether the predicted miRNAs reported on Supplementary Table VI were binding or not in an evolutionary conserved site. Again, for *KRAS* none of the predicted miRNAs was found in TargetScan database while the three miRNAs were binding to *PARP1* on a poorly conserved site (7- to 8-mer binding).

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

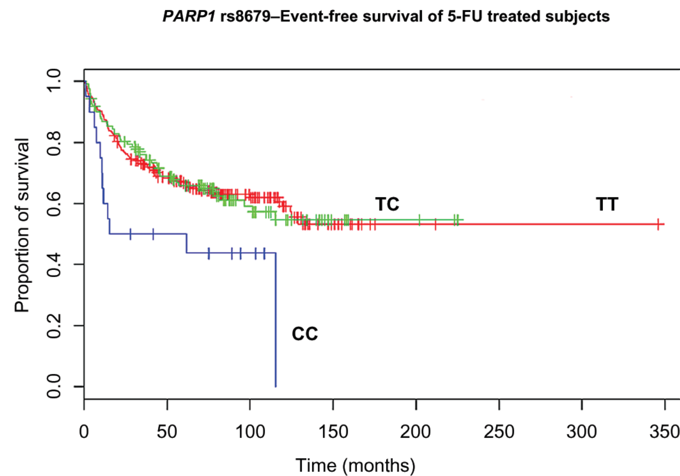


Figure 2. Kaplan–Meier EFS curves in CRC patients undergoing 5-FU-based chemotherapy stratified for rs8679 in *PARP1* gene.

Discussion

The ability of miRNAs to locate and bind mRNA is critical for regulating transcripts level and protein expression (23). The discovery of the role played by miRNAs in biological processes has also provided evidence that genetic variations affecting miRNA function may have a pathogenic role in cancer. SNPs within miRNA-binding sites have been demonstrated highly penetrant for certain phenotypes (24). On the other hand, miRNA-binding site sequences in the majority of protein-coding genes are highly conserved (25). In the last years, a number of studies have suggested the importance of inherited variants in miRNA target sites to human disease susceptibility and progression (24, 26, 27). In this respect, we have searched for miRSNPs within the 3′UTRs of eight genes commonly mutated in CRC (*APC, ATM, BRAF, KRAS, PARP1, PIK3CA, SMAD7* and *TP53*) to evaluate whether these genetic variations could potentially alter the binding with specific miRNAs. The identified miRSNPs were tested in association with CRC risk and clinical outcomes. There is an increasing interest to identify mutations/variations in genes important in tumourigenesis such as *APC, BRAF* and *KRAS* because they are involved in the Wnt and the Ras-Raf-MEK-MAPK signalling cascades (28). Interestingly, *BRAF, PIK3CA* and *TP53* did not present any miRSNPs while several other polymorphisms were present at a very low allele frequency in the Caucasian population. The most important result was that carriers of at least one C allele in rs8679 located in the 3′UTR of *PARP1* were at a decreased risk of CRC. Similarly, patients with homozygous variant genotype TT for rs712 in *KRAS* gene were found to be associated with a decreased rectal cancer risk.

Teo *et al.* previously observed that individuals homozygous for the variant allele of rs8679 within *PARP1* were at increased risks of both breast and bladder cancers (29). MiR-145-3p is predicted to bind to the region where this polymorphism lies (as reported in Supplementary Table VII). The predicted difference of binding energy according to the allele (ΔΔG −5.71 kJ/mol) implies that the less common C allele ΔG being less negative than the more common T allele (29, 30). This suggests a less efficient binding of miR-145-3p in the presence of the C allele on the 3′UTR of *PARP1* mRNA and implicates a potentially decreased post-transcriptional repression of *PARP1* by this miRNA. Previous studies have reported a downregulation of miR-145 expression in CRC (31–33). In CRC cell lines, an ectopic overexpression of miR-145 inhibited cell proliferation, motility and invasion *in vitro*; on the other hand, a stable overexpression of miR-145 suppressed tumour growth and pulmonary metastasis *in vivo* (31). Another miRNA predicted to bind to rs8679 polymorphism is miR-27b, which is frequently downregulated in several malignancies, including lung, breast, colon and prostate cancer (34, 35). This evidence suggests that miR-27b may function as a tumour suppressor (36–38). However, it is still not clear the mechanisms underlying miR-27b downregulation, as well as its targets in human cancers.

We have also calculated the ΔG for the miRNA target region using RNAcofold software as previously described (39, 40). We have investigated the flanking regions of rs8679 that correspond to the binding sites with the predicted miRNAs. By calculating the ΔG, we can predict the impact of the change of allele to affect or impair the binding with predicted miRNAs evaluating the effect of the presence of the two alleles on miRNA binding. For rs8679, there were six miRNAs that could bind in the same position in presence of both alleles and for which we calculated ΔΔG (as reported in the Supplementary Table VI). rs8679 is located within a predicted target region for miR-2116-3p, miR-335-5p and miR-4469. These miRNAs showed the highest energy needed to bind to the 3′UTR with the common T allele. miR-3074-5p exhibited the opposite trend: the highest binding energy was necessary when binding to the region in the presence of the rare C allele, whereas for miR-27b-5p and miR-188-3p there was almost no difference in the binding between alleles. Thus, the miRNA–mRNA binding when the C allele is present, it might be globally more favourable and, theoretically, in this case, it may result in a more stringent repression of translation (i.e. decreased target gene expression). Interestingly, findings from our association study show that carriers of at least one C allele were at a decreased risk of CRC. With an observed decreased risk of cancer for the C allele carriers, this finding supports once again our initial hypothesis of different allele specificity on miRNA-binding target sites that may be reflected in miRNA regulation.

We have observed that the CC genotype in rs8679 was also associated with an increased risk of recurrence or progression only in patients that received 5-FU-based chemotherapy. In light of our results, genetic variations within *PARP1* gene might lead to a decrease in its activity eventually impacting in the failure of apoptosis and ultimately cancer susceptibility. The effectiveness of apoptotic activity after 5-FU treatment could be then lower. As previously mentioned, miR-335-5p is predicted *in silico* to bind to C allele of

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

rs8679. Recently, increased expression levels of miR-335 were markedly associated with CRC tumour size and differentiation. An overexpression of miR-335 in CRC cells facilitated cell proliferation *in vitro* and tumour growth *in vivo* (41).

The study of the modulating effect of miRNAs on *PARP1* protein levels in CRC tumours is particularly important because of the current interest in the use of *PARP1* inhibitors as a single agent or as a chemo or radiosensitizer in cancer treatment (42). Interestingly, for the same polymorphism, the CC genotype was associated with an increased risk of recurrence or progression only in patients that received 5-FU-based chemotherapy. The close association between *PARP1* gene and 5-FU-based chemotherapy was shown in the study of Cheng and colleagues (43). Authors observed a decrease in *PARP1* expression in mice with induced liver cancer treated by galactosylated chitosan/5-FU. *PARP1* is an enzyme that performs central roles in the repair of damaged DNA since it initiates BER. The activation of *PARP1* following severe DNA damage induced by 5-FU results in depletion of cellular energy. In order to prevent the consumption of NAD+ and adenosine triphosphate, activated caspase-3 cleaves and inactivates *PARP1*, which results in apoptosis (44). In light of our results, genetic variations within *PARP1* gene might lead to a decrease in its activity.

rs61764370, residing in the *lethal-7* (*let-7*) miRNA complementary binding site (LCS6) and rs712 of *KRAS* 3′UTR are two miRSNPs that were previously associated with CRC risk and patients survival (45–50). However, according to a recent meta-analysis *KRAS* rs61764370 is not suitable for personalised therapeutic strategies for CRC outcome (51). Our results support the conclusions from this study. No comprehensive meta-analysis has been conducted on the predictive role of rs712 yet. Concerning this specific miRSNP, our results are in disagreement with a recent study where T allele carriers had an increased risk of developing CRC (49). However, this study comprised a considerably smaller population than the present one (339 CRC patients and 313 age- and sex-matched controls). miR-34 family is predicted to bind to the region where this polymorphism lies (reported in Supplementary Table VII). miR-34a and miR-34b/c are involved in the suppression of epithelial-to-mesenchymal transition (EMT) by directly inhibiting the expression of the EMT-inducing transcription factor (EMT-TF) SNAIL (52). A few studies have already described that miR-34 family is abnormally expressed in several types of cancers and regulate several cellular events, including cell cycle, cell migration and apoptosis (53, 54), and since recently miR-34a is considered as a critical mediator of p53 function (55). miR-34b/c expression was reported to be consistently downregulated in CRC (56, 57). A meta-analysis showed that miR-34 family members could become potential diagnostic and prognostic biomarkers in some types of human cancer (58). In addition to the already discussed meta-analysis (51), studies of Zhang *et al.* (47), Graziano *et al.* (45) and Ruzzo *et al.* (46) focused on patients whose primary tumours were positive for a *KRAS* codon 61 mutation and who were treated with cetuximab in combination with irinotecan. While some authors (45, 46) found that carriers of the *KRAS* (LCS6; rs61764370) variant G allele showed worse OS and progression-free survival (PFS), in the study of Zhang *et al.* (47), patients with the same allele had longer OS and PFS. Other studies reported an improvement in the survival of early-stage CRC cases with the LCS6 variant (59) and a reduced risk of mortality in late-stage CRC for the same carriers (60). No effect of the LCS6 variant allele was observed on response rate in mCRC receiving cetuximab (48). Winder *et al.* (50) found that TT carriers with mutated *KRAS* treated with irinotecan/cetuximab in the EPIC trial had significantly better PFS compared with those harbouring the G allele (TG + GG). These conflicting results refer that the LCS6 variant allele has different predictive values in mCRC patients treated with cetuximab alone or in combination with 5-FU/oxaliplatin than in patients treated with cetuximab in combination with irinotecan. As several of the studies reported significant associations with outcomes in well-conducted, prospective studies, the prognostic value of *KRAS* (LCS6; rs61764370) genotype should be largely dependent on the combination of therapy used in conjunction with the anti-EGFR treatment. Additional studies are required to determine the effectiveness of *KRAS* miRSNPs in the prognosis of patients treated with specific anti-EGFR therapy regimens. Future studies should also analyse the effect of *let-7* and its genetic modulation in early stage patients and those treated with chemotherapy. There are several miRNAs predicted to bind in the region where rs61764370 polymorphism in *KRAS* gene lies. However, for many of them, there is no evidence on their functionality or expression levels in colorectal tissues. In fact, few data and no publications are retrievable for miR-6089, miR-4705, miR-3975 and miR-6134. miR-1972 resulted the miRNA with more impact in the binding with the region hosting rs61764370. miR-1972 was identified as serum diagnostic miRNA in lung (61) and bladder cancer (62).

Only three miRNAs are predicted to bind to the region where rs61764370 lies, independently of the allele present. Interestingly, calculating the ΔΔG for the different alleles, the miRNA–mRNA binding is stronger with the presence of the variant G allele. This results in a stronger negative regulation of the target gene expression. miR-3975 and miR-6134, also binding to this region, did not show any difference in the binding with the target region according to the presence of the different alleles (Supplementary Table VI).

Supposing that miRNA-binding site might result in the change of target gene expression, a cis-eQTL effect was investigated using the GTEx and SCANDB databases. Significant cis-eQTL effect was observed for rs712 and *KRAS* gene in brain, lung, nerve, skin, thyroid and esophagus tissues. For rs8679 polymorphism, *CDCA7*, *ULK2* and *C11orf76* genes resulted differentially expressed. However, mechanisms behind these interactions have to be further clarified. No significant eQTLs were found for rs61764370.

We are aware of some limitations of the present study. For instance, there was a different distribution in age and gender between cases and controls, as well as other parameters such as BMI. We attempted to control the potential age effect by matching cases and controls by age quartiles through bootstrap sampling; however, the results from the MLR did not change considerably between ten different bootstrap rounds of re-sampling.

The principal and novel finding of the present study is an association between rs8679 in *PARP1* and either a decreased risk of CRC or an increased risk of recurrence or progression in patients that received 5-FU-based chemotherapy. Further studies are needed to validate our findings and assess this miRSNP as a predictive biomarker in independent patient cohorts, to functionally characterise this genetic variant and to find the biological mechanisms underlying the associations.

Supplementary Material

Supplementary data is available at *Mutagenesis* online.

Funding

This work was supported by the Grant Agency of the Ministry of Health of the Czech Republic (AZV MZ 15-26535A, COST LD 14050 and partly by AZV

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

MZ 15-27580), Grant Agency of the Czech Republic GA 17-16857S and from the Istituto Toscano Tumori (Italy) (grant n. I56D15000010002). This work was supported by Fondazione Umberto Veronesi 'Post-doctoral fellowship Year 2014, 2015, 2016 and 2017' (BP recipient).

Conflict of interest statement: None declared.

References

1. Haggar, F. A. and Boushey, R. P. (2009) Colorectal cancer epidemiology: Incidence, mortality, survival, and risk factors. *Clin. Colon Rectal Surg.*, 22, 191–197.
2. Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G. S., Barzi, A. and Jemal, A. (2017) Colorectal cancer statistics, 2017. *CA. Cancer J. Clin.*, 67, 177–193.
3. O'Connell, J. B., Maggard, M. A. and Ko, C. Y. (2004) Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst.*, 96, 1420–1425.
4. Patel, O., Clyde, D., Chang, M., Nordlund, M. S., Steel, R., Kemp, B. E., Pritchard, D. M., Shulkes, A. and Baldwin, G. S. (2012) Pro-GRP-derived peptides are expressed in colorectal cancer cells and tumors and are biologically active in vivo. *Endocrinology*, 153, 1082–1092.
5. Fariña-Sarasqueta, A., van Lijnschoten, G., Moerland, E., Creemers, G. J., Lemmens, V. E., Rutten, H. J. and van den Brule, A. J. (2010) The BRAF V600E mutation is an independent prognostic factor for survival in stage II and stage III colon cancer patients. *Ann. Oncol.*, 21, 2396–2402.
6. Yokota, T., Ura, T., Shibata, N., *et al.* (2011) BRAF mutation is a powerful prognostic factor in advanced and recurrent colorectal cancer. *Br. J. Cancer*, 104, 856–862.
7. Rosty, C., Young, J. P., Walsh, M. D., *et al.* (2013) PIK3CA activating mutation in colorectal carcinoma: Associations with molecular features and survival. *PLoS One*, 8, e65479.
8. Bhushan, S., McLeod, H. and Walko, C. M. (2009) Role of pharmacogenetics as predictive biomarkers of response and/or toxicity in the treatment of colorectal cancer. *Clin. Colorectal Cancer*, 8, 15–21.
9. Fernandez-Rozadilla, C., Palles, C., Carvajal-Carmona, L., *et al.*; EPICOLON Consortium. (2013) BMP2/BMP4 colorectal cancer susceptibility loci in northern and southern European populations. *Carcinogenesis*, 34, 314–318.
10. Smith, C. G., Fisher, D., Harris, R., *et al.* (2015) Analyses of 7,635 patients with colorectal cancer using independent training and validation cohorts show that rs9929218 in CDH1 is a prognostic marker of survival. *Clin. Cancer Res.*, 21, 3453–3461.
11. Whiffin, N., Hosking, F. J., Farrington, S. M., *et al.*; Swedish Low-Risk Colorectal Cancer Study Group. (2014) Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum. Mol. Genet.*, 23, 4729–4737.
12. Cipollini, M., Landi, S. and Gemignani, F. (2014) MicroRNA binding site polymorphisms as biomarkers in cancer management and research. *Pharmgenomics Pers. Med.*, 7, 173–191.
13. Naccarati, A., Pardini, B., Stefano, L., *et al.* (2012) Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. *Carcinogenesis*, 33, 1346–1351.
14. Vymetalkova, V., Pardini, B., Rosa, F., *et al.* (2014) Variations in mismatch repair genes and colorectal cancer risk and clinical outcome. *Mutagenesis*, 29, 259–265.
15. Pardini, B., Rosa, F., Barone, E., *et al.* (2013) Variation within 3'-UTRs of base excision repair genes and response to therapy in colorectal cancer patients: A potential modulation of microRNAs binding. *Clin. Cancer Res.*, 19, 6044–6056.
16. Naccarati, A., Rosa, F., Vymetalkova, V., *et al.* (2016) Double-strand break repair and colorectal cancer: Gene variants within 3' UTRs and microRNAs binding as modulators of cancer risk and clinical outcome. *Oncotarget*, 7, 23156–23169.
17. Barenboim, M., Zoltick, B. J., Guo, Y. and Weinberger, D. R. (2010) MicroSNiPer: A web tool for prediction of SNP effects on putative microRNA targets. *Hum. Mutat.*, 31, 1223–1232.
18. Liu, C., Zhang, F., Li, T., Lu, M., Wang, L., Yue, W. and Zhang, D. (2012) MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics*, 13, 661.
19. Thomas, L. F., Saito, T. and Sætrom, P. (2011) Inferring causative variants in microRNA target sites. *Nucleic Acids Res.*, 39, e109.
20. Ziebarth, J. D., Bhattacharya, A., Chen, A. and Cui, Y. (2012) PolymiRTS database 2.0: Linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res.*, 40, D216–D221.
21. Naccarati, A., Rosa, F., Vymetalkova, *et al.* (2015) Double-strand break repair and colorectal cancer: Gene variants within 3' UTRs and microRNAs binding as modulators of cancer risk and clinical outcome. *Oncotarget*, 7, 23156–23169.
22. Landi, D., Gemignani, F., Pardini, B., *et al.* (2012) Identification of candidate genes carrying polymorphisms associated with the risk of colorectal cancer by analyzing the colorectal mutome and microRNAome. *Cancer*, 118, 4670–4680.
23. Preskill, C. and Weidhaas, J. B. (2013) SNPs in microRNA binding sites as prognostic and predictive cancer biomarkers. *Crit. Rev. Oncog.*, 18, 327–340.
24. Ryan, B. M., Robles, A. I., McClary, A. C., *et al.* (2015) Identification of a functional SNP in the 3'UTR of CXCR2 that is associated with reduced risk of lung cancer. *Cancer Res.*, 75, 566–575.
25. Ryan, B. M., Robles, A. I. and Harris, C. C. (2010) Genetic variation in microRNA networks: The implications for cancer research. *Nat. Rev. Cancer*, 10, 389–402.
26. Sethupathy, P. and Collins, F. S. (2008) MicroRNA target site polymorphisms and human disease. *Trends Genet.*, 24, 489–497.
27. Maxwell, E. K., Campbell, J. D., Spira, A. and Baxevanis, A. D. (2015) SubmiRine: Assessing variants in microRNA targets using clinical genomic data sets. *Nucleic Acids Res.*, 43, 3886–3898.
28. Lemieux, E., Cagnol, S., Beaudry, K., Carrier, J. and Rivard, N. (2015) Oncogenic KRAS signalling promotes the Wnt/β-catenin pathway through LRP6 in colorectal cancer. *Oncogene*, 34, 4914–4927.
29. Teo, M. T., Landi, D., Taylor, C. F., *et al.* (2012) The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes. *Carcinogenesis*, 33, 581–586.
30. Vodicka, P., Pardini, B., Vymetalkova, V. and Naccarati, A. (2016) Polymorphisms in non-coding RNA genes and their targets sites as risk factors of sporadic colorectal cancer. *Adv. Exp. Med. Biol.*, 937, 123–149.
31. Feng, Y., Zhu, J., Ou, C., Deng, Z., Chen, M., Huang, W. and Li, L. (2014) MicroRNA-145 inhibits tumour growth and metastasis in colorectal cancer by targeting fascin-1. *Br. J. Cancer*, 110, 2300–2309.
32. Ramzy, I., Hasaballah, M., Marzaban, R., Shaker, O. and Soliman, Z. A. (2015) Evaluation of microRNAs-29a, 92a and 145 in colorectal carcinoma as candidate diagnostic markers: An Egyptian pilot study. *Clin. Res. Hepatol. Gastroenterol*, 39, 508–515.
33. Gattolliat, C. H., Uguen, A., Pesson, M., Trillet, K., Simon, B., Doucet, L., Robaszkiewicz, M. and Corcos, L. (2015) MicroRNA and targeted mRNA expression profiling analysis in human colorectal adenomas and adenocarcinomas. *Eur. J. Cancer*, 51, 409–420.
34. Navon, R., Wang, H., Steinfeld, I., Tsalenko, A., Ben-Dor, A. and Yakhini, Z. (2009) Novel rank-based statistical methods reveal microRNAs with differential expression in multiple cancer types. *PLoS One*, 4, e8003.
35. Taylor, B. S., Schultz, N., Hieronymus, H., *et al.* (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell*, 18, 11–22.
36. Wan, L., Zhang, L., Fan, K. and Wang, J. (2014) MiR-27b targets LIMK1 to inhibit growth and invasion of NSCLC cells. *Mol. Cell. Biochem.*, 390, 85–91.
37. Ishihara, T., Seki, N., Inoguchi, S., *et al.* (2014) Expression of the tumor suppressive miRNA-23b/27b cluster is a good prognostic marker in clear cell renal cell carcinoma. *J. Urol.*, 192, 1822–1830.
38. Matsuyama, R., Okuzaki, D., Okada, M. and Oneyama, C. (2016) MicroRNA-27b suppresses tumor progression by regulating ARFGEF1 and focal adhesion signaling. *Cancer Sci.*, 107, 28–35.
39. Naccarati, A., Pardini, B., Stefano, L., *et al.* (2012) Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. *Carcinogenesis*, 33, 1346–1351.

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

40. Landi, D., Gemignani, F., Pardini, B., *et al*. (2012) Identification of candidate genes carrying polymorphisms associated with the risk of colorectal cancer by analyzing the colorectal mutome and microRNAome. *Cancer*, 118, 4670–4680.
41. Lu, Y., Yang, H., Yuan, L., *et al*. (2016) Overexpression of miR-335 confers cell proliferation and tumour growth to colorectal carcinoma cells. *Mol. Cell. Biochem*., 412, 235–245.
42. Mégnin-Chanet, F., Bollet, M. A. and Hall, J. (2010) Targeting poly(ADP-ribose) polymerase activity for cancer therapy. *Cell. Mol. Life Sci*., 67, 3649–3662.
43. Cheng, M., He, B., Wan, T., *et al*. (2012) 5-Fluorouracil nanoparticles inhibit hepatocellular carcinoma via activation of the p53 pathway in the orthotopic transplant mouse model. *PLoS One*, 7, e47115.
44. Parikh, N., Koshy, C., Dhayabaran, V., Perumalsamy, L. R., Sowdhamini, R. and Sarin, A. (2007) The N-terminus and alpha-5, alpha-6 helices of the pro-apoptotic protein bax, modulate functional interactions with the anti-apoptotic protein Bcl-xL. *BMC Cell Biol*., 8, 16.
45. Graziano, F., Canestrari, E., Loupakis, F., *et al*. (2010) Genetic modulation of the let-7 microRNA binding to KRAS 3'-untranslated region and survival of metastatic colorectal cancer patients treated with salvage cetuximab-irinotecan. *Pharmacogenomics J*., 10, 458–464.
46. Ruzzo, A., Canestrari, E., Galluccio, N., Santini, D., Vincenzi, B., Tonini, G., Magnani, M. and Graziano, F. (2011) Role of KRAS let-7 LCS6 SNP in metastatic colorectal cancer patients. *Ann. Oncol*., 22, 234–235.
47. Zhang, W., Labonte, M. J. and Lenz, H. J. (2011) KRAS let-7 LCS6 SNP predicts cetuximab efficacy in KRASwt metastatic colorectal cancer patients: Does treatment combination partner matter? *Ann. Oncol*., 22, 484–485.
48. Kjersem, J. B., Ikdahl, T., Guren, T., *et al*. (2012) Let-7 miRNA-binding site polymorphism in the KRAS 3'UTR; colorectal cancer screening population prevalence and influence on clinical outcome in patients with metastatic colorectal cancer treated with 5-fluorouracil and oxaliplatin +/- cetuximab. *BMC Cancer*, 12, 534.
49. Pan, X. M., Sun, R. F., Li, Z. H., Guo, X. M., Zhang, Z., Qin, H. J., Xu, G. H. and Gao, L. B. (2014) A let-7 KRAS rs712 polymorphism increases colorectal cancer risk. *Tumour Biol*., 35, 831–835.
50. Winder, T., Zhang W., El-Khoueiry, A., Yang, D., Pohl, A., Lurje, G., Rowinsky, E., Khambata-Ford, S., Langer, C., Awad, M., Lenz, H.; USC Norris Comprehensive Cancer Center, Los Angeles, CA; ImClone Systems, New York, NY; Bristol-Myers Squibb Pharmaceutical Research Institute, New York, NY (2009) *Association of a germ-line variant in the K-ras 3' untranslated region with response and progression-free survival in patients with mCRC treated with single-agent cetuximab (IMCL-0144) or in combination with cetuximab (EPIC) independent of K-ras mutation status*.
51. Langevin, S. M. and Christensen, B. C. (2014) Let-7 microRNA-binding-site polymorphism in the 3'UTR of KRAS and colorectal cancer outcome: A systematic review and meta-analysis. *Cancer Med*., 3, 1385–1395.
52. Rokavec, M., Öner, M. G., Li, H., *et al*. (2014) IL-6R/STAT3/miR-34a feedback loop promotes EMT-mediated colorectal cancer invasion and metastasis. *J. Clin. Invest*., 124, 1853–1867.
53. Raver-Shapira, N., Marciano, E., Meiri, E., Spector, Y., Rosenfeld, N., Moskovits, N., Bentwich, Z. and Oren, M. (2007) Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol. Cell*, 26, 731–743.
54. Migliore, C., Petrelli, A., Ghiso, E., Corso, S., Capparuccia, L., Eramo, A., Comoglio, P. M. and Giordano, S. (2008) MicroRNAs impair MET-mediated invasive growth. *Cancer Res*., 68, 10128–10136.
55. Navarro, F. and Lieberman, J. (2015) miR-34 and p53: New insights into a complex functional relationship. *PLoS One*, 10, e0132767.
56. Toyota, M., Suzuki, H., Sasaki, Y., Maruyama, R., Imai, K., Shinomura, Y. and Tokino, T. (2008) Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. *Cancer Res*., 68, 4123–4132.
57. Schetter, A. J., Leung, S. Y., Sohn, J. J., *et al*. (2008) MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA*, 299, 425–436.
58. Wang, L., Yu, J., Xu, J., Zheng, C., Li, X. and Du, J. (2015) The analysis of microRNA-34 family expression in human cancer studies comparing cancer tissues with corresponding pericarcinous tissues. *Gene*, 554, 1–8.
59. Smits, K. M., Paranjape, T., Nallur, S., *et al*. (2011) A let-7 microRNA SNP in the KRAS 3'UTR is prognostic in early-stage colorectal cancer. *Clin. Cancer Res*., 17, 7723–7731.
60. Ryan, B. M., Robles, A. I. and Harris, C. C. (2012) KRAS-LCS6 genotype as a prognostic marker in early-stage CRC—letter. *Clin. Cancer Res*., 18, 3487–3488; author reply 3489.
61. Keller, A., Leidinger, P., Gislefoss, R., Haugen, A., Langseth, H., Staehler, P., Lenhof, H. P. and Meese, E. (2011) Stable serum miRNA profiles as potential tool for non-invasive lung cancer diagnosis. *RNA Biol*., 8, 506–516.
62. Armstrong, D. A., Green, B. B., Seigne, J. D., Schned, A. R. and Marsit, C. J. (2015) MicroRNA molecular profiling from matched tumor and bio-fluids in bladder cancer. *Mol. Cancer*, 14, 194.

Downloaded from https://academic.oup.com/mutage/article-abstract/32/5/533/4555222
by AV CR Fyziologicky ustav SVI knihov user
on 13 November 2017

Manuscript III

Vymetalkova V, Pardini B, Rosa F, **Jiraskova K**, Di Gaetano C, Bendova P, Levy M, Veskrnova V, Buchler T, Vodickova L, Naccarati A, Vodicka P.

Polymorphisms in microRNA binding sites of mucin genes as predictors of clinical outcome in colorectal cancer patients

Carcinogenesis (2017) 38(1):28-39

Carcinogenesis, 2017, Vol. 38, No. 1, 28–39

doi:10.1093/carcin/bgw114
Advance Access publication January 2, 2017
Original Manuscript

OXFORD

ORIGINAL MANUSCRIPT

Polymorphisms in microRNA binding sites of mucin genes as predictors of clinical outcome in colorectal cancer patients

Veronika Vymetalkova[1,2,*,†], Barbara Pardini[3,†], Fabio Rosa[3], Katerina Jiraskova[1,2], Cornelia Di Gaetano[3,4], Petra Bendova[1,2,5], Miloslav Levy[6], Veronika Veskrnova[7], Tomas Buchler[7], Ludmila Vodickova[1,2,5], Alessio Naccarati[1,3] and Pavel Vodicka[1,2,5]

[1]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, 14200 Prague, Czech Republic, [2]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, 11000 Prague, Czech Republic, [3]Human Genetics Foundation, 10126 Turin, Italy, [4]Department of Medical Sciences, University of Turin, 10126 Turin, Italy, [5]Biomedical Centre, Faculty of Medicine in Pilsen, Charles University in Prague, 30605 Pilsen, Czech Republic, [6]Department of Surgery, First Faculty of Medicine, Charles University and Thomayer University Hospital, 14200 Prague, Czech Republic and [7]Department of Oncology, Thomayer Hospital and First Faculty of Medicine, Charles University, 11000 Prague, Czech Republic

*To whom correspondence should be addressed. Department of Molecular Biology of Cancer, Videnska 1083, 14200 Prague, Czech Republic. Tel: +420 2296 2251; Fax: +420 2 41062782; Email: vpolakova@biomed.cas.cz
†These authors contributed equally to this work.

Abstract

Polymorphisms in microRNA (miRNA) binding sites may affect miRNA/target gene interaction, resulting in differential mRNA/protein expression and susceptibility to common diseases. Mucins have been identified as markers of adverse prognosis. We hypothesized that genetic variations in miRNA binding sites located in mucin genes may modulate signaling response and the maintenance of genomic stability ultimately affecting cancer susceptibility, efficacy of chemotherapy and survival. In this study, we analyzed the association of single nucleotide polymorphisms in predicted miRNA target sites (miRSNPs) of mucin genes with colorectal cancer (CRC) risk and clinical outcome. Thirteen miRSNPs in 9 genes were assessed in 1111 cases and 1469 controls. No strongly significant associations were observed in the case–control study. Patients carrying the CC genotype of rs886403 in *MUC21* displayed a shorter survival and higher recurrence risk when compared with TT carriers [overall survival (OS): hazard ratios (HR) 1.69; 95% confidence intervals (CI) 1.13–2.46; $P = 0.01$ and event-free survival (EFS): HR 1.99; 95% CI 1.38–2.84; $P = 0.0002$, respectively]. The observed associations were more striking after stratification for tumor site (in patients with colon cancer, OS: HR 2.63; 95% CI 1.69–4.10; $P < 0.0001$ and EFS: HR 2.65; 95% CI 1.72–4.07; $P < 0.0001$). In contrast, rectal cancer cases carrying the CC genotype of rs4729655 in *MUC17* displayed a longer survival (OS: HR 0.27; 95% CI 0.14–0.54; $P = 0.0002$) than those with the most common genotype. To our knowledge, this is the first study investigating miRSNPs potentially affecting miRNA binding to mucin genes and revealing their impact on CRC susceptibility or patient's survival.

Introduction

One of the most common malignancies, colorectal cancer (CRC), accounts annually for almost half a million deaths worldwide. Mortality of cases results from uncontrolled metastatic disease, particularly in peritoneum, lymph nodes and liver. Tumor metastasis itself is responsible for ~90% of all CRC-related deaths (1,2). The molecular basis of CRC pathogenesis and progression is complicated and poorly understood at present.

Mucinous colorectal carcinoma is generally defined as having greater than 50% of the tumor area with a mucinous

Received: March 29, 2016; **Revised:** September 20, 2016; **Accepted:** October 25, 2016

© The Author 2017. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com.

Abbreviations

BMI	body mass index
CRC	colorectal cancer
EFS	event-free survival
HRs	hazard ratios
miRNA	microRNA
OS	overall survival
ORs	odds ratios

differentiation by histologic examination. The incidence of mucinous carcinoma is about 12% among CRC patients, ranging from 6 to 19% in different studies (3). The increased incidence occurs in proximal colon and among younger patients, as compared to nonmucinous adenocarcinoma. Mucinous CRCs have been found to have a higher Duke stage at diagnosis and, consequently, a lower survival (4).

Mucins are large extracellular glycoproteins produced by epithelial cells, which are heavily glycosylated with complex oligosaccharides (5,6). The core proteins for human mucins (MUC1-MUC8, MUC12, MUC13, MUC15-17 and MUC19-21) have already been identified (7). Many mucins are abnormally expressed and aberrantly glycosylated in adenocarcinomas. In general, increased levels of mucin genes have been associated with increased cancer risk, tumor invasion, and a poor patient outcome (8–11). Previous studies have found that an upregulation of *MUC1* and *MUC5AC* and a down-regulation of *MUC2* are involved in the development and progression of CRC (3,12,13). Recently, the overexpression of *MUC20* was considered as a predictor of poor prognosis in CRC (10). In CRC, the overexpression of transmembrane mucins suggests their role in signaling cell growth and survival. Mucins have thus been identified as markers of adverse prognosis (6).

Interestingly, recent studies have unraveled a role of microRNAs (miRNAs) also in the regulation of various mucins (14). MiRNAs are short (20–22 nucleotide) non-coding RNAs that regulate gene expression by binding mainly to the 3′ untranslated regions (3′ UTR) of target mRNA thereby hampering protein translation or inducing mRNA destabilization. Aberrant miRNA expression and/or function are frequently observed in many malignancies, including CRC (15). The increasing need for newer diagnostic strategies to target tumors has led to the appearance of miRNAs as potential cancer therapeutics of new generation. Moreover, genetic variations in the 3′ UTR of target genes may affect miRNA binding, ultimately adding additional variability in the differential mRNA and protein expressions.

We recently reported associations between single nucleotide polymorphisms (SNPs) in miRNA target regions of DNA repair pathway genes and CRC risk or clinical outcome (16–19). In the present study, we hypothesized that genetic variations in mucin genes may also affect cancer susceptibility, patient's survival and efficacy of chemotherapy. SNPs in miRNA target regions (miRSNPs) of mucin genes might affect the efficiency of translation of corresponding proteins, thus affecting individual's susceptibility to cancer. So far, the impact of miRSNPs in mucin genes on the CRC risk has not been tested yet. We have studied this assumption in patients with CRC from the Czech Republic, a country with one of the highest incidence worldwide for this cancer (20).

Material and methods

Study population and data collection

Blood samples were collected from 1111 patients with histologically confirmed CRC attending several oncological departments in the Czech Republic [Prague (3 hospitals), Benesov, Brno, Liberec, Ples, Pribram, Usti nad Labem and Zlin] between September 2003 and October 2010. 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 gastroenterological departments that had negative colonoscopy results for malignancy or idiopathic bowel diseases (Control group 1). The reasons for undergoing the colonoscopy were: (1) positive fecal occult blood test, (2) hemorrhoids, (3) abdominal pain of unknown origin and (4) macroscopic bleeding. The second group of controls consisted of 781 healthy blood donor volunteers (Control group 2) collected from a blood donor centre 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. Details of CRC cases and controls have been reported previously (18).

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 local Ethics Committee approved the design of the study. 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 CRC.

Follow up of patients

Eight hundred sixty-six CRC cases and a second group consisting of 232 CRC patients recruited later on were monitored and followed up until 31st March 2013. For all subjects, clinical data at the time of diagnosis, including location of the tumor, UICC (International Union Against Cancer) tumor-node metastasis stage system, grade and adjuvant chemotherapy treatment were collected, along with information about distant metastasis, relapse and date of death (21). In our patients dataset, a subgroup of patients ($n = 47$) was diagnosed with mucinous cancer.

Four hundred and eleven CRC cases received a 5-FU-based adjuvant regimen as first-line postoperative therapy. The therapy was based on either a Mayo regimen, delivered as a bolus infusion of 5-FU (425 mg/m^2) and leucovorin (10 mg/m^2) for five days every four weeks six times or a simplified De Gramont regimen which consisted of a 2-h intravenous (i.v.) infusion of leucovorin (200 mg/m^2), then a 5-FU i.v. bolus (400 mg/m^2) followed by a 46-h 5-FU continuous i.v. infusion (2400–3000 mg/m^2). Four hundred forty subjects did not receive any adjuvant chemotherapy after surgery.

Selection of candidate genes and SNPs in miRNA target binding sites

The list of all mucin genes was retrieved by literature mining (3,6,7). The approach used to select the candidate miRSNPs was similar to the one previously described (18). Briefly, for each gene, SNPs within predicted miRNA binding sites were identified by using the freely available software: *MicroSNiper* (http://epicenter.ie-freiburg.mpg.de/services/microsniper/), *Mirnsnpscore* (http://www.bigr.medisin.ntnu.no/mirsnpscore/) and *Polymirt* (http://compbio.uthsc.edu/miRSNP/). The 59 detected SNPs were then filtered for their minor allele frequency (MAF> 5%) in Caucasian populations (dbSNP; http://www.ncbi.nlm.nih.gov/SNP/) to reach an appropriate representation of all genotypes in our set of cases and controls. The information was primarily derived from 1000genomes project database, phase 1, CEU population; whenever this was not possible, other reference populations were considered (i.e. HAPMAP CEU population). SNPs with the required MAF were further tested for the possibility to be in *linkage disequilibrium* (LD) using HaploView (v. 4.2) with the data from HapMap (v. 3, release R2 in the CEU population).

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 selected SNPs were genotyped using the KASP™ genotyping assay, a competitive allele-specific PCR SNP genotyping system (LGC Genomics, Hoddesdon, Herts, UK). For quality control purposes, duplicate samples (5% of the total numbers of samples) were repeated for each SNP, no template controls were included in each plate (NTCs).

Statistical analyses

Pearson`s chi-square test (1 degree of freedom), with a type-I error threshold set at α = 0.05, was used to verify whether the genotypes were in Hardy–Weinberg equilibrium in control population. SNPs were excluded from further analyses when: (1) the call rate was <95%; (2) polymorphisms deviated from Hardy–Weinberg equilibrium in controls at $P < 0.01$; (3) genotypes were discrepant in more than 2% of duplicate samples. The multivariate logistic regression analysis was used to test the association between genotypes and risk of CRC. The covariates analyzed in the multivariate model were: sex, age, smoking habit (non-smokers vs. smokers and ex-smokers), BMI, familial history of CRC, education level (high, intermediate and low) and living area (country, suburbs, and town). The associations between SNPs and CRC risk were calculated by estimating the odds ratios (ORs) and their 95% confidence intervals (CI), adjusted for both continuous and discrete covariates. For all the genotypes, regression coefficients for additive models were estimated. For all SNPs, the dominant or recessive models were also calculated. The Bonferroni-corrected significance threshold is 0.004 (for 13 SNPs and α = 0.05).

The model with the highest likelihood was additionally checked for the significance of possible interaction terms in the multivariate logistic regression analysis. Statistical analyses were performed using R (http://www.rproject.org).

In this study, the outcome variables measured were overall survival (OS, time from diagnosis until death or censorship), and event-free survival (EFS, time of surgery or end of chemotherapy until date of relapse, death or censorship 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 hazard ratio (HR) using Cox regression (R version 2.14-2, Survival package). Multivariate survival analyses were adjusted for age, gender, smoking and stage.

Results

miRSNP selection

Out of the 20 genes categorized as human mucins, only 12 had miRSNPs in their 3′ UTRs (62 miRSNPs retrieved). For three of them we could not retrieve any information. The remaining 59 miRSNPs were filtered according to the study criteria previously described (i.e. MAF and LD). In total, 46 SNPs were excluded. The remaining 13 miRSNPs in nine genes (*MUC6*, *MUC7*, *MUC13*, *MUC14*, *MUC15*, *MUC17*, *MUC20, MUC21* and *MUC24*) were included in the study (Supplementary Table I, available at *Carcinogenesis* Online).

Case–control study

The characteristics of the study participants are presented in Table 1. Among the 1111 CRC cases, 397 patients were diagnosed with a tumor in colon, 334 in sigmoideum and 377 with rectal cancer (3 cases were missing the information about the site of tumor; however, since they had complete survival data, they remained in the survival analysis). Out of the 1469 controls, 688 were cancer-free colonoscopy inspected controls (Control group 1) and 781 were healthy blood donor volunteers (Control group 2). Compared to subjects of both control groups, CRC cases were more likely to be older, have a slightly higher BMI while, compared to the Control group 2, they were more likely to have a positive family history of CRC and lower formal education (Table 1).

Results of the associations between the investigated miRSNPs and CRC susceptibility are reported in Supplementary Table II, available at *Carcinogenesis* Online. A decreased cancer risk was observed for rs4071 in *MUC14* gene: carriers of the AA genotype were at decreased risk to develop either CRC or rectal cancer (OR 0.57, 95% CI 0.34–0.95, P = 0.03 and OR 0.42, 95% CI 0.18–1.00, P = 0.05, respectively). This association was confirmed in a recessive model as well (OR 0.55, 95% CI 0.34–0.91, P = 0.02 and OR 0.40, 95% CI 0.17–0.94, P = 0.04, respectively). Another SNP in *EMCN/MUC14* gene, rs17552409, was also found associated with decreased risk of CRC (for the AA genotype: OR 0.27, 95% CI 0.09–0.80, P = 0.02; for the A-allele: OR 0.27, 95% CI 0.09–0.79, P = 0.02) and colon cancer (for the AA genotype OR 0.18,

Table 1. Characteristics of the study population

		Cases (%)	Controls (%)	OR	95% CI	P value	Colon cancer cases[a]	Rectal cancer cases[a]
Age (years)	(≤47]	90 (8.1)	596 (40.1)	Ref			67 (9.2)	23 (6.1)
	(47, 55]	208 (18.7)	424 (28.5)	3.24	2.46–4.28	**<0.001**	131 (17.9)	77 (20.4)
	(55, 65]	375 (33.8)	289 (19.4)	8.59	6.56–11.25	**<0.001**	237 (32.4)	138 (36.6)
	>65	438 (39.4)	177 (12.0)	16.39	12.36–21.73	**<0.001**	296 (40.5)	139 (36.9)
Sex	Females	433 (39.0)	672 (45.2)	Ref			317 (43.4)	116 (30.8)
	Males	678 (61.0)	814 (54.8)	1.29	1.10–1.51	**0.002**	414 (56.6)	261 (69.2)
BMI	Underweight ≤18.4	12 (1.4)	5 (0.4)	Ref			8 (1.5)	4 (1.4)
	Normal weight (18.4, 24.9]	278 (33.3)	532 (38.2)	0.22	0.08–0.62	**0.003**	188 (34.2)	90 (31.6)
	Overweight (24.9, 29.9]	388 (46.5)	618 (44.3)	0.26	0.09–0.75	**0.01**	249 (45.4)	138 (48.4)
	Obese >29.9	157 (18.8)	239 (17.1)	0.27	0.09–0.79	**0.02**	104 (18.9)	53 (18.6)
Smoking	Non smokers	541 (51.9)	821 (58.2)	Ref			385 (56.0)	155 (43.8)
	Smokers	161 (15.4)	330 (23.4)	0.74	0.60–0.92	**0.006**	89 (12.9)	72 (20.3)
	Ex smokers	341 (32.7)	260 (18.4)	0.42	0.32–0.56	**<0.001**	214 (31.1)	127 (35.9)
Family history CRC	No	736 (83.4)	1215 (89.4)	Ref			468 (81.7)	267 (86.7)
	Yes	146 (16.6)	144 (10.6)	1.67	1.31–2.14	**<0.001**	105 (18.3)	41 (13.3)
Living area	Town	520 (58.3)	957 (68.0)	Ref			334 (57.3)	186 (60.4)
	Town and country	128 (14.3)	177 (12.6)	1.33	1.04–1.71	**0.03**	98 (16.8)	30 (9.7)
	Country	244 (27.4)	273 (19.4)	1.64	1.34–2.02	**<0.001**	151 (25.9)	92 (29.9)
Education	Basic	271 (30.6)	229 (16.3)	Ref			173 (30.0)	98 (31.9)
	High school	473 (53.4)	827 (58.9)	0.48	0.39–0.60	**<0.001**	302 (52.3)	170 (55.4)
	University	141 (15.9)	347 (24.7)	0.34	0.26–0.45	**<0.001**	102 (17.7)	39 (12.7)

Significant results in bold. Numbers may not add up to 100% of available subjects because of missing data.
[a]Three cases missed tumor location information.

95% CI 0.04–0.81, P = 0.02; for the variant A allele OR 0.18, 95% CI 0.04–0.80, P = 0.02). An association with CRC risk was observed for a SNP in *MUC13*. The variant AA genotype of rs1532602 was in fact associated with a decreased risk of cancer (OR 0.75, 95% CI 0.56–1.00, P = 0.05). After stratification according to the tumor site, the association was still observed in colon cancer patients (OR 0.72, 95% CI 0.51–1.00, P = 0.05), and resulted stronger for the dominant model (OR 0.79, 95% CI 0.63–0.99, P = 0.04). Finally, a decreased risk of CRC was observed in a recessive model for rs974034 in *MUC24* (OR 0.76, 95% CI 0.58–1.00, P = 0.05).

None of the above associations remained significant after applying correction for multiple testing (Bonferroni's correction).

We have also repeated the analyses in the subgroup of patients with diagnosed mucinous CRC histology (47 patients). However, in this subgroup of patients, we did not find any significant association with the CRC risk for any of the analyzed SNPs (data not shown). This last result should be cautiously considered due to the low frequency of mucinous CRC in our study group.

Survival analysis

The mean (median) OS and EFS for patients were 86.5 (80.5) and 72.6 (62.4) months, respectively. Age, gender, T, N, M status, chemotherapy treatment and CRC stage were associated with OS and EFS in the preliminary univariate assessment of covariates (Table 2). Advanced age, male gender and current smoking status were related to a shorter OS. Likewise, men were also at higher risk of relapse or metastasis (OS: HR 1.54; 95% CI 1.23–1.92; P = 0.0001; EFS: HR 1.35; 95% CI 1.09–1.68; P = 0.006). Four established prognostic factors (T, N, M status and stage) were associated with decreased patients' survival and increased risk of recurrence. Moreover, adjuvant chemotherapy was also associated with survival (Table 2).

After adjusting for sex, age, smoking and CRC stage, the strongest association with patient´s survival was observed for rs886403 in *MUC21* (Tables 3 and 4). CRC patients carrying the CC genotype displayed a shorter survival and higher recurrence risk (OS: HR 1.69; 95% CI 1.13–2.46; P = 0.01 and EFS: HR 1.99; 95% CI 1.38–2.84; P = 0.0002, respectively) when compared with carriers of the most frequent genotype. The observed associations were more striking, after stratification for tumor site, in patients with colon cancer (OS: HR 2.63; 95% CI 1.69–4.10; P < 0.0001 and EFS: HR 2.65; 95% CI 1.72–4.07; P = 0.0001 and in the recessive model OS: HR 2.70; 95% CI 1.77–4.12; P < 0.0001 and EFS: HR 2.43; 95% CI 1.61–3.64; P < 0.0001). Particularly for EFS, CRC patients also showed a similar significant trend across genotypes in the Kaplan–Meier curves (log-rank test P = 0.03; Median survival time for TT carriers=353 months; MST for CT =231 months; MST for CC carriers=79 months; Figure 1A). A similar trend was found also for colon cancer patients (log-rank test P = 0.03; MST for CC carriers = 115 months; MST not reached for the other genotypes; Figure 1B) but not for rectal cancer patients (Figure 1C).

Among colon cancer patients, carriers of the TC genotype of *MUC6* rs4077531 showed a decreased survival (OS: HR 1.38; 95% CI 1.00–1.90; P = 0.05). On the other hand, individuals with rectal cancer and carrying variant CC genotype of rs4729655 in *MUC17* displayed a longer survival when compared with the reference genotype (OS: HR 0.27; 95% CI 0.14–0.54; P = 0.0002). Overall, rectal cancer patients also showed a similar significant trend across genotypes in the Kaplan–Meier curves (log-rank

Table 2. Clinical and anamnestic characteristics significantly affecting overall survival (OS) and event free survival (EFS) of the CRC patients with complete follow up (Cox regression)

			OS	EFS	OS	
		N[a]	HR (95% CI)	*P*	HR (95% CI)	*P*
Sex	Females	427	Ref		Ref	
	Males	656	1.54 (1.23–1.92)	**0.0001**	1.35 (1.09–1.68)	**0.006**
Age (years)	≤55	293	Ref		Ref	
	56–62	248	1.43 (1.05–1.95)	**0.02**	1.41 (1.06–1.87)	**0.02**
	63–70	294	1.39 (1.04–1.88)	**0.03**	1.19 (0.90–1.58)	0.22
	>70	248	2.02 (1.50–2.72)	**<0.0001**	1.04 (0.76–1.42)	0.80
Smoking habit[b]	No	533	Ref		Ref	
	Yes	493	1.26 (1.02–1.56)	**0.03**	1.14 (0.93–1.41)	0.19
pT	1	50	Ref		Ref	
	2	166	2.64 (0.94–7.40)	0.06	2.18 (0.85–5.55)	0.10
	3	535	5.84 (2.17–15.71)	**0.0005**	5.58 (2.30–13.53)	**0.0001**
	4	136	9.21 (3.36–25.26)	**<0.0001**	6.96 (2.80–17.27)	**<0.0001**
pN	0	498	Ref		Ref	
	1	260	2.17 (1.69–2.79)	**<0.0001**	1.87 (1.46–2.41)	**<0.0001**
	2	68	3.40 (2.35–4.91)	**<0.0001**	3.43 (2.45–4.81)	**<0.0001**
pM	0	725	Ref		Ref	
	1	177	4.80 (3.83–6.02)	**<0.0001**	4.56 (3.68–5.65)	**<0.0001**
5FU-based	Yes	411	Ref		Ref	
chemotherapy	No	440	1.42 (1.13–1.790)	**0.003**	0.85 (0.68–1.06)	0.14
Stage	1	149	Ref		Ref	
	2	293	2.14 (1.32–3.48)	**0.002**	2.47 (1.51–4.05)	**0.0003**
	3	244	3.75 (2.33–6.03)	**<0.0001**	3.87 (2.38–6.31)	**<0.0001**
	4	177	11.87 (7.44–18.95)	**<0.0001**	11.86 (7.42–18.98)	**<0.0001**

Significant results in bold.

[a]Numbers may not add up to 100% of available subjects because of missing information.

[b]Ex-smokers included in non-smokers.

Table 3. miRSNPs associated with OS of patients divided for cancer sites and in the pooled population (Cox regression for adjusted estimates)

		All patients					Rectal cancer patients					Colon cancer patients				
Gene dbSNP ID	Genotype	N[a]	Events	Expected	HR (95% CI)[b]	P	N[a]	Events	Expected	HR (95% CI)[b]	P	N[a]	Events	Expected	HR (95% CI)[b]	P
MUC6	TT	316	115	122.5	Ref		101	41	43.8	Ref		215	74	78.0	Ref	
rs4077531	TC	437	169	160.8	1.27 (0.98–1.65)	0.07	145	62	59.5	1.00 (0.63–1.58)	0.99	292	107	101.6	1.38 (1.00–1.90)	**0.05**
	CC	168	63	63.7	1.25 (0.90–1.74)	0.18	62	27	26.6	1.21 (0.71–2.07)	0.49	105	36	37.4	1.19 (0.78–1.81)	0.42
	TC+CC	605	232	224.5	1.27 (0.99–1.62)	0.06	207	89	86.1	1.06 (0.69–1.63)	0.79	397	143	139.0	1.33 (0.98–1.80)	0.07
	TT+TC	753	284	283.3	Ref		246	103	103.3	Ref		207	181	179.6	Ref	
	CC	168	63	63.7	1.08 (0.81–1.45)	0.58	62	27	26.6	1.21 (0.77–1.92)	0.41	105	36	37.4	0.99 (0.68–1.44)	0.95
MUC7	AA	608	238	225.0	Ref		199	87	81.2	Ref		408	151	143.9	Ref	
rs3733492	AG	281	95	104.3	0.95 (0.74–1.22)	0.71	93	33	39.0	0.91 (0.59–1.40)	0.66	188	62	65.5	0.94 (0.69–1.28)	0.69
	GG	25	6	9.7	0.43 (0.18–1.05)	0.07	9	3	2.8	0.77 (0.23–2.50)	0.66	16	3	6.6	0.26 (0.06–1.06)	0.06
	AG + GG	306	101	114.0	0.90 (0.70–1.15)	0.38	102	36	41.8	0.89 (0.59–1.36)	0.60	204	65	72.1	0.87 (0.64–1.18)	0.36
	AA + AG	889	333	329.3	Ref		292	120	120.2	Ref		596	213	209.4	Ref	
	GG	25	6	9.7	0.44 (0.18–1.07)	0.07	9	3	2.8	0.79 (0.24–2.56)	0.69	16	3	6.6	0.27 (0.07–1.08)	0.06
MUC13	TT	598	227	222.3	Ref		194	86	81.9	Ref		404	141	139.7	Ref	
rs12732	TC	304	111	117.2	0.91 (0.71–1.17)	0.45	113	45	46.8	0.84 (0.56–1.26)	0.40	190	66	71.3	0.94 (0.68–1.29)	0.70
	CC	37	14	12.5	0.92 (0.53–1.58)	0.76	9	2	4.3	0.42 (0.10–1.72)	0.23	28	12	8.1	1.13 (0.62–2.06)	0.70
	TC+CC	341	124	129.7	0.91 (0.72–1.15)	0.44	122	47	51.1	0.80 (0.53–1.19)	0.26	218	78	79.4	0.97 (0.72–1.30)	0.83
	TT+TC	902	338	339.5	Ref		307	131	128.7	Ref	0.26	594	205	211.0	Ref	
	CC	37	14	12.5	0.95 (0.55–1.62)	0.84	9	2	4.3	0.44 (0.11–1.81)		28	12	8.1	1.15 (0.63–2.08)	0.65
MUC13	GG	337	117	128.4	Ref		109	49	46.6	Ref		228	68	81.7	Ref	
rs1532602	GA	448	177	167.8	1.20 (0.93–1.55)	0.16	152	62	64.6	0.97 (0.63–1.48)	0.87	295	115	103.4	1.33 (0.97–1.84)	0.08
	AA	139	53	50.8	1.0 (0.73–1.47)	0.84	45	18	17.9	0.87 (0.48–1.56)	0.63	93	35	32.8	1.15 (0.74–1.79)	0.53
	GA + AA	587	230	218.6	1.16 (0.91–1.47)	0.24	197	80	82.5	0.94 (0.63–1.41)	0.77	388	150	136.2	1.29 (0.95–1.75)	0.10
	GG + GA	785	294	296.2	Ref		261	111	111.2	Ref		523	183	185.1	Ref	
	AA	139	53	50.8	0.93 (0.68–1.27)	0.65	45	18	17.9	0.89 (0.52–1.51)	0.65	93	35	32.8	0.97 (0.66–1.43)	0.87
MUC14/ EMCN	GG	563	212	209.7	Ref		186	81	81.8	Ref		376	131	128.3	Ref	
rs4071	GA	338	123	123.0	1.07 (0.85–1.36)	0.57	119	48	48.4	0.82 (0.55–1.22)	0.32	219	75	75.1	1.15 (0.85–1.55)	0.36
	AA	28	9	11.2	0.64 (0.30–1.36)	0.25	6	3	1.8	1.00 (0.14–7.31)	1.00	22	6	8.7	0.65 (0.28–1.47)	0.30
	GA+AA	366	132	134.2	1.03 (0.82–1.30)	0.81	125	51	50.2	0.82 (0.55–1.22)	0.32	241	81	83.8	1.08 (0.81–1.45)	0.60
	GG+GA	901	335	332.7	Ref		305	129	130.2	Ref		595	206	203.4	Ref	
	AA	28	9	11.2	0.62 (0.29–1.32)	0.22	6	3	1.8	1.07 (0.15–7.81)	0.95	22	6	8.7	0.62 (0.27–1.39)	0.24
MUC14/ EMCN	GG	743	278	274.9	Ref		246	107	103.4	Ref		496	171	171.6	Ref	
rs17552409	GA	185	66	68.5	1.00 (0.75–1.34)	0.99	61	22	24.3	0.79 (0.46–1.34)	0.38	124	44	43.9	1.11 (0.78–1.56)	0.57
	AA	4	1	1.5	0.87 (0.12–6.24)	0.89	2	0	1.2	—	—	2	1	0.5	1.68 (0.23–12.1)	0.61
	GA+AA	189	67	70	1.00 (0.75–1.33)	1.00	63	22	25.5	0.76 (0.45–1.30)	0.32	126	45	44.4	1.12 (0.79–1.57)	0.53
	GG+GA	928	344	343.4	Ref		307	129	127.7	Ref		620	215	215.5	Ref	
	AA	4	1	1.5	0.87 (0.12–6.23)	0.89	2	0	1.2	—	—	2	1	0.5	1.65 (0.23–11.9)	0.62
MUC15	AA	552	206	205.8	Ref		187	76	79.2	Ref		364	130	126.8	Ref	
rs17243454	AG	356	132	133.7	0.89 (0.70–1.12)	0.32	117	53	50.4	1.06 (0.71–1.58)	0.77	239	79	83.3	0.82 (0.61–1.10)	0.18
	GG	36	15	13.5	0.96 (0.55–1.66)	0.88	11	5	4.5	1.24 (0.45–3.43)	0.64	25	10	8.9	0.85 (0.44–1.64)	0.63
	AG+GG	392	147	147.2	0.89 (0.71–1.12)	0.34	128	58	54.9	1.08 (0.73–1.58)	0.71	264	89	92.2	0.82 (0.62–1.09)	0.17
	AA+AG	908	338	339.5	Ref		304	129	129.6	Ref		603	209	210.1	Ref	
	GG	36	15	13.5	1.00 (0.59–1.73)	0.99	11	5	4.5	1.22 (0.45–3.33)	0.70	25	10	8.9	0.92 (0.49–1.76)	0.81

Table 3. *Continued*

Gene dbSNP ID	Genotype	All patients					Rectal cancer patients					Colon cancer patients				
		N[a]	Events	Expected	HR (95% CI)[b]	P	N[a]	Events	Expected	HR (95% CI)[b]	P	N[a]	Events	Expected	HR (95% CI)[b]	P
MUC17	TT	252	107	92.8	Ref		83	44	29.0	Ref		169	63	62.4	Ref	
rs4729655	TC	459	167	170.8	1.02 (0.78–1.33)	0.89	157	65	68.7	0.72 (0.47–1.12)	0.15	301	102	102.8	1.14 (0.82–1.60)	0.44
	CC	215	69	79.4	0.74 (0.53–1.03)	0.08	70	20	31.3	0.27 (0.14–0.54)	**0.0002**	145	49	48.8	1.06 (0.72–1.58)	0.76
	TC+CC	674	236	250.2	0.92 (0.72–1.18)	0.53	227	85	100	0.57 (0.37–0.86)	**0.008**	446	151	165.2	1.12 (0.81–1.53)	0.58
	TT+TC	711	274	263.6	Ref		240	109	97.7	Ref		470	165	151.6	Ref	
	CC	215	69	79.4	0.73 (0.55–0.98)	**0.03**	70	20	31.3	0.33 (0.18–0.63)	**0.0007**	145	49	48.8	0.98 (0.70–1.37)	0.91
MUC20	AA	912	345	341.1	Ref		306	132	130.1	Ref		606	213	211.5	Ref	
rs7631009	AG	32	9	12.9	0.63 (0.33–1.23)	0.18	10	3	4.9	0.44 (0.14–1.41)	0.17	21	6	7.5	0.74 (0.33–1.68)	0.48
	GG	0	0	0	—	—	0	0	0	—	—	0	0	0	—	—
	AG+GG	32	9	12.9	0.63 (0.33–1.23)	0.18	10	3	4.9	0.44 (0.14–1.41)	0.17	21	6	7.5	0.74 (0.33–1.68)	0.48
	AA+AG	944	354	354.0	Ref		316	135	135	Ref		627	219	219	Ref	
	GG	0	0	0	—	—	0	0	0	—	—	0	0	0	—	—
MUC20	AA	601	217	227.9	Ref		197	81	83.8	Ref		403	136	143.7	Ref	
rs6782006	AG	294	112	108.8	103 (0.81–1.31)	0.82	95	41	41.1	1.18 (0.78–1.79)	0.44	199	71	67.9	0.99 (0.73–1.34)	0.94
	GG	41	21	13.4	1.80 (1.10–2.97)	**0.02**	18	9	6.2	1.60 (0.73–3.51)	0.24	23	12	7.5	1.90 (1.00–3.69)	**0.05**
	AG+GG	335	133	122.2	1.10 (0.87–1.39)	0.42	113	50	47.3	1.23 (0.83–1.83)	0.30	222	83	75.4	1.06 (0.79–1.42)	0.69
	AA+AG	895	329	336.7	Ref		292	122	124.9	Ref		602	207	211.6	Ref	
	GG	41	21	13.4	1.79 (1.09–2.92)	**0.02**	18	9	6.2	1.51 (0.70–3.28)	0.30	23	12	7.5	1.91 (1.01–3.63)	**0.05**
MUC21	TT	479	185	181.1	Ref		175	79	71.0	Ref		303	106	110.6	Ref	
rs886403	TC	373	129	141.1	0.86 (0.67–1.09)	0.21	116	47	52.6	0.73 (0.49–1.09)	0.12	257	82	88.2	0.94 (0.69–1.28)	0.71
	CC	80	36	27.8	1.69 (1.13–2.46)	**0.01**	24	8	10.4	0.53 (0.21–1.31)	0.17	56	28	17.2	2.63 (1.69–4.10)	**<0.0001**
	TC+CC	453	165	168.9	0.96 (0.76–1.20)	0.70	140	55	63.0	0.70 (0.47–1.03)	0.07	313	110	105.4	1.13 (0.85–1.50)	0.41
	TT+TC	852	314	322.2	Ref		291	126	123.6	Ref		560	188	198.8	Ref	
	CC	80	36	27.8	1.78 (1.23–2.60)	**0.003**	24	8	10.4	0.60 (0.25–1.48)	0.27	56	28	17.2	2.70 (1.77–4.12)	**<0.0001**
MUC24/ CD164	TT	339	137	120.6	Ref		118	49	45.8	Ref		221	88	75.0	Ref	
rs974034	TC	457	153	17.3	0.86 (0.67–1.11)	0.24	158	63	69.0	1.01 (0.67–1.50)	0.96	298	90	108.5	0.79 (0.58–1.09)	0.15
	CC	125	52	44.1	1.14 (0.81–1.60)	0.45	34	17	14.2	1.30 (0.71–2.41)	0.40	91	35	29.5	1.06 (0.70–1.60)	0.79
	TC+CC	583	205	61.4	0.92 (0.73–1.16)	0.49	192	80	83.2	1.06 (0.72–1.58)	0.77	389	125	138	0.85 (0.63–1.14)	0.28
	TT+TC	796	290	137.9	Ref		276	112	114.8	Ref		519	178	183.5	Ref	
	CC	125	52	44.1	1.24 (0.91–1.69)	0.18	34	17	14.2	1.30 (0.73–2.29)	0.37	91	35	29.5	1.21 (0.83–1.76)	0.33
MUC24/ CD164	TT	449	172	164.6	Ref		153	67	59.9	Ref		295	105	103.8	Ref	
rs7372	TC	408	144	156.5	0.95 (0.75–1.21)	0.69	134	52	61.1	0.85 (0.57–1.27)	0.43	274	92	96.0	1.33 (0.76–1.38)	0.87
	CC	71	30	24.9	0.95 (0.63–1.43)	0.80	24	12	10.0	0.98 (0.49–1.97)	0.95	47	18	15.2	0.90 (0.53–1.51)	0.68
	TC+CC	479	174	181.4	0.95 (0.76–1.19)	0.67	158	64	71.1	0.87 (0.59–1.28)	0.48	321	110	111.2	1.00 (0.75–1.33)	0.99
	TT+TC	857	316	321.1	Ref		287	119	121.0	Ref		569	197	199.8	Ref	
	CC	71	30	24.9	0.97 (0.65–1.45)	0.88	24	12	10.0	1.05 (0.53–2.08)	0.89	47	18	15.2	0.89 (0.54–1.47)	0.64

Significant results in bold.

[a]Numbers may not add up to 100% of available subjects because of missing information.

[b]Adjusted for sex, age, smoking, and stage.

Table 4. miRSNPs associated with EFS of patients divided for cancer sites and in the pooled population (Cox regression for adjusted estimates)

Gene		All patients					Rectal cancer patients					Colon cancer patients				
dbSNP ID	Genotype	N[a]	Events	Expected	HR (95% CI)[b]	P	N[a]	Events	Expected	HR 95% CI)[b]	P	N[a]	Events	Expected	HR (95% CI)[b]	P
MUC6	TT	312	121	126.6	Ref		99	35	39.7	Ref		213	86	86.6	Ref	
rs4077531	TC	423	170	162.7	1.23 (0.96–1.57)	0.10	139	59	51.0	1.27 (0.80–2.01)	0.31	284	111	110.9	1.22 (0.91–1.64)	0.19
	CC	160	62	63.7	1.12 (0.81–1.54)	0.51	57	20	23.3	1.19 (0.67–2.11)	0.56	102	41	40.5	1.04 (0.70–1.54)	0.84
	TC + CC	583	232	226.4	1.20 (0.95–1.51)	0.13	196	79	74.3	1.24 (0.81–1.92)	0.33	386	152	151.4	1.17 (0.88–1.54)	0.27
	TT + TC	735	291	289.3	Ref		238	94	90.7	Ref		497	197	197.5	Ref	
	CC	160	62	63.7	0.99 (0.74–1.32)	0.94	57	20	23.3	1.04 (0.63–1.71)	0.89	102	41	40.5	0.93 (0.65–1.33)	0.70
MUC7	AA	591	220	226.3	Ref		190	64	72.0	Ref		400	155	153.7	Ref	
rs3733492	AG	274	113	106.3	1.08 (0.86–1.37)	0.51	91	40	33.3	1.33 (0.87–2.03)	0.19	183	73	72.9	0.97 (0.73–1.30)	0.85
	GG	24	9	9.2	0.88 (0.45–1.72)	0.71	8	4	2.7	1.01 (0.35–2.92)	0.98	16	5	6.4	0.75 (0.31–1.83)	0.53
	AG + GG	298	122	115.5	1.03 (0.85–1.34)	0.60	99	44	36	1.29 (0.86–1.95)	0.22	199	78	79.3	0.95 (0.72–1.26)	0.74
	AA + AG	865	333	332.6	Ref		281	104	105.3	Ref		583	228	226.6	Ref	
	GG	24	9	9.2	0.86 (0.44–1.67)	0.65	8	4	2.7	0.92 (0.32–2.61)	0.87	16	5	6.4	0.76 (0.31–1.84)	0.54
MUC13	TT	582	227	232.5	Ref		186	68	73.0	Ref		396	159	158.9	Ref	
rs12732	TC	295	118	115.2	1.06 (0.84–1.33)	0.65	108	45	40.8	1.09 (0.73–1.62)	0.68	186	72	74.1	1.01 (0.75–1.36)	0.93
	CC	37	16	13.3	1.00 (0.59–1.70)	1.00	9	4	3.2	1.09 (0.34–3.55)	0.89	28	12	10.0	0.96 (0.53–1.73)	0.89
	TC + CC	332	134	128.5	1.05 (0.84–1.31)	0.68	117	49	44	1.09 (074–1.61)	0.68	214	84	84.1	1.00 (0.76–1.33)	0.98
	TT + TC	877	345	347.7	Ref		294	113	113.8	Ref		582	231	233	Ref	
	CC	37	16	13.3	0.98 (0.58–1.66)	0.95	9	4	3.2	1.06 (0.33–3.40)	0.93	28	12	10.0	0.95 (0.53–1.72)	0.87
MUC13	GG	327	120	129.9	Ref		103	38	39.2	Ref		224	82	89.9	Ref	
rs1532602	GA	436	173	168.7	1.02 (0.80–1.30)	0.90	147	52	57.1	0.84 (0.54–1.31)	0.45	288	120	111.5	1.08 (0.80–1.44)	0.63
	AA	136	58	52.4	0.99 (0.71–1.37)	0.93	44	22	15.7	1.01 (0.58–1.75)	0.98	92	36	36.6	0.92 (0.60–1.39)	0.68
	GA + AA	572	231	221.1	1.01 (0.80–1.27)	0.95	191	74	72.8	0.89 (059–1.34)	0.57	380	156	148.1	1.04 (0.78–1.37)	0.80
	GG + GA	763	283	298.6	Ref		250	90	96.3	Ref		512	202	201.4	Ref	
	AA	136	58	52.4	0.98 (0.7–1.31)	0.87	44	22	15.7	1.11 (0.68–1.82)	0.66	92	36	36.6	0.88 (0.60–1.28)	0.50
MUC14/ EMCN	GG	550	216	213.5	Ref		178	73	68.1	Ref		371	142	144.9	Ref	
rs4071	GA	326	124	124.0	1.08 (0.85–1.36)	0.54	115	39	43.2	0.82 (0.54–1.26)	0.37	211	85	80.6	1.16 (0.88–1.54)	0.30
	AA	27	9	11.4	0.86 (0.44–1.67)	0.65	5	1	1.7	0.82 (0.11–6.04)	0.85	22	8	9.6	0.89 (0.43–1.81)	0.74
	GA + AA	353	133	135.4	1.06 (0.84–1.32)	0.64	120	40	44.9	0.82 (0.54–1.25)	0.36	233	93	90.2	1.13 (0.86–1.48)	0.39
	GG + GA	876	340	337.5	Ref		293	112	111.3	Ref		582	227	225.5	Ref	
	AA	27	9	11.4	0.84 (0.43–1.62)	0.60	5	1	1.7	0.88 (0.12–6.44)	0.90	22	8	9.6	0.84 (0.42–1.71)	0.63
MUC14/ EMCN	GG	728	288	282.2	Ref		238	92	89.2	Ref		489	195	191.8	Ref	
rs17552409	GA	176	64	70.3	0.97 (0.73-1.27)	0.80	58	19	21.8	0.90 (0.53—1.52)	0.69	118	45	48.4	0.97 (0.70-1.35)	0.85
	AA	4	2	1.7	1.22 (0.30–4.90)	0.79	2	1	1.0	1.57 (0.21–11.6)	0.66	2	1	0.8	0.97 (0.13–6.97)	0.97
	GA + AA	180	66	72	0.97 (0.74–1.28)	0.84	60	20	22.8	0.92 (0.55–1.54)	0.76	120	46	49.2	0.97 (0.70–1.34)	0.85
	GG + GA	904	352	352.5	Ref		296	111	111	Ref		607	240	240.2	Ref	
	AA	4	2	1.7	1.22 (0.30–4.93)	0.78	2	1	1	1.60 (0.22–11.8)	0.65	2	1	0.8	0.97 (0.14–7.00)	0.98

Table 4. *Continued*

Gene dbSNP ID	Genotype	All patients N[a]	Events	Expected	HR (95% CI)[b]	P	Rectal cancer patients N[a]	Events	Expected	HR 95% CI)[b]	P	Colon cancer patients N[a]	Events	Expected	HR (95% CI)[b]	P
MUC15	AA	541	216	211.0	Ref		180	74	69.4	Ref		360	141	141.2	Ref	
rs17243454	AG	342	130	133.4	0.95 (0.75–1.19)	0.63	112	40	43.5	0.92 (0.61–1.38)	0.67	230	90	89.3	0.99 (0.75–1.30)	0.93
	GG	35	13	14.6	0.73 (0.41–1.31)	0.29	10	3	4.1	0.95 (0.30–3.02)	0.92	25	10	10.5	0.90 (0.36–1.38)	0.31
	AG + GG	377	143	148	0.92 (0.74–1.15)	0.46	122	43	47.6	0.92 (0.62–1.37)	0.67	252	100	99.8	0.95 (0.73–1.24)	0.71
	AA + AG	883	346	344.4	Ref		292	114	112.9	Ref		590	231	230.5	Ref	
	GG	35	13	14.6	0.75 (0.42–1.33)	0.32	10	3	4.1	0.98 (0.31–3.09)	0.97	25	10	10.5	0.71 (0.36–1.38)	0.31
MUC17	TT	245	101	94.2	Ref		80	33	27.0	Ref		165	68	66.6	Ref	
rs4729655	TC	448	178	174.1	1.02 (0.79–1.33)	0.86	152	60	58.0	0.95 (0.60–1.52)	0.84	295	114	115.7	1.05 (0.77–1.44)	0.75
	CC	207	72	79.7	0.97 (0.71–1.33)	0.86	65	18	26.1	0.72 (0.39–1.34)	0.30	142	54	53.7	1.09 (0.76–1.59)	0.63
	TC + CC	655	250	253.8	101 (0.79–1.29)	0.95	217	78	84.1	0.89 (0.57–1.39)	0.59	437	168	169.4	1.07 (0.79–1.43)	0.67
	TT + TC	693	279	268.3	Ref		232	93	85	Ref		460	182	182.3	Ref	
	CC	207	72	79.7	0.96 (0.73–1.25)	0.75	65	18	26.1	0.74 (0.43–1.27)	0.28	142	54	53.7	1.06 (0.77–1.45)	0.72
MUC20	AA	887	344	344.7	Ref		293	113	114.0	Ref		594	231	232.5	Ref	
rs7631009	AG	31	15	11.6	1.18 (0.69–2.03)	0.55	10	5	4.0	1.01 (0.41–2.50)	0.98	20	9	7.5	1.23 (0.63-2.42)	0.54
	GG	0	0	0	—	—	0	0	0	—	—	0	0	0	—	—
	AG + GG	31	15	11.6	1.18 (0.69–2.03)	0.55	10	5	4.0	1.01 (0.41–2.50)	0.98	20	9	7.5	1.23 (0.63–2.42	0.54
MUC20	AA	585	219	231.3	Ref		188	69	72.6	Ref		396	149	158.2	Ref	
rs6782006	AG	285	121	112.0	1.04 (0.83–1.31)	0.73	92	40	36.7	0.98 (0.65–1.48)	0.92	193	81	75.0	1.08 (0.82–1.43)	0.59
	GG	40	17	13.8	1.45 (0.84–2.50)	0.18	17	6	5.7	1.21 (0.48-3.05)	0.95	23	11	7.9	1.58 (0.80–3.11)	0.19
	AG + GG	325	138	125.8	1.08 (0.86–1.34)	0.52	109	46	42.4	1.00 (0.67–1.49)	0.99	216	92	82.9	1.12 (0.85–1.47)	0.42
	AA + AG	870	340	343.3	Ref		280	109	109.3	Ref		589	230	233.2	Ref	
	GG	40	17	13.8	1.43 (0.83–2.45)	0.19	17	6	5.7	1.22 (0.49–3.04)	0.67	23	11	7.9	1.54 (0.78–3.02)	0.21
MUC21	TT	464	169	184.6	Ref		168	62	64.0	Ref		295	106	120.2	Ref	
rs886403	TC	366	145	142.1	1.10 (0.87–1.39)	0.43	112	44	44.0	1.01 (0.67–1.51)	0.97	254	101	97.5	1.20 (0.90–1.59)	0.22
	CC	76	40	27.3	1.99 (1.38–2.84)	**0.0002**	22	11	9.0	1.05 (0.52–2.14)	0.89	54	29	18.2	2.65 (1.72–4.07)	**0.00001**
	TC + CC	442	185	169.4	1.21 (0.98–1.51)	0.08	134	55	53.0	1.02 (0.69–1.49)	0.94	308	130	115.7	1.36 (1.04–1.78)	**0.03**
	TT + TC	830	314	326.7	Ref		280	106	108.0	Ref		549	207	217.7	Ref	
	CC	76	40	27.3	1.91 (1.35–2.70)	**0.0003**	22	11	9.0	1.05 (0.53–2.09)	0.89	54	29	18.2	2.43 (1.61–3.64)	**0.00002**
MUC24/ CD164	TT	327	129	125.1	Ref		115	45	42.7	Ref		212	84	82.2	Ref	
rs974034	TC	451	171	179.6	1.01 (0.79–1.28)	0.95	152	56	59.2	1.06 (0.70–1.62)	0.77	298	114	119.9	0.98 (0.73–1.32)	0.89
	CC	119	49	44.3	1.12 (0.79–1.57)	0.53	31	12	11.1	1.46 (0.74–2.88)	0.27	88	37	32.8	1.05 (0.70–1.57)	0.80
	TC + CC	570	220	183.9	1.03 (0.82–1.29)	0.80	183	68	70.3	1.12 (0.75–1.67)	0.59	386	151	152.7	1.00 (0.75–1.32)	0.98
	TT + TC	778	300	304.7	Ref		267	101	101.9	Ref		510	198	202.1	Ref	
	CC	119	49	44.3	1.11 (0.81–1.52)	0.51	31	12	11.1	1.41 (0.75–2.67)	0.29	88	37	32.8	1.07 (0.74–1.53)	0.73

Table 4. *Continued*

Gene dbSNP ID	Genotype	All patients N[a]	Events	Expected	HR (95% CI)[b]	P	Rectal cancer patients N[a]	Events	Expected	HR 95% CI)[b]	P	Colon cancer patients N[a]	Events	Expected	HR (95% CI)[b]	P
MUC24/ CD164	TT	433	177	164.6	Ref		146	62	52.1	Ref		286	114	111.7	Ref	
rs7372	TC	406	149	163.2	0.89 (0.71–1.12)	0.31	132	45	55.3	0.70 (0.46–1.06)	0.09	274	104	107.8	0.97 (0.73–1.28)	0.83
	CC	65	27	25.1	0.83 (0.55–1.26)	0.38	21	8	7.6	0.92 (0.42–2.04)	0.84	44	19	17.4	0.82 (0.50–1.35)	0.44
	TC + CC	471	176	188.3	0.88 (0.71–1.09)	0.25	153	53	62.9	0.73 (0.49–1.08)	0.11	318	123	125.5	0.94 (0.72–1.23)	0.67
	TT + TC	839	326	237.8	Ref		278	107	107.4	Ref		560	218	125.2	Ref	
	CC	65	27	25.1	0.88 (0.59–1.32)	0.53	21	8	7.6	1.08 (0.50–2.34)	0.84	44	19	17.4	0.84 (0.52–1.35)	0.46

Significant results in bold.
[a]Numbers may not add up to 100% of available subjects because of missing information.
[b]Adjusted for sex, age, smoking, and stage.

test P = 0.002; MST for TT carriers = 67 months; MST for TC carriers=152; MST for CC carriers=176; Figure 2C). For the same variant genotype, CRC patients showed also better survival in a recessive model (OS: HR 0.73; 95% CI 0.55–0.98; $P = 0.03$). Conversely, no significant changes in survival were observed in CRC patients and in colon cancer patients in the univariate model (Figure 2A and B).

Finally, CRC and colon cancer patients carrying variant GG genotype of rs6782006 in *MUC20* showed worse survival (OS: HR 1.80; 95% CI 1.10–2.97; $P = 0.02$ and HR 1.90; 95% CI 1.00–3.69; $P = 0.05$, respectively).

Discussion

Mucins are high molecular weight glycoproteins predominantly expressed at the epithelial surface of tissues that provide protection for colon surface under normal physiological conditions (22). Several lines of evidence point towards a biological role of mucins in CRC (22,23). These include: observations on mucinous CRC in *in vitro* and *in vivo* experiments and alterations of mucin structure in polyps and cancer. Although mucinous CRC have a higher Dukes stage at diagnosis and show a worse prognosis, the cause of such negative phenotypes is not comprehensively elucidated. Mucins have an important role in epithelial cell protection and maintenance of homeostasis by covering human colon surface by gel mucous layer. Lower cell adhesion may contribute to incorrect cellular organization and structure, proliferation and survival, and ultimately gene expression alteration. The appropriate cell adhesion is necessary for numerous physiological processes and can be deranged in many diseases, including thrombosis, inflammation, and cancer (24). Recently, miRNAs have emerged as important regulators responsible for an altered mucin expression during the malignant development (14). To date, no study has investigated the impact of miRSNPs within mucin genes on CRC susceptibility or clinical outcome. In the present study, we have found that some miRSNPs in mucin genes were associated with CRC risk and they had an impact on patient´s survival. Interestingly, a reduced CRC risk was observed for individuals bearing homozygous variant genotypes of *MUC13* rs1532602, *EMCN/MUC14* rs4071 and rs17552409 and *MUC24* rs974034. After stratification for tumor localization, some of these results were confirmed in colon cancer patients (rs1532602 in *MUC13*, in a dominant model) or rectal cancer patients (rs4071 in *EMCN/MUC14* in recessive model). After applying correction for multiple testing, these associations were lost. On the other hand, such conservative correction may not be required considering: (1) the exploratory nature of our study, (2) the fact that all the SNPs were selected for their high prior probability of functional significance, and (3) based on differential binding of miRNAs to their predicted polymorphic target sites. Both *EMCN/MUC14* rs4071 and rs17552409 polymorphisms were also tested for their potential SNP-SNP interaction on CRC susceptibility. No effect on CRC risk was observed when these two SNP were analyzed together. Similarly, we have also repeated the analyses in the subgroup of patients with diagnosed mucinous CRC histology. However, due to the low representation of mucinous CRC in this subgroup of patients (47 patients), we did not find any significant association with the CRC risk for any of the analyzed SNPs. Mucinous CRC tends to occur in younger patients, are often seen in the proximal colon, are more diagnosed at an advanced stage and are more frequently associated with hereditary non-polyposis colorectal cancer (HNPCC) and young-age sporadic colorectal cancer. The low proportion of mucinous CRC patients in the present study could be inflicted

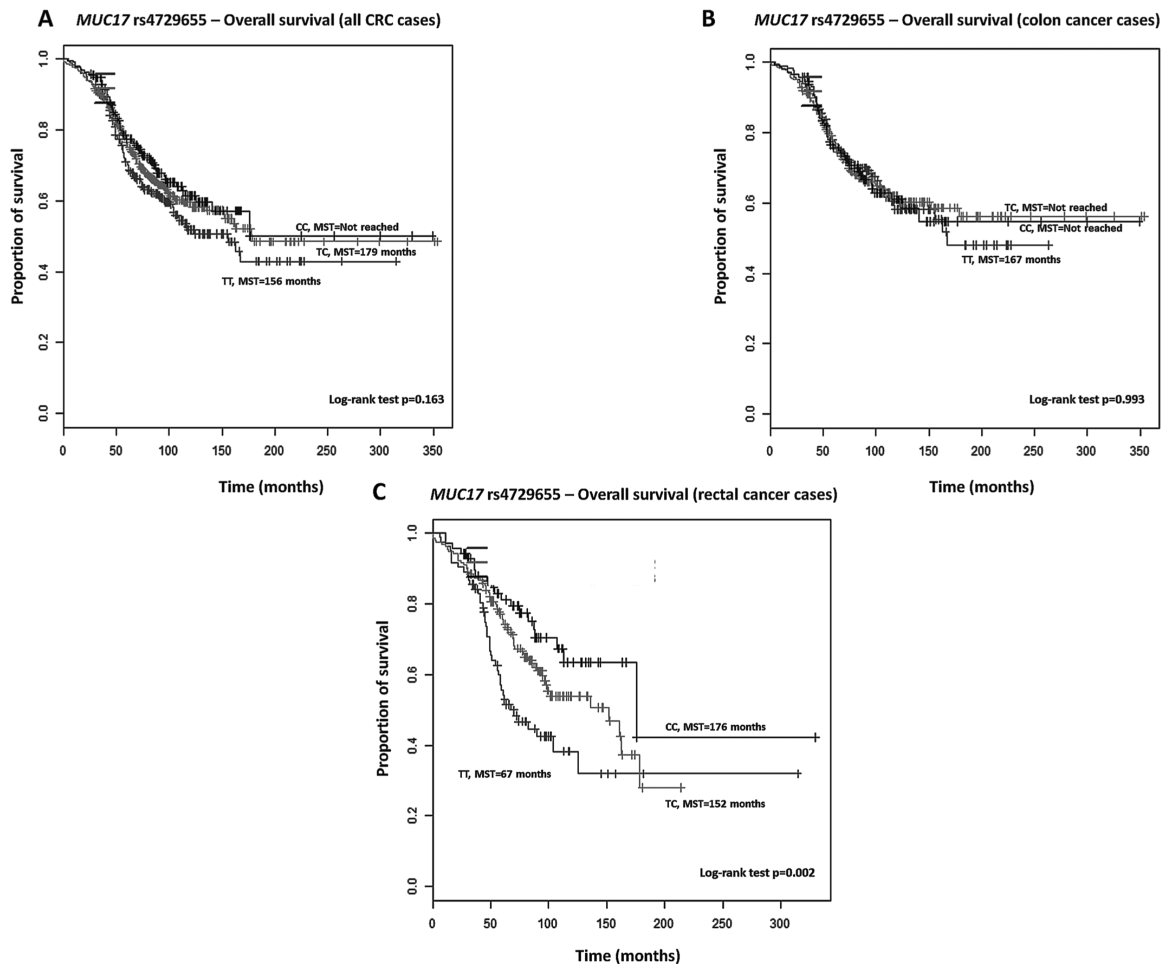


Figure 1. Kaplan–Meier EFS curves stratified for rs886403 in *MUC21* gene in (**A**) all CRC patients, (**B**) colon cancer patients and C) rectal cancer patients.

by the fact that CRC cases comprise rather older patients (age range 18–47 comprises 86 CRC patients while age range 65–91 comprises 427 CRC patients).

The stronger and novel finding of this study is represented by the associations of some miRSNPs with clinical outcome. In particular, CRC patients carrying the CC genotype for rs886403 in *MUC21* displayed a shorter survival and higher recurrence risk. The observed association was strikingly pronounced in colon cancer patients. *MUC21* has been identified quite recently (25). The gene encodes for a transmembrane mucin related to the biosynthesis of N-glycan precursor (dolichol lipid-linked oligosaccharide, LLO). According to STRING Interaction Network (www.genecards.org), there is a close cooperation among *MUC21* and *MUC17, MUC20, MUC16, GALNT5* and *ST3GAL3*. Scarce information is available for this gene, thus the postulation of hypothesis that this gene might be associated with patient´s survival is quite difficult. Cells transfected by MUC21 were significantly less adherent to each other and to extracellular matrix components than control cells, suggesting that MUC21 prevents integrin-mediated cell adhesion to extracellular matrix components (26). Cell adhesion and cell-cell interactions also play vital roles in many later steps in cancer progression, facilitating the entry and survival of cancer cells into the bloodstream, and their arrest and establishment at distant organs (24). The miRNA-mediated regulation of *MUC21* expression has not been investigated yet. Several miRNAs (such as miR-4647, miR-588, miR-125 and let-7) are predicted to bind in the region surrounding rs886403. However, none of them have been validated in relation to *MUC21,* so far.

In this study, rectal cancer patients carrying the CC genotype of *MUC17* rs4729655 displayed a better OS when compared with the carriers of the other genotypes. MUC17, an intestinal membrane-bound mucin, has been shown to enhance mucosal restitution by stimulation of cell migration and inhibition of apoptosis (27). MUC17 is highly expressed on the surface epithelium of normal colonic mucosa but its expression becomes altered in colorectal neoplasia. Interestingly, an increased expression of MUC17 was associated with a longer OS in patients with stage III and IV colorectal adenocarcinomas (28). These results point to its possible role in cancer progression and prognosis. Although the physiological function of *MUC17* is still unclear, it may serve as a physical barrier against microorganisms and as cell-surface sensor. MUC17 may also conduct signals in response to external stimuli that lead to cellular responses, including proliferation, differentiation, apoptosis or secretion of cellular products such as other membrane-bound mucins (29). Kitamoto *et al.* (29) proposed several miRNAs as potential regulators of *MUC17* expression, but none of them has been validated *in vivo*.

As there is an established interplay between *MUC21* and *MUC17* genes according to STRING, the polymorphisms significantly associated emerging from the follow up study were also

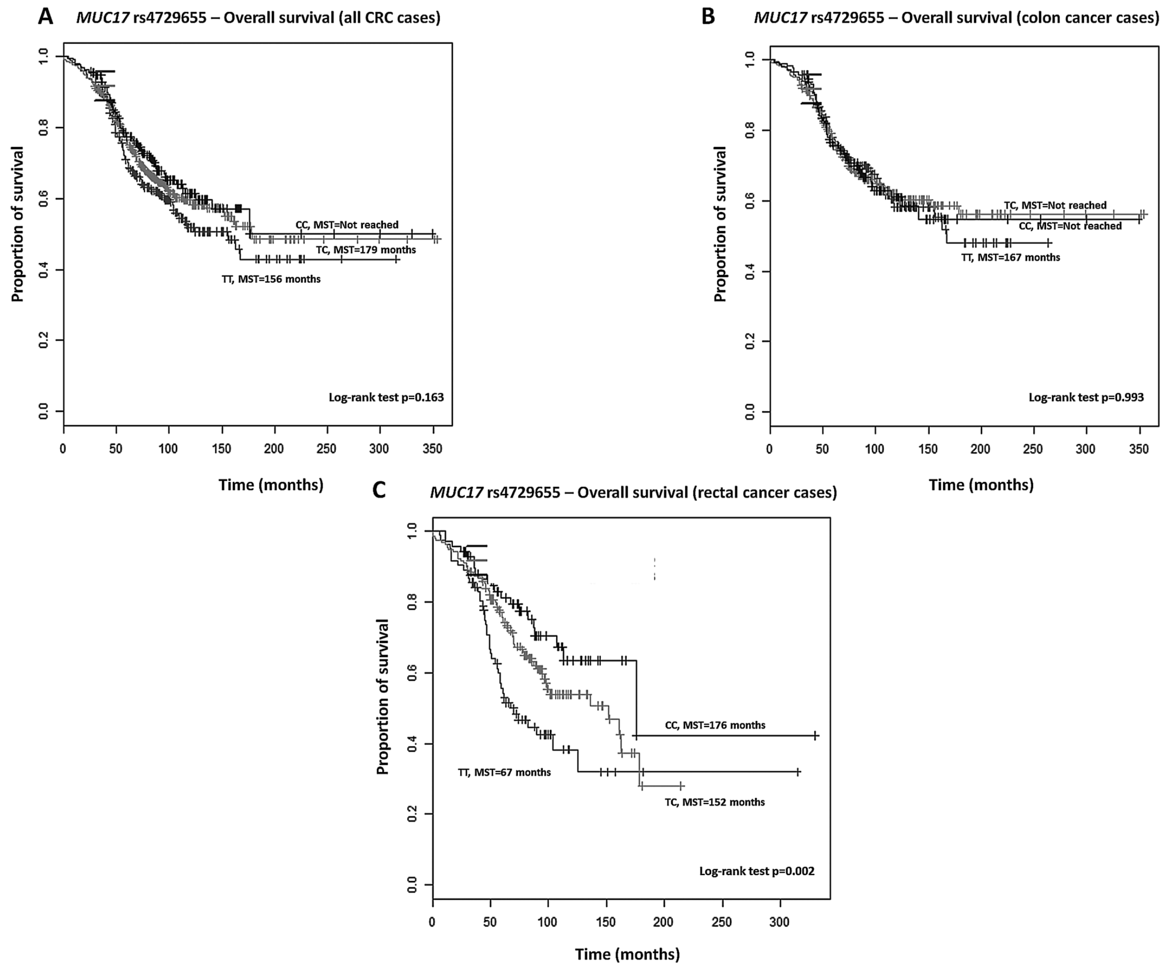


Figure 2. Kaplan–Meier OS curves stratified for rs4729655 in *MUC17* gene in (**A**) all CRC patients, (**B**) colon cancer patients and (**C**) rectal cancer patients.

explored for their potential SNP–SNP interaction on patients survival. However, no effect was observed for these two particular genes in combination.

A strong association with either shorter OS or EFS was observed for *MUC20* gene (rs6782006). CRC patients, particularly those with colon cancer, carrying the variant GG genotype of this miRSNP had worse survival. MUC20 is a novel mucin protein highly expressed in kidney and colon tissues. Based on whole-genome expression profiling of CRC, *MUC20* was significantly upregulated in CRC patients with poor prognosis (10). A relationship between *MUC20* overexpression and poor survival was found in many human tumors, including ovarian cancer (30), non-small cell lung cancer (31), and gastric cancer (32). *MUC20* overexpression predicts poor prognosis in endometrial cancer and enhances EGF-triggered invasive behavior through activation of EGFR–STAT3 pathway (33). Increased expression levels of *MUC20* promoted metastasis of CRC cells, whereas knockdown of this gene attenuated migration and invasion abilities of CRC cells (10). Like for the other mucin genes, the number of studies regarding miRNA-mediated control of the expression of *MUC20* is still scarce.

From a clinical point of view, malignancies in the colon and the rectum represent two distinct entities that require different treatment strategies. The distinction between colon and rectum is largely anatomical but it impacts both surgical and radiotherapeutic management with often different prognoses. This has been already observed by us in a previous study on rectal cancer (34) but it emerged also from the results of the present study. There are in fact different survival rates associated to the different kind of cancer site and specific miRSNPs. Our data contributes to improved understanding of the role of specific miRSNPs in rectal and colon cancer pathogenesis. Treating the 2 sites as independent entities may improve discovery of biomarkers used for early detection and prognosis.

We are aware of certain limitations of the present investigation. In the case-control study, healthy subjects differed from cases in age and gender distribution, as well as other parameters such as BMI. However, we attempted to control tentative age effect by matching cases and controls by age quartiles through bootstrap sampling (830 cases and 830 controls). In particular, we obtained similar results to those presented in this work in 8 out of 10 resamplings. Moreover, patients were collected from the same centers (with follow up data collected by the same physicians) and were highly homogeneous for their ancestry, thus with the exclusion of possible population stratifications. In addition, the inclusion of 'colonoscopy negative' individuals ensured disease-free control individuals because a negative colonoscopy result is the best available proof of the CRC absence (35). Nevertheless, there is concern that the colonoscopy negative control group is based on existing medical conditions of patients, which required examination and thus may carry unknown CRC risk factors. Since this

group of individuals may not necessarily represent the general population, we included also healthy cancer-free individuals recruited among volunteers from blood centers. However, there were no statistically significant differences in genotypes frequencies of mucin genes between the two control groups, which allowed pooling of the controls to increase the statistical power of the study.

Mucin synthesis and secretion by CRC affects the phenotype of the disease, as demonstrated *in vitro* and *in vivo*. Expanding our knowledge on mucin involvement in CRC may help us to better understand the etiopathogenesis of this disease and thereby contribute to the development of new treatment strategies. It is well known that mucinous CRC have a higher Dukes stage at diagnosis and a worse prognosis. The present results identified plausible candidate SNPs potentially affecting miRNA binding in mucin genes that were related either to CRC susceptibility or to patient's survival. Moreover, our study supports the emerging idea of a 'miRNA network' that may contribute to CRC. Further studies are needed to replicate these SNPs as predictive biomarkers in independent populations, to functionally characterize the significant genetic variants and to find the biologic mechanisms underlying the associations.

Supplementary material

Supplementary Tables 1 and 2 can be found at http://carcin.oxfordjournals.org/

Funding

Internal Grant Agency of the Ministry of Health of the Czech Republic (IGA NT 13424 and AZV MZ 15-26535A) and the Grant Agency of the Charles University GAUK 112515.
Conflict of Interest Statement: None declared.

References

1. Mehlen, P. et al. (2006) Metastasis: a question of life or death. Nat. Rev. Cancer, 6, 449–58.
2. Roessler, S. et al. (2010) A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res., 70, 10202–10212.
3. Byrd, J.C. et al. (2004) Mucins and mucin binding proteins in colorectal cancer. Cancer Metastasis Rev., 23, 77–99.
4. Debunne, H. et al. (2013) Mucinous differentiation in colorectal cancer: molecular, histological and clinical aspects. Acta Chir. Belg., 113, 385–390.
5. Hollingsworth, M.A. et al. (2004) Mucins in cancer: protection and control of the cell surface. Nat. Rev. Cancer, 4, 45–60.
6. Kufe, D.W. (2009) Mucins in cancer: function, prognosis and therapy. Nat. Rev. Cancer, 9, 874–885.
7. Yonezawa, S. et al. (2011) Mucins in human neoplasms: clinical pathology, gene expression and diagnostic application. Pathol. Int., 61, 697–716.
8. Yonezawa, S. et al. (2008) Precursor lesions of pancreatic cancer. Gut Liver, 2, 137–154.
9. Yonezawa, S. et al. (2010) Significance of mucin expression in pancreatobiliary neoplasms. J. Hepatobiliary. Pancreat. Sci., 17, 108–124.
10. Xiao, X. et al. (2013) Role of MUC20 overexpression as a predictor of recurrence and poor outcome in colorectal cancer. J. Transl. Med., 11, 151.
11. Yonezawa, S. et al. (2008) Expression profiles of MUC1, MUC2, and MUC4 mucins in human neoplasms and their relationship with biological behavior. Proteomics, 8, 3329–3341.
12. Kang, H. et al. (2011) Loss of E-cadherin and MUC2 expressions correlated with poor survival in patients with stages II and III colorectal carcinoma. Ann. Surg. Oncol., 18, 711–719.
13. Matsuda, K. et al. (2000) Clinical significance of MUC1 and MUC2 mucin and p53 protein expression in colorectal carcinoma. Jpn. J. Clin. Oncol., 30, 89–94.
14. Macha, M.A. et al. (2015) Emerging potential of natural products for targeting mucins for therapy against inflammation and cancer. Cancer Treat. Rev., 41, 277–288.
15. Slattery, M.L. et al. (2011) MicroRNAs and colon and rectal cancer: differential expression by tumor location and subtype. Genes. Chromosomes Cancer, 50, 196–206.
16. Naccarati, A. et al. (2012) Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. Carcinogenesis, 33, 1346–1351.
17. Vymetalkova, V. et al. (2014) Variations in mismatch repair genes and colorectal cancer risk and clinical outcome. Mutagenesis, 29, 259–265.
18. Pardini, B. et al. (2013) Variation within 3'-UTRs of base excision repair genes and response to therapy in colorectal cancer patients: A potential modulation of microRNAs binding. Clin. Cancer Res., 19, 6044–6056.
19. Naccarati, A. et al. (2016) Double-strand break repair and colorectal cancer: gene variants within 3' UTRs and microRNAs binding as modulators of cancer risk and clinical outcome. Oncotarget, 7, 23156–23169.
20. Pardini, B. et al. (2009) NBN 657del5 heterozygous mutations and colorectal cancer risk in the Czech Republic. Mutat. Res., 666, 64–67.
21. Pardini, B. et al. (2015) Polymorphisms in microRNA genes as predictors of clinical outcomes in colorectal cancer patients. Carcinogenesis, 36, 82–86.
22. Gupta, B.K. et al. (2012) Increased expression and aberrant localization of mucin 13 in metastatic colon cancer. J. Histochem. Cytochem., 60, 822–831.
23. Niv, Y. (2000) Mucin and colorectal cancer. Isr. Med. Assoc. J., 2, 775–777.
24. Labelle, M. et al. (2012) The initial hours of metastasis: the importance of cooperative host-tumor cell interactions during hematogenous dissemination. Cancer Discov., 2, 1091–1099.
25. Itoh, Y. et al. (2008) Identification and expression of human epiglycanin/MUC21: a novel transmembrane mucin. Glycobiology, 18, 74–83.
26. Yi, Y. et al. (2010) Mucin 21/epiglycanin modulates cell adhesion. J. Biol. Chem., 285, 21233–21240.
27. Ho, S.B. et al. (2010) Activity of recombinant cysteine-rich domain proteins derived from the membrane-bound MUC17/Muc3 family mucins. Biochim. Biophys. Acta, 1800, 629–638.
28. Li, K. et al. (2011) Aberrant Expression of MUC17 Correlates With Survival in Advanced Stage Colorectal Adenocarcinomas. Gastroenterology, 140, S-145.
29. Kitamoto, S. et al. (2011) DNA methylation and histone H3-K9 modifications contribute to MUC17 expression. Glycobiology, 21, 247–256.
30. Vlad, A.M. et al. (2006) MUC1 in endometriosis and ovarian cancer. Immunol. Res., 36, 229–236.
31. Woenckhaus, M. et al. (2008) Prognostic value of FHIT, CTNNB1, and MUC1 expression in non-small cell lung cancer. Hum. Pathol., 39, 126–136.
32. Marín, F. et al. (2012) Genetic variation in MUC1, MUC2 and MUC6 genes and evolution of gastric cancer precursor lesions in a long-term follow-up in a high-risk area in Spain. Carcinogenesis, 33, 1072–1080.
33. Chen, C.H. et al. (2013) MUC20 overexpression predicts poor prognosis and enhances EGF-induced malignant phenotypes via activation of the EGFR-STAT3 pathway in endometrial cancer. Gynecol. Oncol., 128, 560–567.
34. Vymetalkova, V. et al. (2016) Epigenome-wide analysis of DNA methylation reveals a rectal cancer-specific epigenomic signature. Epigenomics, 8, 1193–1207.
35. Slattery, M.L. et al. (2009) A comparison of colon and rectal somatic DNA alterations. Dis. Colon Rectum, 52, 1304–1311.

<u>Manuscript IV</u>

Huhn S, da Silva Filho M, Sanmuganatham T, Pichulik T, Catalano C, Pardini B, Naccarati A, Polakova-Vymetálkova V, **Jiraskova K**, Vodickova L, Vodicka P, Löffler MW, Courth L, Wehkamp J, VN Din F, Timofeeva M, Jansen L, Hemminki K, Chang-Claude J, Brenner H, Hoffmeister M, Dunlop M, Weber ANR, Försti A.

Coding variants in NOD-like receptors: An association study on risk and survival of colorectal cancer

PlosOne (2018) 13(6)

PLOS ONE

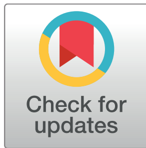
RESEARCH ARTICLE

Coding variants in NOD-like receptors: An association study on risk and survival of colorectal cancer

Stefanie Huhn[1,2]*, Miguel I. da Silva Filho[1], Tharmila Sanmuganantham[3], Tica Pichulik[3], Calogerina Catalano[1], Barbara Pardini[4,5], Alessio Naccarati[4,5], Veronika Polakova-Vymetálkova[5,6], Katerina Jiraskova[5,6], Ludmila Vodickova[5,6,7], Pavel Vodicka[5,6,7], Markus W. Löffler[3,8], Lioba Courth[9], Jan Wehkamp[9], Farhat V. N. Din[10], Maria Timofeeva[10], Susan M. Farrington[10], Lina Jansen[11], Kari Hemminki[1,12], Jenny Chang-Claude[13], Hermann Brenner[11,14,15], Michael Hoffmeister[11], Malcolm G. Dunlop[10], Alexander N. R. Weber[3]*, Asta Försti[1,12]

1 Department of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, **2** Department of Multiple Myeloma, Internal Medicine V: Hematology, Oncology and Rheumatology, Heidelberg University Hospital, Heidelberg, Germany, **3** Interfaculty Institute for Cell Biology, Department of Immunology, University of Tübingen, Tübingen, Germany, **4** Italian Institute for Genomic Medicine (IIGM), Turin, Italy, **5** Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic, **6** Institute of Biology and Medical Genetics, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, **7** Biomedical Centre, Faculty of Medicine Pilsen, Charles University Prague, Pilsen, Czech Republic, **8** Department of General, Visceral and Transplant Surgery, University Hospital Tübingen, Tübingen, Germany, **9** Department of Internal Medicine I, University Hospital Tübingen, Tübingen, Germany, **10** Colon Cancer Genetics Group, MRC Human Genetics Unit, The University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom, **11** Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, **12** Center for Primary Health Care Research, Clinical Research Center, Lund University, Malmö, SE, Sweden, **13** Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, **14** Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Im Neuenheimer Feld 460, Heidelberg, Germany, **15** German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

* Stefanie.Huhn@med.uni-heidelberg.de (SH); alexander.weber@uni.tuebingen.de (AW)



OPEN ACCESS

Citation: Huhn S, da Silva Filho MI, Sanmuganantham T, Pichulik T, Catalano C, Pardini B, et al. (2018) Coding variants in NOD-like receptors: An association study on risk and survival of colorectal cancer. PLoS ONE 13(6): e0199350. https://doi.org/10.1371/journal.pone.0199350

Editor: Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

Received: March 12, 2018

Accepted: June 6, 2018

Published: June 21, 2018

Copyright: © 2018 Huhn et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. All genotype data pertinent to this study are presented in the paper. Survival analysis was performed as described. Individual clinical and follow-up level data for this analysis are only relevant to the SNPs of interest and so provided in the analysis.

Funding: This work was supported by the COST Action BM1206 to AF, COST-CZ: LD14050, AZV

Abstract

Nod-like receptors (NLRs) are important innate pattern recognition receptors and regulators of inflammation or play a role during development. We systematically analysed 41 non-synonymous single nucleotide polymorphisms (SNPs) in 21 NLR genes in a Czech discovery cohort of sporadic colorectal cancer (CRC) (1237 cases, 787 controls) for their association with CRC risk and survival. Five SNPs were found to be associated with CRC risk and eight with survival at 5% significance level. In a replication analysis using data of two large genome-wide association studies (GWASs) from Germany (DACHS: 1798 cases and 1810 controls) and Scotland (2210 cases and 9350 controls) the associations found in the Czech discovery set were not confirmed. However, expression analysis in human gut-related tissues and immune cells revealed that the NLRs associated with CRC risk or survival in the discovery set were expressed in primary human colon or rectum cells, CRC tissue and/or cell lines, providing preliminary evidence for a potential involvement of NLRs in general in CRC development and/or progression. Most interesting was the finding that the enigmatic development-related *NLRP5* (also known as *MATER*) was not expressed in normal colon

15-26535A, 15-27580A and Progres Q28/LF1 to PV. TS, TP, MWL and ANRW were supported by the DFG collaborative research center SFB685 "Immunotherapy", the University of Tübingen, the Federal State Baden-Württemberg Junior-Professor-Program, the German Cancer Consortium (DKTK) and/or the Henriette-and-Ottmar-Eier Stiftung. DACHS: This work was further supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1) to HB, the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B) to HB, and from the National Institutes of Health (USA) (U01 CA137088) to HB. In Scotland, the work was supported by Programme Grant funding from Cancer Research UK [C348/A18927] to MGD which funded genotyping, phenotyping and follow up of cases and matched population controls. Additional genotyping and phenotyping of Scottish population control subjects received grant support as follows: Lothian Birth Cohorts (Ian J Deary and Sarah E Harris) - Biotechnology and Biological Sciences Research Council (BBSRC; BB/F019394/1, 15/SAG09977) to MGD, The Royal Society, The Chief Scientist Office of the Scottish Government, Age UK-The Disconnected Mind project, and the Centre for Cognitive Ageing and Cognitive Epidemiology (co-funded by BBSRC and Medical Research Council G0700704/84698 and MR/K026992/1) to MGD. Generation Scotland received core funding from the Chief Scientist Office of the Scottish Government Health Directorate CZD/16/6 and the Scottish Funding Council HR03006 to MGD. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z) to MGD.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: ASC, Apoptosis-associated speck-like protein containing a CARD; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CI, confidence interval; CRC, colorectal cancer; EFS, event-free survival; HNPCC, hereditary-nonpolyposis colorectal cancer; HR, hazard ratio; HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; MoDC, monocyte-derived dendritic cells; MoMacs, monocyte-derived macrophages; NLRs, Nod-like receptors; nsSNPs, non-synonymous single nucleotide polymorphisms; OR, Odds ratio; OS,

tissue but in colon cancer tissue and cell lines. Future studies may show whether regulatory variants instead of coding variants might affect the expression of NLRs and contribute to CRC risk and survival.

Introduction

Within the last few years it has become evident that the interplay between pattern recognition receptors (PRRs), such as Toll-like receptors (TLRs) or Nod-like receptors (NLRs), and the gut microbiota has a profound influence on the homeostasis of the immune system and therefore on many important aspects of human health [1–3]. If undisturbed and well-regulated, this symbiosis is beneficial for the human host. A disruption of the underlying regulatory pathways can, however, result in the development of local and chronic inflammation, inflammatory bowel disease (IBD) and/or colorectal cancer (CRC) [4, 5]. Gut homeostasis is maintained by a physical separation of the microbial community from the gut epithelia by the mucosa and a mucus layer. PRRs monitor the integrity of this barrier and the adjacent microbial community by detecting microbe-associated molecular patterns (MAMPs) as well as endogenous damage associated molecular patterns (DAMPs), and consequently controlling antimicrobial responses that contribute to an equilibrium between microbes and host [2, 4–6]. The activation of PRRs TLRs and/or NLRs by MAMPs or DAMPs results in the activation of multiple signaling pathways including nuclear factor-κB (NF-κB), mitogen-activated protein kinases (MAPKs), and the type I interferon (IFN) response, with subsequent induction of an inflammatory and antimicrobial response that includes secretory IgA, antimicrobial peptides, pyroptosis and autophagy [1, 2, 7]. Some NLRs, such as NLRP1, 3, 6, 12 and NLRC4, form so-called "inflammasome" complexes, comprising of the respective NLRs, the adaptor Apoptosis-associated speck-like protein containing a CARD (ASC) and pro-caspase-1. Inflammasome assembly initiates inflammatory and antimicrobial response via the autoproteolytic cleavage of caspase-1, catalysing the proteolytic conversion of pro-interleukin-1β (IL-1ß) and other IL-1 family members into biologically active cytokines which drive inflammation [2, 8, 9].

We recently reported an impact of TLR polymorphisms on CRC survival [10, 11]. Given the suggested concerted action of TLR and NLR signaling [1, 2], the connection between NLRs and CRC seemed of special interest. Provoking studies in mice and association studies in humans have suggested that NLR signaling is involved in inflammatory bowel disease, chronic inflammation and gastrointestinal cancers, including CRC [6, 11, 12]. Apart from various reports on mice, convincing data directly connecting NLRs and human CRC are available only for *NOD1*, *NOD2* and *NLPR3*, which were found to be associated with susceptibility, progression and treatment of sporadic CRC, colitis and/or colitis-associated CRC [13, 14]. In general, it is unclear whether and how other NLRs contribute to human CRC development or progression. Additionally, the functional importance of NLRs that have embryonic lethal phenotypes in mice–so-called "reproduction-related NLRs" like NLRP2, 5 and 13 –in human immunity and/or tumorigenesis remains an unresolved question [8].

In order to systematically investigate the influence of potentially functional coding polymorphisms in the NLR genes on sporadic CRC risk and survival, we conducted a case-control study with replications in 2 large genome-wide association studies (GWASs), covering the majority of known non-synonymous single nucleotide polymorphisms (nsSNPs) in 21 genes across different NLR signalling pathways. *In silico* analysis was done on selected nsSNPs in the NLR gene family. Furthermore, RNA expression of selected genes was measured to assess mRNA expression in immune cells, biopsies and/or CRC cell lines.

overall survival; pM 0, no metastasis upon diagnosis; pM1, metastasis upon diagnosis; PRRs, pattern recognition receptors; SNPs, single nucleotide polymorphisms; TLRs, Toll-like receptors; TNM, UICC TNM staging: size or direct extent of the primary tumor (T); degree of spread to regional lymph nodes (N); presence of metastasis (M).

Methods and material

Ethical approval

The Czech study: Ethics Committee of the Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, 26.3.2004; Ethics Committee of the Institute of Clinical and Experimental Medicine and Faculty Thomayer Hospital, Prague, Czech Republic, 29.4.2009; and Ethics Committee of the General University Hospital, Prague, Czech Republic, 4.4.2011.

For the work in Tübingen: All patients or healthy blood donors included in gene expression analyses for this study provided their written informed consent before study inclusion. Approval for use of their biomaterials was obtained by the local ethics committee at the University of Tübingen, in accordance with the principles laid down in the Declaration of Helsinki. Terminal ileum/ colon biopsies were obtained from patients undergoing routine colonoscopy at the University Hospital Tübingen, buffy coats obtained from blood donations of healthy donors were received from the Center for Clinical Transfusion Medicine (ZKT) at the University Hospital Tübingen and whole blood from voluntary healthy donors was obtained at the University of Tübingen, Department of Immunology.

The DACHS study was approved by the ethics committee of the Medical Faculty of the University of Heidelberg (no. 310/2001). The DACHS study is registered: StudyBox no. ST-066, DRKS no. DRKS00011793

The work in Scotland was approved by the UK National Health Service Research Ethics Committee (approval references 13/SS/0248; 11/SS/0109 and 01/0/05).

SNP selection and in silico analysis of conservation and functional relevance

21 candidate NLR genes (NLRP1-14, NLRC4 and 5, NOD1 and 2, NAIP, RIPK2 and ASC [PYCARD]) were screened for non-synonymous variants. Thirteen of the 21 genes harboured validated missense variants with a minor allele frequency (MAF) > 0.01 in the CEU reference panels (Source: 1000Genomes, HapMap, dbSNP). Choosing only one SNP per linkage block ($r^2 \geq 0.8$), 41 SNPs were selected for genotyping (Table A in S2 File & Fig 1).

To gain additional insight into a possible functional relevance, all genotyped and linked SNPs were mapped to their location in the respective proteins. 19 of the 41 genotyped SNPs are located in defined NLR protein domains [3] (Fig 1 & Table A in S2 File): 11 in the NACHT domain, seven in the LRR domain and one in the PYD domain. The remaining genotyped SNPs mapped to linker regions. SIFT (sift.jcvi.org) and PolyPhen2 (genetics.bwh.harvard.edu/pph2) databases were used to assess possible effects of the SNPs on the protein. 23 of the genotyped SNPs are predicted to be deleterious or damaging, and/or to result in non-sense mediated decay or retained introns (Fig 1 & Table A in S2 File). Assessments of evolutionary conservation of the selected variants was performed by three software suites namely Genomic Evolutionary Rate Profiling (GERP [15]), PhastCons [16] and phylogenetic p-value (PhyloP [17]). The GERP score of >2.0 and the PhastCons score (values between 0–1) of >0.3 indicate a good level of conservation of the variants. Positive PhyloP scores (values between −14 and +6) are predicted to be conserved. Higher values of these tools reflect the probability that the nucleotide is located at a conserved position, based on the multiple alignment of genome sequences of 100 different vertebrates. Lower values of these tools reflect fast-evolving variant positions.

Discovery set—Czech Republic

Study population. The study was carried out on a Czech CRC case-control population of patients (n = 1237; median age 63 years; 61.7% males) with colon or rectal malignancy—

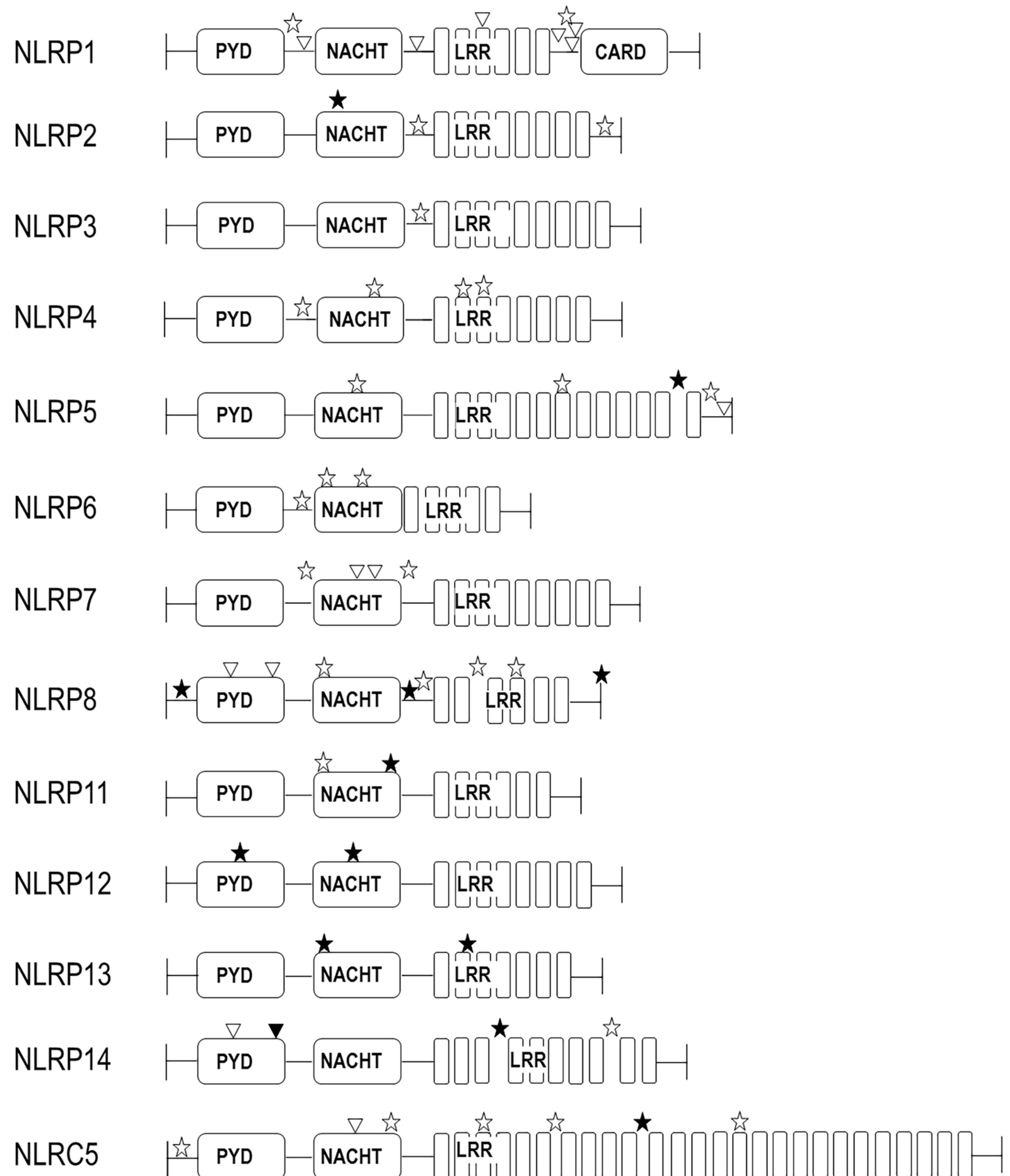


Fig 1. Protein structure of the candidate genes with genotyped SNPs (open and filled star symbols ☆/★), and all linked missense SNPs (triangle △; $r^2 \geq 0.8$). Filled star (★) and triangle (▲): SNPs predicted damaging (SIFT) or deleterious (Polyphen).

https://doi.org/10.1371/journal.pone.0199350.g001

excluding hereditary-nonpolyposis colorectal cancer (HNPCC)—and healthy blood donors (n = 787; median age 47 years, 55.4% males, all cancer-free at the time of sampling), (Table B in S2 File.) [18]. For overall survival (OS) 477 incident CRC cases with information about age, sex, TNM staging, tumor grade, date of death or end of follow-up (August 31, 2011; median follow-up 58 months) were available. For event-free survival (EFS) in patients with non-metastatic disease at the time of diagnosis (n = 325), date of distant metastasis, tumor recurrence, death, or end of follow-up was used as the study end point (median follow-up 55 months).

SNP genotyping. TaqMan SNP Genotyping Assays (Applied Biosystems) or KASP genotyping assays (LGC Genomics) were used for the analysis of the SNPs. Case and control samples were amplified simultaneously in 384 well format (Hydrocycler 16 (LGC Genomics), using 3 ng whole genome amplified DNA from blood). Endpoint genotype detection was carried out on the ViiA 7 Real-Time PCR System (Applied Biosystems). Call rates for 40 out of 41 SNPs were 94–99%. Internal quality controls showed a concordance rate of $\geq$ 99%. Samples with $<$ 50% call rate over all assays were excluded from the study.

Replication sets—Germany and Scotland

For replication, all SNPs associated with CRC in the Czech population ($p<0.05$) were tested in two large European genome-wide association studies (GWASs) carried out in Germany („Darmkrebs: Chancen der Verhütung durch Screening Study"–DACHS) [19, 20] and in Scotland (Table B in S2 File) [21, 22].

Germany. The sample set used as the replication set is part of the still on-going DACHS project and comprised 1796 CRC patients (median age 69 years; 58.6% males) who received in-patient treatment due to a first diagnosis of CRC in 22 hospitals of the Rhine-Neckar-Odenwald region of Germany. The 1810 community-based controls were randomly selected from population registries matched for gender, 5-year age groups and county of residence (median age 70 years, 59.6% males, cancer-free at the time of sampling), (Table B in in S2 File) [19, 20]. Cases and controls genotyped in the present study were recruited between January 01, 2003 and December 31, 2007. For overall survival (OS) analysis, 1794 incident CRC cases with information about age, sex, tumor stage, and a median follow-up time of 48.4 months in men and 49.9 months in women were available [23].

Cases and controls were genotyped on the Illumina HumanCytoSNP or Illumina HumanOmniExpress platform [24]. Imputation was performed for autosomal SNPs to the CEU population in HapMap II release 24 using MACH (available at: www.sph.umich.edu/csg/abecasis/MACH/tour/) [24] with MAF ($<$0.01) and imputation accuracy ($R2 < 0.3$) excluded from the analysis [25].

Scotland. The Scottish study series comprised 2115 cases (median age 57 years, 57% males) from the Scottish colorectal cancer study (SOCCS) [21] and 95 cases (median age 67 years, 66% males) from Ninewells Hospital, Dundee and Perth Royal Infirmary collected between 1997 and 2000 [26]. SOCCS is a case-control study designed to identify genetic and environmental factors associated with non-hereditary CRC risk and survival outcome. Population controls with no personal history of cancer were ascertained from four cohorts including 8533 (42% males, mean age 55.4 yrs)—from Generation Scotland-Scottish Family Health Study [27, 28]; 513 (41% males, mean age 79 years) and 1004 (50.6% males, mean age 70 years) from the Lothian Birth Cohorts 1921 and 1936, respectively; and 262 Dundee controls (50% males) were recruited through the same General Practice surgeries as cases or from spouses/friends of cases [29]. The detailed information on genotyping cases and controls and data quality control is described elsewhere [22]. 2210 cases and 9350 controls were included in the final analysis. The survival analysis was performed in a subset of SOCCS study comprising 1402 patients (median follow up 107 months, recruited between 2001 and 2006) with colorectal adenocarcinoma confirmed by pathological assessment. Participants completed a detailed lifestyle questionnaire and a semi-quantitative food frequency and supplements questionnaire (http://www.foodfrequency.org). Genotyping was performed using the Infinium Human Exome BeadChip 12v1.0 or 12v1.1 (Illumina), with genotype calling using Illumina GenCall for HumanExome-12v1.0 and HumanExome-12v1.1 versions called separately. Generation Scotland controls and a subset of the cases from the SOCCS study were genotyped using OmniExpressExome BeadChip 8v1.1 or 8v1.2 (Illumina).

In accordance with the Declaration of Helsinki, all participants provided written informed consent. The studies were approved by the local ethics committees.

Statistical analysis—Discovery set

Genotype frequencies in controls were tested for Hardy-Weinberg equilibrium (HWE; Pearson's goodness-of-fit χ^2 test, deviation assumed at $p < 0.001$). *NLRP11* rs12461110 was excluded for violation of HWE.

Single variant associations with CRC risk, overall and event-free survival. Odds ratios (ORs) and 95% confidence intervals (CIs) for associations between genotypes and CRC risk were estimated by logistic regression (PROC LOGISTIC, SAS V9.3; SAS Institute, Cary, NC) and refer to the minor allele. P values were considered nominally significant at $p \leq 0.05$, with a study-wide significance level at $p \leq 0.001$ considering Bonferroni correction for multiple testing (0.05/39 = 0.0012). ORs were adjusted for age and sex. The estimated power was > 95% for OR ≥1.5 (MAF > 5%; p = 0.05; dominant model) [30].

Differences in OS and EFS between genotypes were estimated by hazard ratios (HRs) and 95% CIs using Cox regression (PROC PHREG, SAS V9.3) adjusting for age, sex, tumor grade and tumor stage. The estimated power was > 90% for HR ≥ 2.0 (MAF > 10%, p = 0.05). OS was calculated for all patients ($OS_{(pM0\&1)}$); for patients with non-metastatic disease at the time of diagnosis OS ($OS_{(pM0)}$) and EFS ($EFS_{(pM0)}$) were calculated. Kaplan-Meier plots were generated, estimating the differences between the survival functions by log-rank test (PROC LIFETEST, SAS V9.3).

Additive SNP associations with CRC risk and survival. Additive influence of the risk alleles ($p \leq 0.05$) on CRC risk and survival identified in the Czech population was estimated (risk: five SNPs, 0–10 risk alleles per individual; survival: eight SNPs, 0–16 risk alleles). For each SNP the allele associated with a higher OR or HR was designated the "risk allele". Patients were grouped into equally sized groups of risk alleles (risk: 0-3/4-5/6-10; survival: 0-5/6-7/8-12) and ORs and HRs were calculated, adjusting for age and sex, HRs also for tumor grade and stage. Kaplan-Meier plots were generated for the additive survival model and the log-rank test was performed. The same analysis was conducted separately for the three *NLRP5* risk SNPs ($p \leq 0.05$) (0–6 risk alleles).

Statistical analysis—Replication sets. Data provided by the DACHS study consisted mostly of imputed genotypes, all in dosage format referring to the number of copies of minor allele. To permit direct comparison with the German data set genotype data from the Czech study was coded as 0, 1, or 2 copies of the minor allele. For these two data sets, association between SNPs and risk for CRC was obtained by applying logistic regression considering a log-additive genetic effect model (PROC LOGISTIC, SAS Version 9.2; SAS Institute). HRs (PROC PHREG, SAS version 9.2, SAS Institute) were calculated via Cox regression with a model that included the SNP coded as number of copies of the minor allele, age, sex and tumor stage for both sample sets.

ORs and 95% CIs for association between each of the genotypes and risk of CRC in Scotland were estimated using unconditional logistic regression adjusted for age and gender. HRs and corresponding 95%CI for overall survival analysis in Scotland was calculated for each of the genotyped SNPs and dominant model using Cox regression adjusted for age, gender and TNM stage. OS was calculated for all patients ($OS_{(pM0\&1)}$) and for patients with non-metastatic disease at the time of diagnosis OS ($OS_{(pM0)}$). No event-free survival analysis was performed in the Scottish data. All analysis was performed in R v3.1.0 (R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2014).

Gene expression

For expression analyses, patients providing biopsy material were recruited at the University Hospital Tübingen. Healthy blood donors were recruited at the Center for Clinical Transfusion Medicine (ZKT), University Hospital Tübingen and respective buffy coats obtained from blood donations. All patients/ healthy blood donors included in gene expression analyses for this study provided their written informed consent before study inclusion. Approval for use of their biomaterials was obtained by the local ethics committee at the University of Tübingen, in accordance with the principles laid down in the Declaration of Helsinki. Terminal ileum/ colon biopsies were obtained from patients undergoing routine colonoscopy at the University Hospital Tübingen, buffy coats obtained from blood donations of healthy donors were received from the Center for Clinical Transfusion Medicine (ZKT) at the University Hospital Tübingen and whole blood from voluntary healthy donors was obtained at the University of Tübingen, Department of Immunology.

Cell lines, primary human immune cells and biopsy material. HCT116, DLD-1 and Caco2 cells were grown and sourced as described [31], without re-authentication. Primary leukocytes were isolated from buffy coats (Tübingen University Hospital, Center for Clinical Transfusion Medicine (ZKT)) using Ficoll (GE Healthcare) density gradient purification and CD14+ monocytes were isolated using MACS (Miltenyi) magnetic beads to a purity of $> 95\%$ (anti-CD14-PE flow cytometry, BD). Subsequently, cells were differentiated into monocyte-derived dendritic cells (MoDC) or monocyte-derived macrophages (MoMacs) by culture in RPMI 1640 medium (Invitrogen) supplemented with 10% fetal calf serum (FCS) in the presence of 40 ng/ml IL-4 and 25 ng/ml GM-CSF (Peprotech) or with 25 ng/ml GM-CSF for 6 days, respectively. Neutrophils were isolated from the Ficoll pellet after NH_4Cl lysis of erythrocytes. All cells were grown at 37˚C and 5% CO_2. Biopsies from the terminal ileum or colon (n = 12; median age 46; 56% males) were obtained during routine colonoscopy at the University Hospital Tübingen and stored in liquid nitrogen until analysis [13].

Gene expression analysis. Gene expression analysis was carried out using single-gene TaqMan® Gene Expression Assays (Applied Biosystems) for *NLRP2*, *NLRP3*, *NLRP5*, *NLRP6*, *NLRP13* and *NLRC5*. mRNA was isolated from whole blood or primary blood cells (two donors, #1 and #2, respectively) or THP-1, HCT116, DLD-1 or CaCo2 cell lines using an RNeasy Mini Kit (Qiagen) and commercially available RNA samples for human ovary, duodenum, ileum (sample #7), rectum and colon adenocarcinoma were used (Agilent). RNA from ileum or colon biopsies from six patients (samples #1–6) was isolated using TRIzol Reagent (Life Technologies) according to standard protocols and reverse transcribed into cDNA using oligo(dT)12 primer [13]. Following transcription to cDNA (High Capacity RNA-to-cDNA Kit; Life Technologies), expression was analysed using pre-validated TaqMan® Gene Expression Assays (Applied Biosystems). Data were normalized to *TBP* (TATA box binding protein). The samples were analysed in triplicate using the 7500fast Real-Time System (Applied Biosystems).

Results

NLR variants are associated with CRC risk and survival in the Czech sample set

Nominally significant associations with CRC risk were detected for six SNPs (Table 1; Table C in in S2 File). In an additive risk model combining those six variants, CRC risk increased significantly with increasing numbers of risk alleles, and a maximum for carriers of 6–10 risk alleles (OR 2.10, p = 0.0005; Table 1).

Table 1. CRC risk: genotype distribution of SNPs analyzed in the Czech case-control population for SNPs with $p \leq 0.05$. Amino acid changes are given as <> with the amino acid position indicated. Data adjusted for age at diagnosis and sex. Nominal significance at p≤0.05; significance level corrected for multiple testing (39 genotyped SNPs) at p≤0.001.

Gene	Risk of CRC				
SNP	**Genotype**	**Cases**	**Controls**	**OR (95%CI)**	**P Val**
NLRP2	C/C	427	284	1	
rs1043673	A/C	574	355	1.08 (0.84–1.39)	0.56
1052: A<>E	A/A	203	108	1.41 (1.00–1.99)	**0.05**
	A/C + A/A	777	463	1.16 (0.91–1.47)	0.23
NLRP3	C/C	1114	700	1	
rs35829419	A/C	85	66	0.63 (0.41–0.97)	**0.04**
705: Q<>K	A/A	4	1	0.97 (0.09–10.11)	0.98
	C/A + A/A	89	67	0.64 (0.42–0.98)	**0.04**
NLRP6	G/G	924	629	1	
rs6421985	T/G	252	128	1.36 (1.01–1.83)	**0.04**
163: L<>M	T/T	-	-	-	-
	T/G + T/T	252	128	1.36 (1.01–1.83)	**0.04**
NLRP8	G/G	732	493	1	
rs306457	C/G	415	235	1.13 (0.88–1.45)	0.32
1049: STOP<>Y	C/C	63	26	2.01 (1.09–3.72)	**0.03**
	C/G + C/C	478	261	1.20 (0.95–1.53)	0.13
NLRP11	A/A	1070	662	1	
rs299163	A/C	134	91	0.94 (0.66–1.34)	0.74
188: A<>S	C/C	7	12	0.21 (0.06–0.68)	**0.01**
	A/C + C/C	141	103	0.83 (0.59–1.17)	0.2954
NLRP13	C/C	419	297	1	
rs303997	C/T	564	334	1.37 (1.06–1.77)	**0.02**
247: R<>Q	T/T	212	125	1.54 (1.10–2.16)	**0.01**
	C/T + T/T	776	459	1.42 (1.11–1.80)	**0.005**
No. of	0–3	302 (27.83)	216 (32.34)	1	-
risk alleles [a]	4–5	633 (58.34)	381 (57.04)	1.36 (1.04–1.79)	**0.03**
	6–10	150 (13.82)	71 (10.63)	2.10 (1.38–3.20)	**0.0005**

[a] NLRP11 rs299163 was excluded from the "No. of risk alleles analysis" due to low MAF 0.05: only the rare homozygote genotype was associated with CRC risk, not contributing in risk in the Risk-SNP-Panel.

https://doi.org/10.1371/journal.pone.0199350.t001

Eight SNPs were associated with altered OS and/or EFS (p≤0.05); Table 2; Table C in in S2 File). Strikingly, among them three unlinked SNPs in *NLRP5* ($r^2 < 0.5$) were associated with decreased $OS_{(pM0)}$ and $EFS_{(pM0)}$. The additive survival model showed a nominally significantly decreased OS and EFS with an increasing number of risk alleles (Table 2). The maximum effect was detected for carriers of 8–12 risk alleles ($HR_{OS(pM0\&1)}$ 1.88, p = 0.003; $HR_{OS(pM0)}$ 2.89 p = 0.0008 and $HR_{EFS(pM0)}$ 3.02, p = 0.0003, respectively; Table 2).

GWAS data on the NLRP risk and survival SNPs—Replication sets

In order to validate the results from the Czech cohort, all SNPs included into the additive model for CRC risk (N = 5) and survival (N = 8) were analyzed in two large GWAS sample sets from Germany and Scotland. Complete data was available from the DACHS GWAS. The Scottish GWAS provided data on three CRC risk variants (rs12150220, rs306457 and rs303997) and seven survival variants (Table E in in S2 File). Scottish data was available for

Table 2. Overall survival pM0&1 and pM0, and event-free survival pM0: genotype distribution of SNPs analyzed in the Czech case-control population for SNPs with $p \leq 0.05$. Amino acid changes are given as <> with the amino acid position indicated. Data adjusted for age at diagnosis and sex, tumor grade and tumor stage. Nominal significance at $p \leq 0.05$; significance level corrected for multiple testing at $p \leq 0.001$.

Gene		Overall Survival (pM = 0&1)				Overall Survival (pM = 0)				Event-free Survival (pM = 0)			
SNP	**Genotype**	**Cases**	**Death (%)**	**HR (95%CI)**	**p-val**	**Cases**	**Deaths (%)**	**HR (95%CI)**	**p-val**	**Cases**	**Events (%)**	**HR (95%CI)**	**p-val**
NLRP1	A/A	106	46 (43.40)	1		74	24 (32.43)	1		74	29 (39.19)	1	
rs12150220	A/T	177	86 (48.59)	1.14 (0.79–1.63)	0.49	130	43 (33.08)	1.10 (0.67–1.83)	0.70	130	50 (38.46)	0.95 (0.60–1.51)	0.81
155: H<>L	T/T	78	42 (53.85)	1.57 (1.03–2.40)	**0.04**	59	25 (42.37)	1.52 (0.86–2.68)	0.15	59	27 (45.76)	1.31 (0.77–2.21)	0.32
	A/T+T/T	255	128 (50.20)	1.25 (0.89–1.76)	0.2	189	68 (35.98)	1.23 (0.77–1.97)	0.39	189	77 (40.74)	1.05 (0.68–1.62)	0.82
NLRP2	C/C	125	73 (58.40)	1		87	38 (43.68)	1		87	42 (48.28)	1	
rs1043673	A/C	181	71 (39.23)	0.64 (0.46–0.89)	**0.008**	136	38 (27.94)	0.59 (0.37–0.92)	**0.02**	136	45 (33.09)	0.61 (0.40–0.93)	**0.02**
1052: A<>E	A/A	56	30 (53.57)	0.83 (0.54–1.28)	0.4	40	15 (37.50)	0.77 (0.42–1.42)	0.41	40	16 (40.00)	0.75 (0.42–1.35)	0.34
	A/C + A/A	237	101 (42.62)	0.69 (0.50–0.93)	**0.02**	176	53 (30.11)	0.63 (0.41–0.96)	**0.03**	87	42 (48.28)	0.64 (0.43–0.96)	**0.03**
NLRP5	G/G	197	88 (44.67)	1		143	42 (29.37)	1		143	48 (33.57)	1	
rs10409555	A/G	140	68 (48.57)	1.20 (0.87–1.64)	0.3	104	39 (37.50)	1.53 (0.98–2.38)	0.06	104	47 (45.19)	1.56 (1.04–2.35)	**0.03**
1181: V<>I	A/A	25	18 (72.00)	1.58 (0.93–2.69)	0.09	17	10 (58.82)	3.04 (1.48–6.23)	**0.002**	17	10 (58.82)	2.36 (1.17–4.78)	**0.02**
	A/G + A/A	165	86 (52.12)	1.26 (0.93–1.70)	0.13	121	49 (40.50)	1.69 (1.11–2.58)	**0.02**	121	57 (47.11)	1.66 (1.12–2.45)	**0.01**
NLRP5	C/C	268	122 (45.52)	1		195	61 (31.28)	1		195	69 (35.38)	1	
rs12462795	C/G	90	47 (52.22)	1.28 (0.91–1.81)	0.15	68	28 (41.18)	1.71 (1.07–2.74)	**0.03**	68	34 (50.00)	1.80 (1.17–2.76)	**0.007**
1108: S<>C	G/G	6	5 (83.33)	2.79 (1.13–6.90)	**0.03**	3	2 (66.67)	3.05 (0.73–12.74)	0.13	3	2 (66.67)	2.72 (0.66–11.29)	0.17
	C/G + G/G	96	52 (54.17)	1.36 (0.97–1.89)	0.07	71	30 (42.25)	1.77 (1.12–2.81)	**0.02**	71	36 (50.70)	1.84 (1.21–2.79)	**0.005**
NLRP5	T/T	253	118 (46.64)	1		184	57 (30.98)	1		184	64 (34.78)	1	
rs16986899	C/T	99	54 (54.55)	1.21 (0.87–1.67)	0.26	73	33 (45.21)	1.70 (1.10–2.64)	**0.02**	73	39 (53.42)	1.74 (1.17–2.61)	**0.007**
912: M<>T	C/C	8	2 (25.00)	0.27 (0.07–1.13)	0.07	4	0 (0.00)	0.00 (0.00-.)	0.98	4	0 (0.00)	0.00 (0.00-.)	0.98
	C/T + C/C	107	56 (52.34)	1.08 (0.78–1.49)	0.64	77	33 (42.86)	1.58 (1.02–2.44)	**0.04**	77	39 (50.65)	1.61 (1.08–2.40)	**0.02**
NLRP12	C/C	225	112 (49.78)	1		161	60 (37.27)	1		161	71 (44.10)	1	
rs34436714	A/C	113	50 (44.25)	0.93 (0.66–1.30)	0.67	83	23 (27.71)	0.72 (0.45–1.17)	0.19	83	26 (31.33)	0.67 (0.43–1.06)	0.09
42: K<>N	A/A	16	6 (37.50)	0.76 (0.33–1.73)	0.51	13	3 (23.08)	0.51 (0.16–1.65)	0.26	13	3 (23.08)	0.47 (0.15–1.49)	0.2
	A/C + A/A	129	56 (43.41)	0.91 (0.65–1.26)	0.55	96	26 (27.08)	0.69 (0.43–1.10)	0.12	96	29 (30.21)	0.64 (0.42–0.99)	**0.05**
NLRC5	C/C	192	101 (52.60)	1		133	53 (39.85)	1		133	61 (45.86)	1	
rs289723	A/C	149	65 (43.62)	0.87 (0.64–1.20)	0.4	113	33 (29.20)	0.72 (0.46–1.11)	0.14	113	37 (32.74)	0.61 (0.41–0.93)	**0.02**

(*Continued*)

Table 2. (Continued)

Gene		Overall Survival (pM = 0&1)				Overall Survival (pM = 0)				Event-free Survival (pM = 0)			
SNP	**Genotype**	**Cases**	**Death (%)**	**HR (95%CI)**	**p-val**	**Cases**	**Deaths (%)**	**HR (95%CI)**	**p-val**	**Cases**	**Events (%)**	**HR (95%CI)**	**p-val**
1105: Q<>K	A/A	27	11 (40.74)	0.99 (0.53–1.86)	0.98	23	8 (34.78)	1.13 (0.53–2.41)	0.75	23	10 (43.48)	1.04 (0.53–2.04)	0.92
	A/C + A/A	176	76 (43.18)	0.89 (0.66–1.20)	0.44	136	41 (30.15)	0.77 (0.51–1.17)	0.22	136	47 (34.56)	0.67 (0.46–0.99)	**0.04**
NLRC5	C/C	282	128 (45.39)	1		204	64 (31.37)	1		204	73 (35.78)	1	
rs74439742	C/T	75	43 (57.33)	1.42 (1.00–2.00)	**0.05**	56	26 (46.43)	1.67 (1.05–2.65)	**0.03**	56	31 (55.36)	1.63 (1.06–2.49)	**0.03**
191: P<>L	T/T	10	7 (70.00)	2.39 (1.10–5.19)	**0.03**	7	4 (57.14)	3.07 (1.07–8.80)	**0.04**	7	4 (57.14)	2.71 (0.96–7.67)	0.06
	C/T + T/T	85	50 (58.82)	1.50 (1.08–2.09)	**0.02**	63	30 (47.62)	1.77 (1.13–2.75)	**0.01**	63	35 (55.56)	1.70 (1.13–2.57)	**0.01**
No. of	0–5	97	34 (35.05)	1	-	75	15 (20.00)	1	-	75	16 (21.33)	1	-
risk alleles	6–7	110	52 (47.27)	1.43 (0.92–2.21)	0.11	79	26 (32.91)	1.53 (0.81–2.90)	0.19	79	32 (40.51)	2.06 (1.12–3.77)	**0.02**
	8–12	108	62 (57.41)	1.88 (1.23–2.86)	**0.003**	75	33 (44.00)	2.89 (1.55–5.37)	**0.0008**	75	37(49.33)	3.02 (1.66–5.48)	**0.0003**

https://doi.org/10.1371/journal.pone.0199350.t002

genotypes, DACHS data according to allelic probabilities. Despite the promising initial results, neither the associations for CRC risk nor the associations for CRC survival were replicated in the GWAS sets. We also tested the additive models in the DACHS population, but no association was evident (data not shown).

Divergent expression patterns of NLRP2, 5, 6 and 13 in hematopoietic and non-hematopoietic cells

To investigate whether the NLRs found to be associated with CRC in the Czech discovery set, were expressed in the gut or immune cells, mRNA levels were quantified for selected NLRs in primary tissue samples and cell lines. In gut-related tissues, CRC cell lines, whole blood and neutrophils (PMN), *NLRP2* showed moderate expression, with the lowest expression in rectum (Fig 2A and 2B). Although *NLRP2* showed low expression in CD14$^+$ monocytes, this was increased up to 100-fold in monocyte-derived primary dendritic cells and macrophages (Fig 2B). *NLRP5* (also known as *MATER*) was not detectable in immune cells (not shown) and normal gut tissue but in ovary (Fig 2C), in keeping with its role in oogenesis [8]. However, we observed expression in a primary CRC sample and three CRC cell lines (Fig 2C) but not fibroblast or B lymphocyte cell lines (not shown), in agreement with mRNA expression data from the GENT database (Figure A in S1 File). Consistently with an earlier report [32], *NLRP6* (also known as *PYPAF5*) was highly expressed in neutrophils (PMN), low in monocytes and MoDC but not detectable in gut biopsies (not shown). *NLRP13* was below the level of detection in all analysed samples except for the positive control, ovary and the DLD1 CRC cell line (Fig 2D). *NLRC5* was expressed to varying degrees in healthy gut tissue and most highly in colon adenocarcinoma, and was inducible by IFNγ in HCT116 cells (Fig 2E). *NLRP3* was expressed at considerable levels only in duodenum, rectum and CRC samples, but not in normal ileum and colon biopsies (Fig 2F). While the role of *NLRP13* remains unclear, additional data on the reported occurrence of somatic mutations in these genes in CRC suggest that *NLRP2*, *NLRP3* and *NLRP6* may impact CRC development and survival via immune cells, whereas *NLRP5* might be relevant in gut tissues themselves, possibly experiencing a re-expression after malignant transformation.

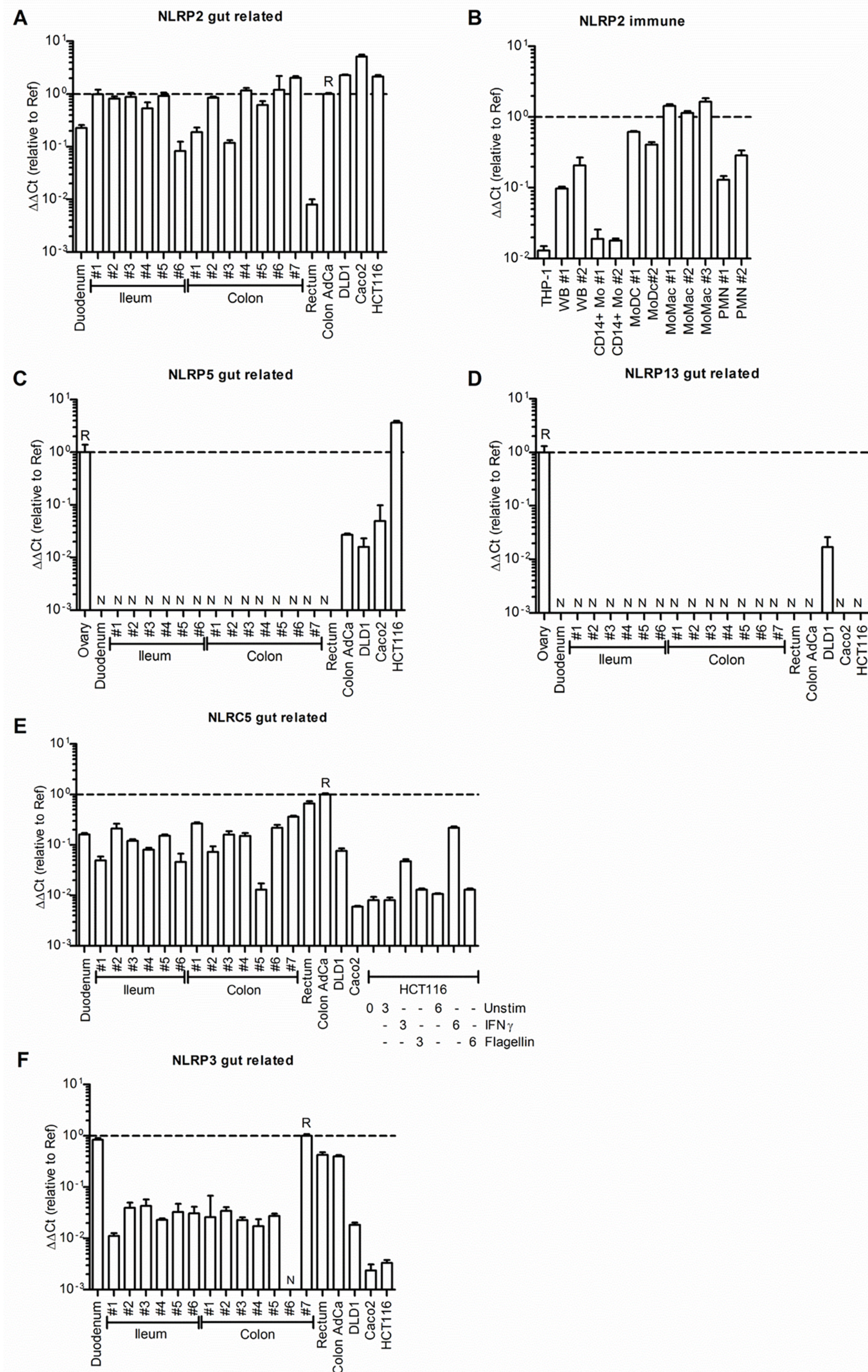


Fig 2. Expression of selected CRC-associated NLRs in immune cells, primary tissue samples or CRC cell lines. mRNA expression of NLRP2 (A,B), NLRP5 (C), NLRP13 (D), NLRC5 (E) and NLRP3 (F) was determined relative to the housekeeper TBP by performing triplicate (means +SD show) qPCR using TaqMan gene-specific primers and probes on the indicated samples (see Methods). In the case of (E) HCT cells were treated with 1000 U/ml IFNγ or 50 ng/ml S. typhimurium Flagellin for 3 or 6 hours as indicated. TBP-relative ΔCt values were normalized to a reference sample (labelled R, ΔΔCt method). N denotes samples in which no expression was detectable above Ct within 40 cycles.

https://doi.org/10.1371/journal.pone.0199350.g002

Discussion

In the discovery set from the Czech Republic, five of 39 successfully tested SNPs were associated with CRC risk, and eight with CRC survival. An additive effect on CRC risk and survival was detected, resulting in a 2-fold increased risk and a 3-fold worse survival for carriers of $\geq$6 and $\geq$8 risk alleles, respectively. Despite these promising results in the Czech population, these associations could not be confirmed in the two large German and Scottish GWAS data sets.

This was surprising taking into account the *in silico* predictions about the functionality of the SNPs and the results of the expression analysis which showed that the genes *NLRP2*, *NLRP3* and *NLRP6* may impact CRC development via immune cells. Accordingly, differential expression of these genes may cause alterations in pathways providing the emerging hallmarks of cancer, such as evading immune clearance and tumor-promoting inflammatory responses [33]. For *NLRC5*, whose expression was induced in HCT116 cells by IFNγ (Fig 2E), one plausible functional outcome may be the modulation of MHC class I expression [34]. The latter strongly correlates with CRC survival due to its effect on CD8 cytotoxic T cell and natural killer cell immuno-surveillance [35]. According to murine data, the *NLRP12* may also affect T cell function in the context of human CRC [36]. Most intriguingly, expression of development-related *NLRP5* was undetectable in normal gut-related tissues but was up-regulated in malignant gut tissue and colon cancer cell lines (*cf*. Fig 2C and Figure A in S1 File.), suggesting for the first time a potential novel role beyond developmental control for this enigmatic NLR in humans [8]. During oogenesis, murine Nlrp5 appears to influence mitochondrial localization and activity, ATP content and Ca^{2+} homeostasis–processes which all have been linked to NLRP3 inflammasome activation and thus inflammation in differentiated cells. NLRP5 may thus act in concert with NLRP3, which is known to be associated with human CRC [14], a speculation warranting further investigation. The concerted association of genes of the NLR family may directly link environmental risk factors, intestinal inflammation, the microbiota and well-described cancer pathways involved in CRC development, such as the MAPK pathway and the NF-κB pathway.

One might argue that the failure to replicate the association results in the Czech discovery set might be due to differences in the clinical composition between the case-control populations of the discovery set and the replication sets (Table B in S2 File). However, data was adjusted for all significantly different parameters except tumor location (colon or rectum; not possible due to incomplete data) suggesting that the detected associations in the discovery set were false positive results. In the light of the supporting gene expression data it is possible that the coding variants analyzed in this study do actually not have an effect on the functionality of the receptor proteins. This assumption is supported by the fact that the majority of variants are not located in evolutionary conserved regions of the genes which allow for natural variability. Further, it is also possible that undetected environmental factors might have biased the results. Especially for immune related genetic variants, interactions with environmental factors or treatment might play a major role enhancing or even enabling effects of SNPs on CRC risk or survival. Based on the interesting expression results, future studies of these genes and their encoded receptors, including the analysis of regulatory genetic variants affecting the gene expression as well as the analysis of the patient specific tumor microenvironment and tumor infiltrating immune cells and immune constitution, may contribute to uncover the still poorly understood role of NLRs within the intestinal immune system, as well as, in CRC development and survival [37]. The integration of different exogenous, endogenous, tumour and immune factors, potentially including the variants in NLR genes studied here, holds promise for future approaches in precision medicine [37].

Supporting information

S1 File. (Figure A) mRNA expression of NLRP5 (MATER) in CRC.
(DOCX)

S2 File. (Table A) Complete list of genotyped SNPs in candidate genes, with information about all linked missense SNPs ($r^2 \geq 0.8$) and the location in protein domains. NMD: non-sense mediated decay; * Variant Effect Predictor by Ensembl http://www.ensembl.org/Homo_sapiens/Tools/VEP (Table B) Population Description. a Z statistics: Wilcoxon Rank-Sum-Test; b Chi-square; event = recurrence, metastasis, death. (Table C) Genotype distribution of all analysed SNPs in the Czech case-control population: Risk and Survival analysis. CRC Risk: Data adjusted for age of diagnosis and sex. Overall Survival and Event free Survival: Data adjusted for age of diagnosis and sex, grade and stage. Nominal significance at $p \leq 0.05$; significance level corrected for multiple testing at $p \leq 0.001$. (Table D) mRNA Expression for the most promising candidate genes: Study data and reported somatic mutations for CRC-associated NLRs. (Table E) CRC risk and Overall survival pM0: Comparison of the SNPs with $p \leq 0.05$ in the Czech discovery set with GWAS results from the Scottish and DACHS replication sets. Amino acid changes are given as <> with the amino acid position indicated. Nominal significance at $p \leq 0.05$.
(DOCX)

Author Contributions

Conceptualization: Stefanie Huhn, Alexander N. R. Weber, Asta Försti.

Data curation: Stefanie Huhn, Tharmila Sanmuganantham, Tica Pichulik, Calogerina Catalano, Barbara Pardini, Alessio Naccarati, Veronika Polakova-Vymetálkova, Katerina Jiraskova, Ludmila Vodickova, Pavel Vodicka, Markus W. Löffler, Lioba Courth, Farhat V. N. Din, Maria Timofeeva, Susan M. Farrington, Lina Jansen.

Formal analysis: Stefanie Huhn, Miguel I. da Silva Filho.

Funding acquisition: Pavel Vodicka, Jan Wehkamp, Kari Hemminki, Jenny Chang-Claude, Hermann Brenner, Michael Hoffmeister, Malcolm G. Dunlop, Alexander N. R. Weber.

Investigation: Stefanie Huhn, Tharmila Sanmuganantham, Tica Pichulik, Calogerina Catalano, Markus W. Löffler, Lioba Courth, Farhat V. N. Din, Maria Timofeeva, Susan M. Farrington, Lina Jansen.

Methodology: Stefanie Huhn, Tharmila Sanmuganantham, Tica Pichulik, Markus W. Löffler, Lioba Courth, Farhat V. N. Din, Maria Timofeeva, Susan M. Farrington, Lina Jansen.

Project administration: Stefanie Huhn, Pavel Vodicka, Jan Wehkamp, Kari Hemminki, Jenny Chang-Claude, Hermann Brenner, Michael Hoffmeister, Malcolm G. Dunlop, Alexander N. R. Weber, Asta Försti.

Resources: Stefanie Huhn, Pavel Vodicka, Jan Wehkamp, Kari Hemminki, Jenny Chang-Claude, Hermann Brenner, Michael Hoffmeister, Malcolm G. Dunlop, Alexander N. R. Weber, Asta Försti.

Supervision: Ludmila Vodickova, Pavel Vodicka, Kari Hemminki, Jenny Chang-Claude, Hermann Brenner, Michael Hoffmeister, Malcolm G. Dunlop, Alexander N. R. Weber, Asta Försti.

Visualization: Stefanie Huhn, Miguel I. da Silva Filho.

Writing – original draft: Stefanie Huhn, Alexander N. R. Weber.

Writing – review & editing: Stefanie Huhn, Miguel I. da Silva Filho, Tharmila Sanmuganantham, Tica Pichulik, Calogerina Catalano, Barbara Pardini, Alessio Naccarati, Veronika Polakova-Vymetálkova, Katerina Jiraskova, Ludmila Vodickova, Pavel Vodicka, Markus W. Löffler, Lioba Courth, Jan Wehkamp, Farhat V. N. Din, Maria Timofeeva, Susan M. Farrington, Lina Jansen, Kari Hemminki, Jenny Chang-Claude, Hermann Brenner, Michael Hoffmeister, Malcolm G. Dunlop, Alexander N. R. Weber, Asta Försti.

References

1. Kawai T, Akira S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. Immunity. 2011; 34(5):637–50. Epub 2011/05/28. https://doi.org/10.1016/j.immuni.2011.05.006 PMID: 21616434.
2. Oviedo-Boyso J, Bravo-Patino A, Baizabal-Aguirre VM. Collaborative action of Toll-like and NOD-like receptors as modulators of the inflammatory response to pathogenic bacteria. Mediators of inflammation. 2014; 2014:432785. https://doi.org/10.1155/2014/432785 PMID: 25525300; PubMed Central PMCID: PMC4267164.
3. Wilmanski JM, Petnicki-Ocwieja T, Kobayashi KS. NLR proteins: integral members of innate immunity and mediators of inflammatory diseases. J Leukoc Biol. 2008; 83(1):13–30. Epub 2007/09/19. doi: jlb.0607402 [pii] https://doi.org/10.1189/jlb.0607402 PMID: 17875812.
4. Abreu MT. Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function. Nature reviews Immunology. 2010; 10(2):131–44. Epub 2010/01/26. https://doi.org/10.1038/nri2707 PMID: 20098461.
5. Carvalho FA, Aitken JD, Vijay-Kumar M, Gewirtz AT. Toll-like receptor-gut microbiota interactions: perturb at your own risk! Annu Rev Physiol. 2012; 74:177–98. Epub 2011/11/01. https://doi.org/10.1146/annurev-physiol-020911-153330 PMID: 22035346.
6. Weber AN, Forsti A. Toll-like receptor genetic variants and colorectal cancer. Oncoimmunology. 2014; 3(1):e27763. Epub 2014/05/03. https://doi.org/10.4161/onci.27763 PMID: 24790794; PubMed Central PMCID: PMC4004619.
7. Kumar H, Kawai T, Akira S. Pathogen recognition in the innate immune response. Biochem J. 2009; 420(1):1–16. Epub 2009/04/23. doi: BJ20090272 [pii] https://doi.org/10.1042/BJ20090272 PMID: 19382893.
8. Lupfer C, Kanneganti TD. Unsolved Mysteries in NLR Biology. Front Immunol. 2013; 4:285. Epub 2013/09/26. https://doi.org/10.3389/fimmu.2013.00285 PMID: 24062750; PubMed Central PMCID: PMC3775000.
9. Fernandes R, Tsuda C, Perumalsamy AL, Naranian T, Chong J, Acton BM, et al. NLRP5 mediates mitochondrial function in mouse oocytes and embryos. Biol Reprod. 2012; 86(5):138, 1–10. https://doi.org/10.1095/biolreprod.111.093583 PMID: 22357545; PubMed Central PMCID: PMC3364921.
10. Castro FA, Forsti A, Buch S, Kalthoff H, Krauss C, Bauer M, et al. TLR-3 polymorphism is an independent prognostic marker for stage II colorectal cancer. Eur J Cancer. 2011; 47(8):1203–10. https://doi.org/10.1016/j.ejca.2010.12.011 PMID: 21239167.
11. Klimosch SN, Forsti A, Eckert J, Knezevic J, Bevier M, von Schonfels W, et al. Functional TLR5 genetic variants affect human colorectal cancer survival. Cancer Res. 2013; 73(24):7232–42. https://doi.org/10.1158/0008-5472.CAN-13-1746 PMID: 24154872.
12. Wang H, Flannery SM, Dickhofer S, Huhn S, George J, Kubarenko AV, et al. A coding IRAK2 protein variant compromises Toll-like receptor (TLR) signaling and is associated with colorectal cancer survival. J Biol Chem. 2014; 289(33):23123–31. https://doi.org/10.1074/jbc.M113.492934 PMID: 24973222; PubMed Central PMCID: PMC4132810.
13. Wehkamp J, Salzman NH, Porter E, Nuding S, Weichenthal M, Petras RE, et al. Reduced Paneth cell alpha-defensins in ileal Crohn's disease. Proc Natl Acad Sci U S A. 2005; 102(50):18129–34. https://doi.org/10.1073/pnas.0505256102 PMID: 16330776; PubMed Central PMCID: PMC1306791.
14. Ungerbäck J, Belenki D, Jawad ul-Hassan A, Fredrikson M, Fransen K, Elander N, et al. Genetic variation and alterations of genes involved in NFkappaB/TNFAIP3- and NLRP3-inflammasome signaling affect susceptibility and outcome of colorectal cancer. Carcinogenesis. 2012; 33(11):2126–34. Epub 2012/07/31. https://doi.org/10.1093/carcin/bgs256 PMID: 22843550.
15. Cooper GM, Stone EA, Asimenos G, Program NCS, Green ED, Batzoglou S, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005; 15(7):901–13. https://doi.org/10.1101/gr.3577405 PMID: 15965027; PubMed Central PMCID: PMC1172034.

16. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005; 15(8):1034–50. https://doi.org/10.1101/gr.3715005 PMID: 16024819; PubMed Central PMCID: PMC1182216.
17. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010; 20(1):110–21. https://doi.org/10.1101/gr.097857.109 PMID: 19858363; PubMed Central PMCID: PMC2798823.
18. Naccarati A, Pardini B, Stefano L, Landi D, Slyskova J, Novotny J, et al. Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. Carcinogenesis. 2012; 33 (7):1346–51. Epub 2012/05/15. https://doi.org/10.1093/carcin/bgs172 PMID: 22581836.
19. Brenner H, Chang-Claude J, Jansen L, Knebel P, Stock C, Hoffmeister M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. Gastroenterology. 2014; 146(3):709–17. https://doi.org/10.1053/j.gastro.2013.09.001 PMID: 24012982.
20. Rudolph A, Hein R, Hoffmeister M, Forsti A, Hemminki K, Risch A, et al. Copy number variations of GSTT1 and GSTM1, colorectal cancer risk and possible effect modification of cigarette smoking and menopausal hormone therapy. Int J Cancer. 2012; 131(5):E841–8. https://doi.org/10.1002/ijc.27428 PMID: 22234881.
21. Theodoratou E, Farrington SM, Tenesa A, McNeill G, Cetnarskyj R, Barnetson RA, et al. Modification of the inverse association between dietary vitamin D intake and colorectal cancer risk by a FokI variant supports a chemoprotective action of Vitamin D intake mediated through VDR binding. Int J Cancer. 2008; 123(9):2170–9. https://doi.org/10.1002/ijc.23769 PMID: 18709640.
22. Timofeeva MN, Kinnersley B, Farrington SM, Whiffin N, Palles C, Svinti V, et al. Recurrent Coding Sequence Variation Explains Only A Small Fraction of the Genetic Architecture of Colorectal Cancer. Sci Rep. 2015; 5:16286. https://doi.org/10.1038/srep16286 PMID: 26553438; PubMed Central PMCID: PMC4639776.
23. Hoffmeister M, Jansen L, Rudolph A, Toth C, Kloor M, Roth W, et al. Statin use and survival after colorectal cancer: the importance of comprehensive confounder adjustment. J Natl Cancer Inst. 2015; 107 (6):djv045. https://doi.org/10.1093/jnci/djv045 PMID: 25770147.
24. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013; 144(4):799–807 e24. https://doi.org/10.1053/j.gastro.2012.12.020 PMID: 23266556; PubMed Central PMCID: PMC3636812.
25. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic epidemiology. 2010; 34(8):816–34. https://doi.org/10.1002/gepi.20533 PMID: 21058334; PubMed Central PMCID: PMC3175618.
26. Barrett JH, Smith G, Waxman R, Gooderham N, Lightfoot T, Garner RC, et al. Investigation of interaction between N-acetyltransferase 2 and heterocyclic amines as potential risk factors for colorectal cancer. Carcinogenesis. 2003; 24(2):275–82. PMID: 12584178.
27. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. International journal of epidemiology. 2013; 42(3):689–700. https://doi.org/10.1093/ije/dys084 PMID: 22786799.
28. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC Med Genet. 2006; 7:74. https://doi.org/10.1186/1471-2350-7-74 PMID: 17014726; PubMed Central PMCID: PMC1592477.
29. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. International journal of epidemiology. 2012; 41(6):1576–84. https://doi.org/10.1093/ije/dyr197 PMID: 22253310.
30. Quanto. Quanto http://hydra.usc.edu/gxe/. Available from: http://hydra.usc.edu/gxe/.
31. Klimosch SN, Forsti A, Eckert J, Knezevic J, Bevier M, von Schonfels W, et al. Functional TLR5 genetic variants affect human colorectal cancer survival. Cancer research. 2013; 73(24):7232–42. Epub 2013/10/25. https://doi.org/10.1158/0008-5472.CAN-13-1746 PMID: 24154872.
32. Grenier JM, Wang L, Manji GA, Huang WJ, Al-Garawi A, Kelly R, et al. Functional screening of five PYPAF family members identifies PYPAF5 as a novel regulator of NF-kappaB and caspase-1. FEBS letters. 2002; 530(1–3):73–8. Epub 2002/10/22. PMID: 12387869.
33. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144(5):646–74. Epub 2011/03/08. https://doi.org/10.1016/j.cell.2011.02.013 PMID: 21376230.
34. Neerincx A, Castro W, Guarda G, Kufer TA. NLRC5, at the Heart of Antigen Presentation. Front Immunol. 2013; 4:397. Epub 2013/12/10. https://doi.org/10.3389/fimmu.2013.00397 PMID: 24319445; PubMed Central PMCID: PMC3837245.

35. Watson NF, Ramage JM, Madjd Z, Spendlove I, Ellis IO, Scholefield JH, et al. Immunosurveillance is active in colorectal cancer as downregulation but not complete loss of MHC class I expression correlates with a poor prognosis. Int J Cancer. 2006; 118(1):6–10. Epub 2005/07/09. https://doi.org/10.1002/ijc.21303 PMID: 16003753.

36. Lukens JR, Gurung P, Shaw PJ, Barr MJ, Zaki MH, Brown SA, et al. The NLRP12 Sensor Negatively Regulates Autoinflammatory Disease by Modulating Interleukin-4 Production in T Cells. Immunity. 2015; 42(4):654–64. Epub 2015/04/19. https://doi.org/10.1016/j.immuni.2015.03.006 S1074-7613(15)00120-X [pii]. PMID: 25888258; PubMed Central PMCID: PMC4412374.

37. Ogino S, Nowak JA, Hamada T, Phipps AI, Peters U, Milner DA Jr., et al. Integrative analysis of exogenous, endogenous, tumour and immune factors for precision medicine. Gut. 2018; 67(6):1168–80. https://doi.org/10.1136/gutjnl-2017-315537 PMID: 29437869; PubMed Central PMCID: PMC5943183.

<u>Manuscript V</u>

Catalano C, da Silva Filho MI, Frank C, **Jiraskova K**, Vymetalkova V, Levy M, Liska V, Vycital O, Naccarati A, Vodickova L, Hemminki K, Vodicka P, Weber ANR, Försti A.

Investigation of single and synergic effects of NLRC5 and PD-L1 variants on the risk of colorectal cancer

PLoS One (2018) 13(2)

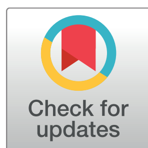
RESEARCH ARTICLE

Investigation of single and synergic effects of *NLRC5* and *PD-L1* variants on the risk of colorectal cancer

Calogerina Catalano[1]*, Miguel Inacio da Silva Filho[1], Christoph Frank[1], Katerina Jiraskova[2,3], Veronika Vymetalkova[2,3], Miroslav Levy[4], Vaclav Liska[4,5], Ondrej Vycital[4,5], Alessio Naccarati[2,6], Ludmila Vodickova[2,3,5], Kari Hemminki[1,7], Pavel Vodicka[2,3,5], Alexander N. R. Weber[8], Asta Försti[1,7]

1 Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, **2** Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic, **3** Institute of Biology and Medical Genetics, 1st Medical Faculty, Charles University, Prague, Czech Republic, **4** Department of Surgery, First Medical Faculty, Charles University and Thomayer Hospital, Prague, Czech Republic, **5** Biomedical Center, Faculty of Medicine in Pilsen, Charles University Prague, Pilsen, Czech Republic, **6** Molecular and Genetic Epidemiology, Italian Institute for Genomic Medicine (IIGM), Turin, Italy, **7** Center for Primary Health Care Research, Clinical Research Center, Lund University, Malmö, Sweden, **8** Department of Immunology, Interfaculty Institute for Cell Biology, University of Tübingen, Tübingen, Germany

* c.catalano@dkfz.de



OPEN ACCESS

Citation: Catalano C, da Silva Filho MI, Frank C, Jiraskova K, Vymetalkova V, Levy M, et al. (2018) Investigation of single and synergic effects of *NLRC5* and *PD-L1* variants on the risk of colorectal cancer. PLoS ONE 13(2): e0192385. https://doi.org/10.1371/journal.pone.0192385

Editor: Valli De Re, Istituto di Ricovero e Cura a Carattere Scientifico Centro di Riferimento Oncologico della Basilicata, ITALY

Received: October 24, 2017

Accepted: January 20, 2018

Published: February 6, 2018

Copyright: © 2018 Catalano et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The work was supported by Internal Grant Agency of the Ministry of Health of the Czech Republic AZV 15-27580A (PI: PV) and AZV 15-26535A (PI: VV) (http://www.azvcr.cz/o-azv-cr) and the Grant Agency of the Charles University (GAUK 112515) (PI: KJ) (http://www.cuni.cz/ukeng-33.html). AW was supported by the University of

Abstract

Constitutive activation of interferon signaling pathways has been reported in colorectal cancer (CRC), leading to a strong $CD8^+$ T cell response through stimulation of *NLRC5* expression. Primed $CD8^+$ T cell expansion, however, may be negatively regulated by *PD-L1* expression. Additionally, aberrant *PD-L1* expression enables cancer cells to escape the immune attack. Our study aimed to select potential regulatory variants in the *NLRC5* and *PD-L1* genes by using several online *in silico* tools, such as UCSC browser, HaploReg, Regulome DB, Gtex Portal, microRNA and transcription factor binding site prediction tools and to investigate their influence on CRC risk in a Czech cohort of 1424 CRC patients and 1114 healthy controls. Logistic regression analysis adjusted for age and gender reported a moderate association between rectal cancer risk and two *NLRC5* SNPs, rs1684575 T>G (OR: 1.60, 95% CI: 1.13–2.27, recessive model) and rs3751710 (OR: 0.70, 95% CI: 0.51–0.96, dominant model). Given that a combination of genetic variants, rather than a single polymorphism, may explain better the genetic etiology of CRC, we studied the interplay between the variants within *NLRC5*, *PD-L1* and the previously genotyped *IFNGR1* and *IFNGR2* variants, to evaluate their involvement in the risk of CRC development. Overall we obtained 18 pairwise interactions within and between the *NLRC5* ad *PD-L1* genes and 6 more when *IFNGR* variants were added. Thirteen out of the 24 interactions were below the threshold for the FDR calculated and controlled at an arbitrary level q*<0.10. Furthermore, the interaction *IFNGR2* rs1059293 C>T—*NLRC5* rs289747 G>A (P<0.0001) remained statistically significant even after Bonferroni correction. Our data suggest that not only a single genetic variant but also an interaction between two or more variants within genes involved in immune regulation may play important roles in the onset of CRC, providing therefore novel biological

Tübingen, the Federal State Baden-Württemberg Junior Professor Program (https://www.medizin.uni-tuebingen.de/en/Research/Research+Funding/Gender+Equality+Prize.html). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AIC, Akaike information criterion; CI, confidence interval; CRC, Colorectal cancer; HWE, Hardy-Weinberg equilibrium; IBD, inflammatory bowel disease; IFN, interferon; IFNγ, Interferon gamma; IFNGR, interferon gamma receptor; LD, linkage disequilibrium; MHC, major histocompatibility complex; MAF, minor allele frequency; Nf-kB, nuclear factor kappa B; NLRC5, NLR family, CARD domain containing 5; NLRs, Nod-like receptors; OR, Odds ratio; PD-1, programmed death protein 1; PD-L1, programmed death-ligand 1; PRRs, pattern recognition receptors; SNPs, single nucleotide polymorphisms; TAF, TATA-box binding protein associated factor 1; TLRs, Toll-like receptors; TSS, transcription start site; UTR, untranslated regions.

information, which could eventually improve CRC risk management but also PD-1-based immunotherapy in CRC.

Introduction

Colorectal cancer (CRC) is the third most common cancer and the fourth leading cause of cancer mortality worldwide [1]. CRC represents a paradigm for the link between inflammation and cancer [2]. The intestinal tract is continuously exposed to both potential pathogens and beneficial commensal microorganisms; therefore the homeostatic balance between tolerance and immunity represents a regulatory challenge to the mucosal immune system [3]. In this context a pivotal role is played by the epithelial cells that monitor the intestinal microenvironment for pathogenic and commensal microorganisms via so-called pattern recognition receptors (PRRs), e.g. Toll-like receptors (TLRs), and in turn influence the function of antigen presenting cells and lymphocytes [3,4]. Additionally, the gut microbiota provides crucial health benefits to its host by contributing to the regulation of the intestinal immune homeostasis [3, 5]. Recently, it has become obvious that alterations of the regulatory pathways that maintain this homeostasis can result in the development of local and chronic inflammation, inflammatory bowel disease (IBD) and CRC [5]. Aberrant activation of nuclear factor kappa B (Nf-kB) and interferon (IFN) signaling pathways have been reported to play a pivotal role in CRC by triggering the production of several proinflammatory mediators [6–8]. Particularly IFNγ signaling pathway is known to play an important role in controlling the $CD8^+$ T cell expansion through the stimulation of *NLRC5* (NOD-like receptor C5) expression, a major histocompatibility complex (MHC) class I transactivator [9,10]. *NLRC5* is a member of the Nod-like receptor (NLR) family of PRR proteins. It contains a nucleotide-binding domain and leucine-rich repeats, which are conserved in PRRs that regulate inflammatory responses and cell death. Given its role in the transcription of MHC class I genes, it is reasonable to think that NLRC5 may play a prominent role in antitumor immunity and its loss may promote tumor immune evasion [11]. Moreover, cytokine production and $CD8^+$ T cell expansion is necessary for generating an effective immune defense against invading harmful pathogens [12].

By assuming the importance of a balanced immune response, a physiological feedback mechanism played by PD-L1 (Programmed death-ligand 1) is necessary for terminating the immune responses in a proper way and for maintaining self- tolerance [13]. However, it has recently been shown that IFNγ is also involved in promoting *PD-L1* expression in tumor cells [11]. This results in an aberrant *PD-L1* expression that allows cancer cells to escape the antitumor immune response by suppressing the $CD8^+$ T cell expansion [13–15]. This escape mechanism is reversed by immune checkpoint inhibitors targeting the PD-1/PD-L1 interaction and restores anti-tumor immunity [16]. Thus the possibility of PD-1/PD-L1-based therapies has received much attention in many tumor entities including CRC.

To gain further evidence about the potential role of SNPs within *NLRC5* and *PD-L1* genes, we genotyped a set of 16 potential regulatory single nucleotide polymorphisms (SNPs) in a case-control study of 1424 CRC patients and 1114 healthy controls from the Czech Republic and evaluated their association with CRC risk. Moreover, given the opposite actions of these two proteins on the $CD8^+$ T cell expansion, we investigated whether pair-wise interactions between all the investigated SNPs and the previously genotyped SNPs in the *IFNGR1* and *IFNGR2* genes exist [17], which may have interactive effects on the risk of CRC. This strategy has the potential to identify complex biological links among cancer-related immunity genes

and processes they are involved in, and could provide novel information for a better basic understanding, risk-management and therapy of CRC.

Materials and methods

Ethics statement

Written informed consent was given by all participants enrolled in the current research study in accordance with the Helsinki declaration. The project was approved by the ethical committees of the participating institutes, the Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic (who issued the Institutional Certification for Multicenter Studies on July 16th 2015 covering all studies between 2004–2015) and the Institute for Clinical and Experimental Medicine and Faculty Thomayer Hospital, Prague, Czech Republic (786/09 (G09-04-09) and 622/11(G11-04-09)).

Study population

The case group contained 1424 CRC patients recruited between the years 2004 and 2013 by several oncological departments in the Czech Republic (Table 1)[17]. Their mean age was 62.7 years, and 61.8% of them were men. The patients showed positive colonoscopic results for malignancy, histologically confirmed as colon or rectal carcinomas. Patients with any previous history of cancer or who met the Amsterdam criteria I or II for hereditary nonpolyposis colorectal cancer were not included in the study. General information about gender and age at diagnosis was available for all patients. The control group contained 1114 healthy individuals recruited by the blood-donor centers in Kralovske Vinohrady Hospital and Vojkov hospital in Prague [17,18]. Their mean age was 47.1 years, and 53.4% of them were men.

SNP selection

A total of 16 SNPs, which captured 32 potential regulatory SNPs ($r^2 > 0.89$), were selected for genotyping within the *NLRC5* (NLR family, CARD domain containing 5) and *CD274* (also known as PD-L1, programmed death ligand 1) genes according to the following selection criteria: non-coding SNPs in the 5’ flanking region (up to 1kb from the transcription start site (TSS) containing the promoter, enhancer or other transcription factor binding sites), 5′ and 3′ untranslated regions (UTRs), and SNPs regulating the expression of the selected genes (eQTL

Table 1. Characteristics of the colorectal cancer patients.

CRC risk analysis		Cases	Controls	p-value
All patients		1424	1114	
Age at diagnosis	Mean (range)	62.7 (24–90)	47.1 (18–94)	**< .0001**[a]
	Median	63	47	
Gender	Male	880 (61.8%)	595 (53.4%)	**2.6e-05**[b]
	Female	544 (38.2%)	519 (46.6%)	
Tumour location	-	-		
	Colon	889 (62.4%)		
	Rectum	398 (27.9%)		
	missing information	137 (9.6%)		

[a]: Z statistics: Wilcoxon Rank-Summ-Test;

[b]: Chi-square.

$P < 0.05$ are in bold.

https://doi.org/10.1371/journal.pone.0192385.t001

SNPs) with a minor allele frequency (MAF) $\geq$ 0.10 in the CEU population validated by 1000 Genomes and with a pairwise linkage disequilibrium (LD) $r^2 \leq 0.80$ (S1 Table).

In-silico analysis

SNPs were selected using several in silico tools, such as UCSC browser (https://genome-euro.ucsc.edu/) to collect all potential functional SNPs in the regulatory regions, HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) and Regulome DB (http://www.regulomedb.org/) to explore the chromatin state, conservation, and regulatory motif alterations within sets of genetically linked variants, Gtex Portal (https://gtexportal.org/home/) to identify all cis-eQTL SNPs that affect the expression of genes of our interest and microRNA binding site prediction tools (http://www.microrna.org/microrna/home.do, http://epicenter.iefreiburg.mpg.de/services/microsniper/) to investigate the 3'-UTR and to predict if a SNP within the target site will disrupt/eliminate or enhance/create a microRNA binding site. PERFECTOS-APE (http://opera.autosome.ru/perfectosape/scan) and s-Transcription factor Affinity Prediction (s-TRAP, http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi) were used to identify transcription factors whose binding sites can be significantly affected by a given polymorphism. LD and the haplotype blocks within the genes were examined based on r^2 (S1 Table).

Genotyping

In this project, genomic DNA from peripheral blood leukocytes was used. The KASP (LGC genomics, Hoddesdon, Hertfordshire, UK) and the TaqMan (Thermo Fisher Scientific, Darmstadt, Germany) allelic discrimination methods were used to genotype the selected SNPs. The genotyping was performed blinded by the case–control status of each sample. DNA amplification was performed according to the LGC genomics' and TaqMan´s PCR conditions. Genotype detection was performed using ViiA 7 Real-Time PCR System (Thermo Fisher Scientific). The sample set contained 142 duplicated samples as quality controls. The genotype correlation between the duplicate samples was > 90%. Genotype call rate ranged between 94.0 and 100%.

Statistical analysis

The observed genotype frequencies in the controls were tested for Hardy–Weinberg equilibrium (HWE) using the chi-square test. Odds ratios (ORs) and 95% confidence intervals (CIs) for associations between genotypes and CRC risk were calculated by logistic regression (SAS Version 9.3; SAS Institute, Cary, NC), and adjusted for age and gender. The estimated power was >98% for OR $\geq$ 1.5 (MAF > 0.10; p = 0.05; dominant model) and >98% for OR $\geq$ 1.5 (MAF > 0.50; p = 0.05; recessive model) (Quanto: http://hydra.usc.edu/gxe/).

All possible SNP combinations were evaluated in binary interaction to find the SNP-SNP interactions that best predict the disease risk. In addition to the SNPs genotyped in the current study, we also included all SNPs in the *IFNGR1/2* genes genotyped previously in 1327 CRC patients and 758 controls from the same Czech cohort [17]. Four different modes of inheritance were calculated and tested for each pair: the so called "three genotypes model" whereby each SNP was treated as a categorical variable with three levels (genotypes); the "log additive model" whereby SNPs were modeled as a continuous variable and genotypes were converted into 0, 1 or 2 risk alleles; the "dominant model" whereby AA was used as reference and AB and BB as the test group; and "the recessive model" whereby AA and AB were used as reference group and BB as the test group. Likelihood ratio tests were performed to assess whether including the SNP–SNP interaction term led to a significantly better fit of the data. The SNPs that significantly interacted with each other according to several competing models were ranked

according to Akaike information criterion (AIC). The smaller the value of AIC, the better the model data fit. To assess the contribution of all genetic components (both SNPs and interaction term) to the model, likelihood ratio test-based P-values were computed. For the best model of each SNP pair, the corresponding ORs and the Wald estimates for their confidence intervals (CIs) and P-values were calculated. Altogether, 120 (16 SNPs*(16–1)/2) independent tests were carried out, leading to a Bonferroni corrected p-value of 0.05/120 = 0.0004. In addition, as an alternative approach, we controlled the false discovery rate (FDR) using the Benjamini-Hochberg procedure. The p-values were sorted from the smallest to the largest and ranked in ascending order. The false discovery rate was calculated and controlled at an arbitrary level $q^* < 0.10$, defining $q = mP_{(1)}/i$, where m is the number of multiple tests, P the p-value of each interaction and i the Rank. Analysis was performed using R version 3.3.2.

Results

CRC risk

As shown in Table 1, there was a significant difference in the age and sex distribution between the cases and controls (p-value <0.0001 and p-value 2.6e-05, respectively). The genotype distribution of all 16 genotyped polymorphisms was consistent with HWE in the control group ($P > 0.05$). Logistic regression analysis adjusted for age and sex reported an association between rectal cancer risk and 2 *NLRC5* SNPs, rs1684575 (OR: 1.60, 95% CI: 1.13–2.27, recessive model) and rs3751710 (OR: 0.70, 95% CI: 0.51–0.96, dominant model) (S2 Table). The other genotyped SNPs did not show any association with CRC risk (S2 Table).

Possible effect of SNP-SNP interactions on CRC risk

We further investigated whether SNP-SNP interactions among these 16 SNPs within *NLRC5* and *PD-L1* genes could affect colorectal cancer risk. Eighteen interactions, including interactions between SNPs both within a gene and between the two genes, were detected at a significance level of p-value <0.05 (Table 2), however, none of these interaction term p-values survived the conservative Bonferroni multiple testing correction (p-value< 0.0004); although the global null hypothesis test was highly significant (p-value< 0.0001). When we calculated and controlled the FDR at an arbitrary level $q^* < 0.10$, a total of 12 of these interactions were below the given threshold (S3 Table). For the best model of each SNP-SNP interaction, the association with CRC risk was evaluated (S5 Table).

As shown in the Fig 1 most of the SNPs were interacting with two or more SNPs, either lying within the same or a different gene. Based on our selection criteria, the genotyped SNPs had pairwise LD $r^2 \leq 0.80$. However, some of the interactions can be explained by a lower level of LD (S1 Fig).

Three *NLRC5* SNPs, rs289747, rs289748 and rs56315364, mapping near/in the promoter ($r^2 = 0.42$–0.70), showed an interaction with the same *PD-L1* promoter SNP rs2890657. Of note, we observed an increased risk of CRC development when at least one minor allele of rs2890657 interacted with the GG genotype of rs289747. Conversely, a protective effect was observed when the CC genotype of rs2890657 interacted with the CC genotype of rs289748 and rs56315364, respectively (Supplementary S5 Table).

On the other hand, the *PD-L1* SNP rs2890657 together with another promoter SNP, rs822338 ($r^2 = 0.68$), interacted with the same *NLRC5* promoter SNP rs289747. Similar to rs2890657-rs289747 interaction, an increased risk of CRC was observed when at least one minor allele of rs822338 interacted with the GG genotype of rs289747 (S5 Table).

The two *PD-L1* promoter SNPs also interacted independently with two eQTL SNPs for *NLRC5*. The interaction partner for rs2890657 was rs12445252, whose T allele is predicted to

Table 2. *NLRC5-PD-L1* pair-wise interactions with cases and controls. Only the best genetic model of each SNP pair is shown.

SNP1	SNP2	Mode of inheritance SNP1	Mode of inheritance SNP2	LRT Statistic	DF	p-value based on LRT	LRT Statistic	DF	p-value based on LRT
				(interaction term)			(SNPs total)		
rs27194	rs289726	Three genotypes	Dominant	13	2	**0.002**	15.54	5	**0.008**
rs289726	rs822338	Dominant	Three genotypes	11.73	2	**0.003**	12.12	5	**0.033**
rs12445252	rs43216	Recessive	Dominant	9.34	1	**0.002**	11.12	3	**0.011**
rs3751710	rs4143815	Three genotypes	Recessive	8.63	2	**0.013**	11.52	5	**0.042**
rs2890657	rs289747	Dominant	Dominant	7.96	1	**0.005**	8.52	3	**0.036**
rs2890657	rs56315364	Recessive	Dominant	7.77	1	**0.005**	8.03	3	**0.045**
rs10815225	rs289726	Recessive	Dominant	7.74	1	**0.005**	11.12	3	**0.011**
rs12445252	rs2890657	Dominant	Recessive	7.66	1	**0.006**	7.93	3	**0.048**
rs27194	rs289748	Recessive	Dominant	7.61	1	**0.006**	8.25	3	**0.041**
rs2890657	rs289748	Recessive	Dominant	7.23	1	**0.007**	8.31	3	**0.04**
rs289747	rs822338	Dominant	Dominant	6.98	1	**0.008**	7.98	3	**0.046**
rs10815225	rs4143815	Recessive	Recessive	6.86	1	**0.009**	11.7	3	**0.009**
rs27194	rs56315364	Recessive	Three genotypes	6.79	2	**0.034**	11.88	5	**0.037**
rs27194	rs4143815	Recessive	Dominant	6.62	1	**0.01**	8.02	3	**0.046**
rs10815225	rs1684575	Recessive	Three genotypes	6.08	2	**0.048**	12.45	5	**0.029**
rs158483	rs866066	Recessive	Recessive	5.66	1	**0.017**	8.04	3	**0.045**
rs27194	rs43216	Recessive	Recessive	4.39	1	**0.036**	8.73	3	**0.033**
rs289748	rs56315364	Recessive	Recessive	4.13	1	**0.042**	10.27	3	**0.016**

DF: Degrees of Freedom

LTR: Likelihood Ratio Test

https://doi.org/10.1371/journal.pone.0192385.t002

decrease *NLRC5* expression in the whole blood tissue with an effect size of -0.26 and a p-value of 1.6e-7. rs822338 interacted with rs289726, whose C allele is reported to increase *NLRC5* expression in the whole blood with an effect size of 0.19 and a p-value of 10e-6. Interestingly, we observed a decreased CRC risk when the CC genotype of *PD-L1* rs2890657 interacted with at least one minor allele of *NLRC5* rs12445252, which is related with lower *NLRC5* expression and an increased CRC risk was shown by the interactions between the CC genotypes of both *PD-L1* rs822338 and *NLRC5* rs289726 and between the TT genotype of *PD-L1* rs822338 and the genotypes with at least one minor allele T of *NLRC5* rs289726 (S5 Table).

Two 3'UTR SNPs in *NLRC5*, rs43216 and rs27194, which map within a genetic block of 305 bp (r^2 = 0.43), revealed a decreased risk of CRC when at least one major allele of rs27194 interacted with the minor allele genotype of rs43216. Additionally, both of them interacted independently with the two *NLRC5* eQTL SNPs, rs12445252 and rs289726, respectively. We observed an increased risk when the GG genotype of rs43216, which binds a lower number of miRNAs, interacted with the TT genotype of rs12445252, related to a lower expression of NLRC5, as well as when the TT genotype of rs27194, which also binds a lower number of miRNAs, interacted with the CC genotype of rs289726, which instead is related to a higher *NLRC5* expression. Moreover, the same genotype of rs27194 showed an increased risk also when it interacted with the CC genotype of the *NLRC5* flanking SNP rs289748 (S5 Table).

When we included the *IFNGR* genes variants previously analyzed in an older version of our cohort with a lower number of individuals (1327 cases and 758 healthy controls [17]) to our analysis, we observed 6 additional interactions including three SNPs within *IFNGR1* (rs2234711, rs1327474 and rs17181457) and two within *IFNGR2* (rs17882748 and rs1059293) (Table 3). Among them, rs2234711, lying within the 5'UTR of *IFNGR1*, showed complicated

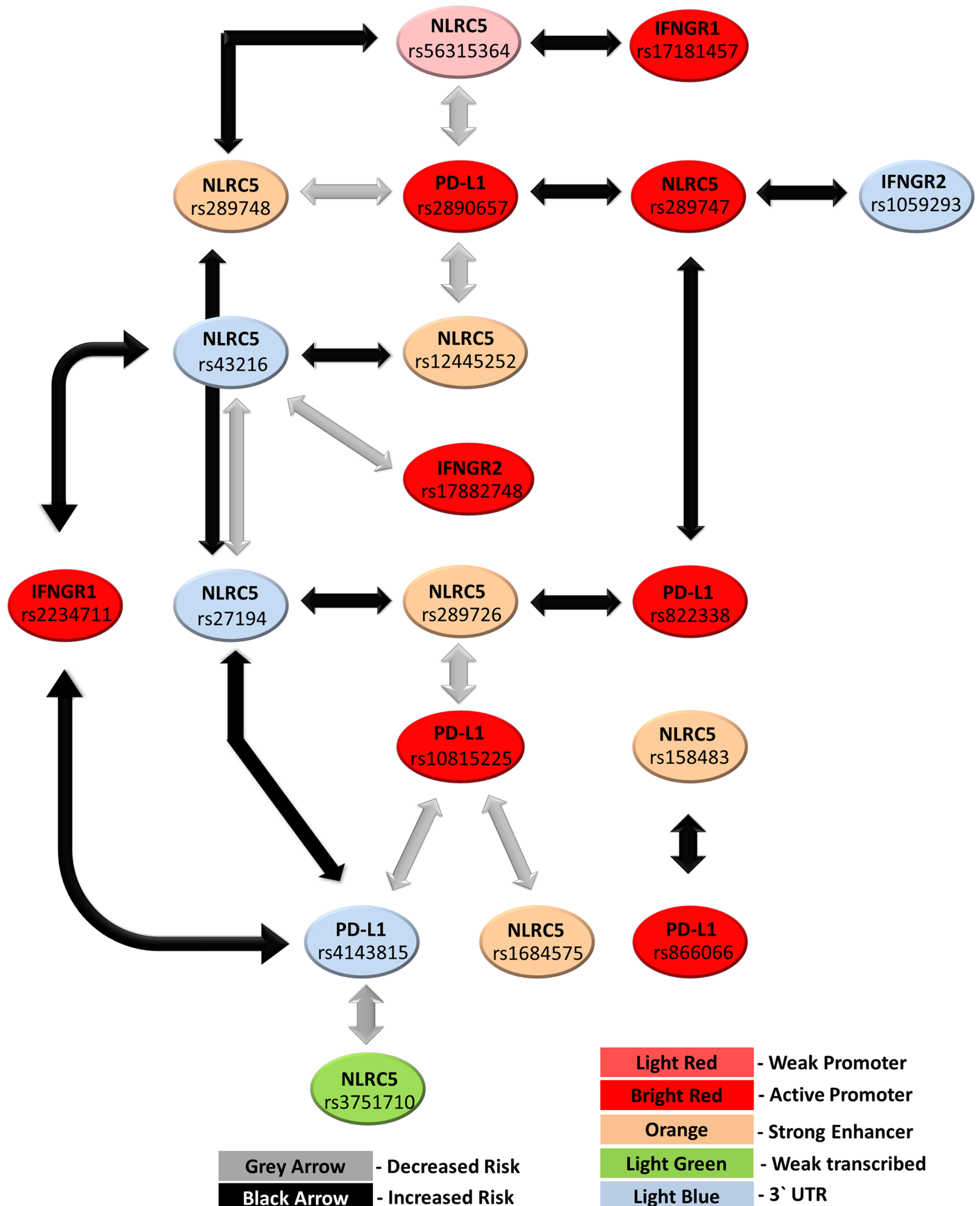


Fig 1. *NLRC5-PD-L1-IFNGR1/2* pair-wise interactions. The color indicates the SNPs' location displayed by UCSC Genome Browser on lymphoblastoid cell lines (GM12878).

https://doi.org/10.1371/journal.pone.0192385.g001

Table 3. *NLRC5-IFNGR1/2* and *PD-L1-IFNGR1/2* pair-wise interactions with cases and controls. Only the best genetic model of each SNP pair is shown.

SNP1	SNP2	Mode of inheritance SNP1	Mode of inheritance SNP2	LRT Statistic	DF	p-value based on LRT	LRT Statistic	DF	p-value based on LRT
				(interaction term)			(SNPs total)		
rs1059293	rs289747	Allele number	Three genotypes	21.17	2	**< .0001**	21.84	5	**0.001**
rs1059293	rs43216	Dominant	Three genotypes	9.64	2	**0.008**	11.64	5	**0.04**
rs17181457	rs56315364	Dominant	Recessive	3.95	1	**0.05**	12.53	3	**0.006**
rs17882748	rs43216	Recessive	Three genotypes	9.53	2	**0.009**	13.47	5	**0.019**
rs2234711	rs4143815	Dominant	Three genotypes	8.7	2	**0.013**	21.3	5	**0.001**
rs2234711	rs43216	Three genotypes	Dominant	7.72	2	**0.021**	19.19	5	**0.002**

DF: Degrees of Freedom

LTR: Likelihood Ratio Test

https://doi.org/10.1371/journal.pone.0192385.t003

interactions with two 3' UTR variants, *NLRC5* rs43216 and *PD-L1* rs4143815 (S6 Table). The strongest interaction was observed between a 3'UTR SNP in *IFNGR2* rs1059293 C>T and the *NLRC5* promoter SNP rs289747 G>A, which also survived the Bonferroni multiple testing correction (interaction term p value < 0.0004) and the FDR controlled at an arbitrary level q^* < 0.10 (S4 Table). Particularly carriers of rs1059293 CC homozygous genotype and rs289747 heterozygous genotype showed an increased risk of CRC development (S6 Table). Curiously, an increased risk was also observed for T allele carriers of rs1059293 and GG genotype carriers of rs289747.

Discussion

In our case-control study comprising up to 1424 cases and 1114 healthy controls, we investigated the role of genetic polymorphisms in the regulatory regions of *NLRC5*, *PD-L1* and the previously genotyped regulatory SNPs in the *IFNGR* genes on the risk of CRC. In the single SNP analysis, only 2 SNPs out of 16, rs1684575 T>G (OR: 1.60, 95% CI: 1.13–2.27, recessive model) and rs3751710 C>T (OR: 0.70, 95% CI: 0.51–0.96, dominant model), both mapping within the *NLRC5* gene, showed a nominal association with rectal cancer risk in the Czech population (p < 0.05). In our previous study on potentially functional *IFNGR* SNPs, rs2234711 in the 5'UTR of *IFNGR1*, was reported to be associated with an increased risk of CRC; particularly the risk allele C was associated with *IFNGR1* gene activity in a context-dependent manner [17,19].

Given that the evaluated 16 SNPs did not show a strong individual association with CRC risk and that SNPs represent common genetic alterations typically characterized by a low level of penetrance, we further evaluated whether their binary interactions might uncover synergistic effects contributing to CRC predisposition. Taking into account that all SNPs were non-coding variants, though located in regulatory regions (promoter, enhancer, 5´and 3´UTR), a possible biological mechanism may be an active involvement in the regulation of gene expression [20]. Perturbations in *NLRC5* and *PD-L1* gene expression may lead, as a consequence, to a dysregulation of the anti-tumor immune response, which in turn may influence CRC development [21–23]. Indeed, the immune infiltration is a major outcome factor in CRC [24,25] and altering immune-regulatory machinery is one of the mechanisms developed by cancer cells to evade the immune system and form a tumor [26,27].

Altogether, we observed 18 interactions between *NLRC5* and *PD-L1*, and further 6 interactions together with *IFNGR1/2* in a smaller sample set. For all interactions, the global null hypothesis test was highly significant (p-value < .0001). Twelve out of the 18 *PD-L1-NLRC5*

interactions were below the threshold for the FDR controlled at an arbitrary level $q^* < 0.10$, while only one out of the 6 *PD-L1-NLRC5-IFNGR1/2* interactions survived both the FDR and the Bonferroni multiple testing correction (p-value< 0.0004). It should be pointed out that we had relatively low power to detect such an association in the first place, due to the limited number of cases in the interacting genotype categories and the high stringency of the Bonferroni correction [28]. Finally, the tests were not completely independent, due to the fact that many SNPs in each gene were studied and there was moderate to low LD between some of the SNPs.

The main interactions included three moderately linked *NLRC5* SNPs rs289747, rs289748, rs56315364, mapping within a genetic block of 3 kb located in the upstream and promoter region of the gene that exhibited a significant interaction with the same *PD-L1* promoter SNP rs2890657. Also another *PD-L1* promoter SNP rs822338 ($r^2 = 0.68$ with rs2890657) interacted with the *NLRC5* promoter SNP rs289747. Interestingly, rs2890657 showed also an interaction with an eQTL SNP for *NLRC5*, rs12445252, while rs822338 interacted with another *NLRC5* eQTL SNP rs289726. Additionally, two 3'UTR SNPs in *NLRC5*, rs43216 and rs27194 ($r^2 = 0.43$), interacted independently with the two *NLRC5* eQTL SNPs, and rs27194 also with another *NLRC5* flanking SNP rs289748. Furthermore, we observed interactions between a 5' UTR SNP in *IFNGR1*, rs2234711, and 3' UTR variants in *NLRC5* (rs43216) and *PD-L1* (rs4143815), respectively, and between a 3' UTR SNP in *IFNGR2* (rs1059293) and a promoter SNP in *NLRC5* (rs289747).

All upstream and/or promoter SNPs in the *NLRC5* and *PD-L1* genes involved in the most significant interactions, and several other SNPs in high LD with them, are located within promoter histone marks and DNase hypersensivity sites. Two of the *NLRC5* SNPs, rs289747 and rs56315364, are predicted to affect in an opposite way the OCT proteins binding site, reflecting the opposite associations that they elicit on the CRC development and supporting the reliability of our interaction analysis. Also the *PD-L1* SNPs rs2890657 and rs822338 are estimated to affect transcription factor binding sites: rs2890657 the c-Myb binding site and rs822338 together with 5 linked SNPs the binding sites of transcription factors such as TAF1 (TATA-box binding protein associated factor 1) and p300. P300 is a histone acetyltransferase that regulates transcription of genes via chromatin remodeling [29]. Members of the TAF transcription factor family may participate in basal transcription, as coactivators, or in promoter recognition or to facilitate complex assembly and transcription initiation [30].

The two *PD-L1* promoter SNPs, rs2890657 and rs822338, also interacted independently with two *NLRC5* eQTL SNPs, rs12445252 and rs289726, respectively. Interestingly, we observed a lower CRC risk when *PD-L1* rs2890657 interacted with the allele of the *NLRC5* rs12445252, which is predicted to decrease *NLRC5* expression (-0.26 and a p-value of 1.6e-7), while the interaction between the different genotype categories of *PD-L1* rs822338 and *NLRC5* rs289726 was more complex, implicating an increased CRC risk for genotype combinations including rs289726 alleles predicted to either increase or decrease NLRC5 expression.

Furthermore, the two 3'UTR SNPs in *NLRC5*, rs43216 and rs27194 ($r^2 = 0.43$), exhibited a decreased risk of CRC, which may be due to the interaction between the alleles that are predicted to bind a higher number of miRNAs than the other allele, leading to a stricter *NLRC5* post-transcriptional repression. Furthermore, both of them were found to be involved in an independent interaction with the two *NLRC5* eQTL SNPs. Particularly, an increased risk was observed when the rs43216 allele binding a lower number of miRNAs interacted with the rs12445252 allele related with a lower expression of *NLRC5*, as well as when the rs27194 allele which has a less strict post-transcriptional repression, interacted with the rs289726 allele related to a higher *NLRC5* expression, again reflecting the complex interactions between the genomic regions. Moreover, rs27194 showed an increased risk also when it interacted with

NLRC5 flanking SNP rs289748, when the allele of rs27194 binding lower number of miRNAs was involved in the interaction. These results suggest that a deregulation in the *NLRC5* expression through complicated interactions between genetic variants may lead to alterations in the downstream pathways and by that influence the risk of CRC.

Additionally, postulating that both, *NLRC5* and *PD-L1*, are downstream targets of IFNγ, we evaluated them in binary interaction with *IFNGR1* and *IFNGR2* variants previously genotyped by us [17]. We observed that the *IFNGR1* 5'UTR SNP, rs2234711, interacted with two 3' UTR variants, *NLRC5* rs43216 and *PD-L1* rs4143815. The association between the *IFNGR1* SNP and the risk of CRC has already been established in our previous study [17]. In the present study the previous association was strengthened when the risk allele of rs2234711 interacted with the variant of rs43216, related with a stricter *NLRC5* post-transcriptional repression and with the allele of rs4143815 whose antisense is targeted by the miR-570, a negative regulator of *PD-L1*, as reported by the online prediction tools.

Finally, a 3'UTR SNP in *IFNGR2* rs1059293 C>T presented an interaction with the *NLRC5* promoter SNP rs289747 G>A. The result pointed to a complicated interaction between the two variants. An increased risk of CRC was observed both for carriers of the CC genotype for *IFNGR2* rs1059293 and the heterozygous GA genotype for *NLRC5* rs289747 and for the T allele carriers of rs1059293 and the GG genotype carriers of rs289747. In this context the C allele of rs1059293 has been reported to bind a lower number of miRNAs than the T allele. On the other hand, the *NLRC5* rs289747 is reported to affect OCT1 binding site, with the G allele showing a nearly inexistent affinity for OCT1, compared to the A allele, which instead is reported to exhibit a consistently increasing affinity. OCT1 is also reported to be overexpressed in many cancers, including CRC [31–33] and the IFNγ promoter has been reported to contain a binding site for Oct proteins [34]. As a consequence, the secretion of IFNγ by Oct proteins might be increased contributing to a dysregulation of the expression of the downstream pathway genes, such as *NLRC5* and *PD-L1* [35,36].

Assuming that NLRC5 has been reported to be the major MHC class I transactivator, a hyper-stimulation of its expression could lead to a strong $CD8^+$ activation. Conversely a lower *NLRC5* expression has been reported to influence the MHC class I expression leading to an impaired ability to elicit $CD8^+$ T-cell activation, which represent a way used by the tumor cells to escape the host immune system [23]. Additionally, recent data suggest that 5-Fluorouracil, a chemotherapeutic frequently used in CRC treatment, impacts on *PD-L1* expression [37]. Therefore the *PD-L1* SNPs studied here and their interactions with *IFNGR* and *NLRC5* variants may also be worth studying with regard to therapy response as well as survival of the CRC patients.

In this study, we included only four of the many immune-related genes for several reasons: first because of the interesting opposite effect that NLRC5 and PD-L1 exert on the regulation of T-cell mediated immunity, second because both of them are downstream targets of IFNγ and third because of the emerging role of these genes on CRC as well as on other cancer types. Furthermore, including a large network of genes would have led to a higher number of multiple tests, increasing the likelihood of chance findings. However, our study serves as a starting point to study the interplay between all the genes involved in the mucosal immune system, which would possibly shed light on the mechanisms underlying CRC development.

In conclusion, we anticipate that the interaction between the inherited genetic variants contributes to signaling defects, which in turn may lead to alteration in the anti-tumor immune response. Defects in the immune responses, especially in the expression of genes involved in immune surveillance, could favor tumorigenesis. Additionally, perturbation of the physiological immune homeostasis may also affect inflammation, another predisposing step for CRC development. It will be interesting to monitor the effect of the variants identified here under

standard therapies for spontaneous and inflammation-related CRC and in ongoing clinical trials with immune check-point inhibitors where effects may be even more pronounced [16,37].

Supporting information

S1 Table. Information about the SNPs evaluated in this study provided by different online tools.
(PDF)

S2 Table. Association between the selected SNPs and the colorectal cancer risk.
(PDF)

S3 Table. False discovery rate for each individual *NLRC5-PD-L1* pair-wise interaction.
(PDF)

S4 Table. False discovery rate for each individual *NLRC5-PD-L1-IFNGR1/2* pair-wise interaction.
(PDF)

S5 Table. *NLRC5-PD-L1* pair-wise interactions. For the best model of each pair, age and sex adjusted ORs and 95% CI, with overall p-values based on the likelihood ratio test, were calculated.
(PDF)

S6 Table. *IFNGR1/2, PD-L1* and *NLRC5* pair-wise interactions. For the best model of each pair, age and sex adjusted ORs and 95% CI, with overall p-values based on the likelihood ratio test, were calculated.
(PDF)

S1 Fig. Haploview plot showing LD and haplotype blocks within *PD-L1* and *NLRC5* genes based on r^2.
(TIF)

Acknowledgments

We thank Dr. Hauke Thomsen for the statistical advice.

Author Contributions

Conceptualization: Alexander N. R. Weber, Asta Försti.

Data curation: Calogerina Catalano, Miguel Inacio da Silva Filho, Christoph Frank, Katerina Jiraskova, Veronika Vymetalkova, Ludmila Vodickova.

Formal analysis: Calogerina Catalano, Miguel Inacio da Silva Filho, Christoph Frank.

Funding acquisition: Katerina Jiraskova, Veronika Vymetalkova, Pavel Vodicka, Alexander N. R. Weber.

Investigation: Calogerina Catalano.

Methodology: Calogerina Catalano.

Project administration: Asta Försti.

Resources: Katerina Jiraskova, Veronika Vymetalkova, Miroslav Levy, Vaclav Liska, Ondrej Vycital, Alessio Naccarati, Ludmila Vodickova, Pavel Vodicka.

Software: Calogerina Catalano.

Supervision: Asta Försti.

Visualization: Calogerina Catalano.

Writing – original draft: Calogerina Catalano, Asta Försti.

Writing – review & editing: Ludmila Vodickova, Kari Hemminki, Pavel Vodicka, Alexander N. R. Weber, Asta Försti.

References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 2017; 66: 683–691. https://doi.org/10.1136/gutjnl-2015-310912 PMID: 26818619
2. Janakiram NB, Rao C V. The role of inflammation in colon cancer. *Adv Exp Med Biol* 2014; 816: 25–52. https://doi.org/10.1007/978-3-0348-0837-8_2 PMID: 24818718
3. Wu H-J, Wu E. The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes* 2012; 3: 4–14. https://doi.org/10.4161/gmic.19320 PMID: 22356853
4. Weber AN, Försti A. Toll-like receptor genetic variants and colorectal cancer. *Oncoimmunology* 2014; 3: e27763. https://doi.org/10.4161/onci.27763 PMID: 24790794
5. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell* 2014; 157: 121–41. https://doi.org/10.1016/j.cell.2014.03.011 PMID: 24679531
6. Hassanzadeh P. Colorectal cancer and NF-κB signaling pathway. *Gastroenterol Hepatol from bed to bench* 2011; 4: 127–32.
7. Cheon H, Borden EC, Stark GR. Interferons and their stimulated genes in the tumor microenvironment. *Semin Oncol* 2014; 41: 156–73. https://doi.org/10.1053/j.seminoncol.2014.02.002 PMID: 24787290
8. Slattery ML, Lundgreen A, Bondurant KL, Wolff RK. Interferon-signaling pathway: associations with colon and rectal cancer risk and subsequent survival. *Carcinogenesis* 2011; 32: 1660–7. https://doi.org/10.1093/carcin/bgr189 PMID: 21859832
9. Welsh RM, Bahl K, Marshall HD, Urban SL. Type 1 interferons and antiviral CD8 T-cell responses. *PLoS Pathog* 2012; 8: e1002352. https://doi.org/10.1371/journal.ppat.1002352 PMID: 22241987
10. Meissner TB, Li A, Kobayashi KS. NLRC5: a newly discovered MHC class I transactivator (CITA). *Microbes Infect* 2012; 14: 477–84. https://doi.org/10.1016/j.micinf.2011.12.007 PMID: 22209772
11. Rodriguez GM, Bobbala D, Serrano D, Mayhue M, Champagne A, Saucier C et al. NLRC5 elicits antitumor immunity by enhancing processing and presentation of tumor antigens to CD8(+) T lymphocytes. *Oncoimmunology* 2016; 5: e1151593. https://doi.org/10.1080/2162402X.2016.1151593 PMID: 27471621
12. Pandiyan P, Hegel JKE, Krueger M, Quandt D, Brunner-Weinzierl MC. High IFN-γ Production of Individual CD8 T Lymphocytes Is Controlled by CD152 (CTLA-4). *J Immunol* 2007; 178.http://www.jimmunol.org/content/178/4/2132.long (accessed 28 Jul2017).
13. Riella L V, Paterson AM, Sharpe AH, Chandraker A. Role of the PD-1 pathway in the immune response. *Am J Transplant* 2012; 12: 2575–87. https://doi.org/10.1111/j.1600-6143.2012.04224.x PMID: 22900886
14. Karwacz K, Bricogne C, MacDonald D, Arce F, Bennett CL, Collins M et al. PD-L1 co-stimulation contributes to ligand-induced T cell receptor down-modulation on CD8+ T cells. *EMBO Mol Med* 2011; 3: 581–92. https://doi.org/10.1002/emmm.201100165 PMID: 21739608
15. Kataoka K, Shiraishi Y, Takeda Y, Sakata S, Matsumoto M, Nagano S et al. Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers. *Nature* 2016; 534: 402–6. https://doi.org/10.1038/nature18294 PMID: 27281199
16. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 2015; 372: 2509–2520. https://doi.org/10.1056/NEJMoa1500596 PMID: 26028255
17. Lu S, Pardini B, Cheng B, Naccarati A, Huhn S, Vymetalkova V et al. Single nucleotide polymorphisms within interferon signaling pathway genes are associated with colorectal cancer susceptibility and survival. *PLoS One* 2014; 9: e111061. https://doi.org/10.1371/journal.pone.0111061 PMID: 25350395
18. Carrai M, Campa D, Vodicka P, Flamini R, Martelli I, Slyskova J et al. Association between taste receptor (TAS) genes and the perception of wine characteristics. *Sci Rep* 2017; 7: 9239. https://doi.org/10.1038/s41598-017-08946-3 PMID: 28835712

19. Jüliger S, Bongartz M, Luty AJF, Kremsner PG, Kun JFJ. Functional analysis of a promoter variant of the gene encoding the interferon-gamma receptor chain I. *Immunogenetics* 2003; 54: 675–80. PMID: 12557053

20. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. *Nat Rev Genet* 2016; 17: 93–108. https://doi.org/10.1038/nrg.2015.17 PMID: 26781813

21. Lynch D, Murphy A. The emerging role of immunotherapy in colorectal cancer. *Ann Transl Med* 2016; 4: 305. https://doi.org/10.21037/atm.2016.08.29 PMID: 27668225

22. Passardi A, Canale M, Valgiusti M, Ulivi P. Immune Checkpoints as a Target for Colorectal Cancer Treatment. *Int J Mol Sci* 2017; 18. https://doi.org/10.3390/ijms18061324

23. Yoshihama S, Roszik J, Downs I, Meissner TB, Vijayan S, Chapuy B et al. NLRC5/MHC class I transactivator is a target for immune evasion in cancer. *Proc Natl Acad Sci U S A* 2016; 113: 5999–6004. https://doi.org/10.1073/pnas.1602069113 PMID: 27162338

24. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol* 2015; 16: 64. https://doi.org/10.1186/s13059-015-0620-6 PMID: 25853550

25. Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D et al. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* 2016; 44: 698–711. https://doi.org/10.1016/j.immuni.2016.02.025 PMID: 26982367

26. Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin Cancer Res* 2015; 21: 687–92. https://doi.org/10.1158/1078-0432.CCR-14-1860 PMID: 25501578

27. Seliger B. Strategies of tumor immune evasion. *BioDrugs* 2005; 19: 347–54. PMID: 16392887

28. Hendricks AE, Dupuis J, Logue MW, Myers RH, Lunetta KL. Correction for multiple testing in a gene region. *Eur J Hum Genet* 2014; 22: 414–8. https://doi.org/10.1038/ejhg.2013.144 PMID: 23838599

29. Zhang X, Ouyang S, Kong X, Liang Z, Lu J, Zhu K et al. Catalytic mechanism of histone acetyltransferase p300: from the proton transfer to acetylation reaction. *J Phys Chem B* 2014; 118: 2009–19. https://doi.org/10.1021/jp409778e PMID: 24521098

30. Lieberman PM, Berk AJ. A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA—promoter DNA complex formation. *Genes Dev* 1994; 8: 995–1006. PMID: 7926793

31. Wang Y-P, Song G-H, Chen J, Xiao C, Li C, Zhong L et al. Elevated OCT1 participates in colon tumorigenesis and independently predicts poor prognoses of colorectal cancer patients. *Tumour Biol* 2016; 37: 3247–55. https://doi.org/10.1007/s13277-015-4080-0 PMID: 26433389

32. Jin T, Branch DR, Zhang X, Qi S, Youngson B, Goss PE. Examination of POU homeobox gene expression in human breast cancer cells. *Int J cancer* 1999; 81: 104–12. PMID: 10077160

33. Almeida R, Almeida J, Shoshkes M, Mendes N, Mesquita P, Silva E et al. OCT-1 is over-expressed in intestinal metaplasia and intestinal gastric carcinomas and binds to, but does not transactivate, CDX2 in gastric cells. *J Pathol* 2005; 207: 396–401. https://doi.org/10.1002/path.1861 PMID: 16278805

34. Mueller K, Quandt J, Marienfeld RB, Weihrich P, Fiedler K, Claussnitzer M et al. Octamer-dependent transcription in T cells is mediated by NFAT and NF-κB. *Nucleic Acids Res* 2013; 41: 2138–54. https://doi.org/10.1093/nar/gks1349 PMID: 23293002

35. Kuenzel S, Till A, Winkler M, Häsler R, Lipinski S, Jung S et al. The Nucleotide-Binding Oligomerization Domain-Like Receptor NLRC5 Is Involved in IFN-Dependent Antiviral Immune Responses. *J Immunol* 2010; 184.http://www.jimmunol.org/content/184/4/1990.long (accessed 2 Aug2017).

36. Wang X, Teng F, Kong L, Yu J. PD-L1 expression in human cancers and its association with clinical outcomes. *Onco Targets Ther* 2016; 9: 5023–39. https://doi.org/10.2147/OTT.S105862 PMID: 27574444

37. Van L, Kraak D, Goel G, Ramanan K, Kaltenmeier C, Zhang L et al. 5-Fluorouracil upregulates cell surface B7- H1 (PD-L1) expression in gastrointestinal cancers. *J Immunother Cancer* 2016; 4. https://doi.org/10.1186/s40425-016-0163-8

Manuscript VI

Catalano C, Silva Filho MI, **Jiraskova K**, Vymetalkova V, Levy M, Liska V, Vycital O, Naccarati A, Vodickova L, Hemminki K, Vodicka P, Weber ANR, Försti A.

Influence of regulatory NLRC5 variants on colorectal cancer survival and 5-fluorouracil-based chemotherapy

Eur J Gastroenterol Hepatol (2018) 30(8):838-842

Influence of regulatory *NLRC5* variants on colorectal cancer survival and 5-fluorouracil-based chemotherapy

Calogerina Catalano[a], Miguel I. da Silva Filho[a], Katerina Jiraskova[c,d], Veronika Vymetalkova[c,d], Miroslav Levy[f], Vaclav Liska[e,f], Ondrej Vycital[e,f], Alessio Naccarati[c,h], Ludmila Vodickova[c,d,e], Kari Hemminki[a,g], Pavel Vodicka[c,d,e], Alexander N.R. Weber[b] and Asta Försti[a,g]

Background NLRC5 is an interferon γ-inducible protein, which plays a role in immune surveillance with a potential influence on cancer survival.
Objective We aimed to evaluate the effect of potential regulatory variants in *NLRC5* on overall survival and survival after 5-fluorouracil (5-FU)-based therapy of colorectal cancer (CRC) patients.
Patients and methods We carried out a case-only study in a Czech population of 589 cases; 232 received 5-FU-based therapy. Eleven variants within *NLRC5* were selected using in-silico tools. Associations between polymorphisms and survival were assessed by Cox regression analysis adjusting for age at diagnosis, sex, and TNM stage. Survival curves were derived using the Kaplan–Meier method.
Results Two variants showed a significant association with survival. All patients and metastasis-free patients at the time of diagnosis (pM0) who were homozygous carriers of the minor allele of rs27194 had a decreased overall survival (OS_{all} and OS_{pM0}) and event-free survival (EFS_{pM0}) under a recessive model (OS_{all} $P=0.003$, OS_{pM0} $P=0.005$, EFS_{pM0} $P=0.01$, respectively). OS was also decreased for all patients and for pM0 patients who carried at least one minor allele of rs289747 (OS_{all} $P=0.03$ and OS_{pM0} $P=0.003$, respectively). Among CRC patients, who underwent a 5-FU-based adjuvant regimen, rs12445252 was associated with OS_{all}, OS_{pM0} and EFS_{pM0}, according to the dosage of the minor allele T (OS_{all} $P=0.0004$, OS_{pM0} $P=0.0001$, EFS_{pM0} $P=0.008$, respectively).
Conclusion Our results showed that polymorphisms in *NLRC5* may be used as prognostic markers of survival of CRC patients, as well as for survival in response to 5-FU treatment. Eur J Gastroenterol Hepatol 00:000–000
Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved.

Introduction

NLRC5 [NOD-like receptor (NLR) family, caspase recruitment domain (CARD) containing 5] is an interferon γ (IFNγ)-inducible protein, which plays a pivotal role in immune surveillance. NLRC5 is not only a transactivator of MHC class I molecules, it is also involved in the activation of genes in the MHC class I antigen-presentation pathway, such as antigen peptide transporter 1 (*TAP1*), proteasome subunit β type-9 (*PSMB9* also known as *LMP2*), and β2-microglobulin (*B2M*) [1–3]. Because of its link with the IFNγ system, NLRC5 might also play a role in the 5-fluorouracil (5-FU)-based therapy. 5-FU is a pyrimidine analog that acts by inhibiting thymidylate synthase, which plays a pivotal role in the conversion of deoxyuridine monophosphate to deoxythymidine monophosphate, an important precursor required for DNA synthesis [4]. It has been shown that the 5-FU-based therapy selectively targets the myeloid-derived suppressor cells (MDSCs) [5–10]. MDSCs are widely present in colorectal cancer (CRC) [11]. Elimination of MDSCs enhances IFNγ secretion by tumor-specific $CD8^+$ cells [8], which in turn leads to *NLRC5* stimulation [1]. On the basis of its role in immune evasion in cancer, NLRC5 might also affect the survival of cancer patients. The aim of our study was to evaluate the effect of potential regulatory variants in *NLRC5* on overall survival and survival in response to 5-FU-based therapy of CRC patients.

European Journal of Gastroenterology & Hepatology 2018, 00:000–000

Keywords: 5-fluorouracil, colorectal cancer, NLRC5, survival analysis

[a]Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, [b]Department of Immunology, Interfaculty Institute for Cell Biology, University of Tübingen, Tübingen, Germany, [c]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, [d]Institute of Biology and Medical Genetics, First Faculty of Medicine, [e]Biomedical Centre, Faculty of Medicine, Pilsen, [f]Department of Surgery, First Medical Faculty, Charles University and Thomayer Hospital, Prague, Czech Republic, [g]Center for Primary Health Care Research, Clinical Research Center, Lund University, Malmö, Sweden and [h]Molecular and Genetic Epidemiology, Italian Institute for Genomic Medicine (IIGM), Turin, Italy

Correspondence to Calogerina Catalano, MSc, Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120, Heidelberg, Germany

Tel: +49 622 142 1805; fax: +49 622 142 1810; e-mail: c.catalano@dkfz.de

Received 13 December 2017 **Accepted** 25 February 2018

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.eurojgh.com.

Patients and methods

In this study, we selected 11 potential regulatory polymorphisms within *NLRC5*, using several in-silico online

0954-691X Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved. DOI: 10.1097/MEG.0000000000001154

Copyright © 2018 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

tools, including Ensembl (*http://www.ensembl.org/index.html*), UCSC Genome Browser (*https://genomeeuro.ucsc.edu/*), HaploReg (*http://www.broadinstitute.org/mammals/haploreg/haploreg.php*), Regulome DB (*http://www.regulomedb.org/*), Gtex Portal (*https://gtexportal.org/home/*), MicroSNiPer (*http://epicenter.iefreiburg.mpg.de/services/microsniper/*) and Transcription factor Affinity Prediction (*http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi*) tools. The selected polymorphisms were located in promoter (rs289747), promoter-flanking region (rs3751710, rs56315364, rs289748, rs7197864), intron (rs158483 and rs1684575), and 3′-UTR (rs27194 and rs43216) or they were intronic eQTLs (rs12445252 and rs289726) (Supplementary Table S1, Supplemental digital content 1, *http://links.lww.com/EJGH/A288*).

Patients

CRC patients were recruited within an ongoing study by several oncological departments in the Czech Republic. For a total of 589 CRC patients, recruited between 2003 and 2013, clinicopathological data and information about recurrence, distant metastasis, date of death, or date of the last contact with the treating physician were available (Table 1) [12]. Survival time was measured from the date of surgery to the end of follow-up or to two different events: death by any cause for overall survival among all patients (OS_{all}) and among patients with nonmetastatic disease at the time of diagnosis (OS_{pM0}); recurrence, distant metastasis, or death, whichever came first, for event-free survival (EFS_{pM0}).The median age of our cohort at the time of diagnosis was 65 years (SD: 11; range: 27–90), and 61.8% of the patients were men. The median follow-up time was 46 months (SD: 31; range: 1–143); during the follow-up time, 140 patients experienced a recurrence or distant metastasis. The median follow-up time for 427 patients who were metastasis-free at the time of diagnosis (pM0) was 52 months (SD: 32; range: 1–143) for death and 43 months (SD: 34; range: 1–143) for event, respectively; 112 patients recorded a recurrence or distant metastasis during the follow-up period. Of the 589 CRC patients, 232 received a 5-FU-based adjuvant regimen (supplemented with leucovorin and/or oxaliplatin), as the first-line postoperative therapy (Table 1). Throughout the follow-up period (median = 48.5 months), 65 patients experienced a recurrence or metastasis. The median follow-up time for 186 patients whose disease was not metastatic at the time of diagnosis was 58 months for death and 47.5 months for event, respectively; 55 patients reported a postsurgical recurrence or distant metastasis, after which 36 patients died.

The project design was approved by the Ethical Committee of the Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, in accordance with the Declaration of Helsinki. Written informed consent was obtained from all the patients who participated in this study.

Genotyping method

Genomic DNA from peripheral blood leukocytes was used to genotype the selected polymorphisms. The TaqMan (Thermo Fisher Scientific, Darmstadt, Germany) and the KASP (LGC genomics; Hoddesdon, Hertfordshire, UK) allelic discrimination methods were used for genotyping. DNA amplification was performed according to the TaqMan and LGC genomics' PCR conditions. Genotype detection was performed using the ViiA 7 Real-Time PCR System (Thermo Fisher Scientific). The sample set contained 34 duplicated samples as quality controls. The genotype correlation between the duplicate samples was more than 90%. Genotype call rate ranged between 94.0 and 100%.

Statistical analysis

Associations between the polymorphisms and clinical outcome were assessed by Cox regression analysis, adjusting for age at diagnosis, sex and TNM stage (PROC PHREG, SAS software version 9.3; SAS Institute Inc., Cary, North Carolina, USA). Survival curves were derived using the Kaplan–Meier (K–M) method (PROC LIFETEST, SAS version 9.3; SAS Institute Inc.).

Results

Cox regression analysis, in line with the K–M analysis, showed that two polymorphisms were associated with the survival outcome (Fig. 1). Carriers of the minor allele T homozygotes, ($n = 37$ OS_{all}, $n = 26$ OS_{pM0} and EFS_{pM0}), of rs27194 had a decreased survival under a recessive model [hazard ratio (HR): 1.92, 95% confidence interval (CI): 1.25–2.93, HR: 2.31, 95% CI: 1.30–4.13, HR: 2.04, 95% CI: 1.17–3.54, respectively]. Carriers of at least one minor allele of rs289747 ($n = 343$ OS_{all}, $n = 244$ OS_{pM0} and EFS_{pM0}) had a decreased survival outcome, in the OS_{all} and OS_{pM0} analysis, under a dominant model (HR: 1.32, 95% CI: 1.02–1.70, HR: 1.73, 95% CI: 1.21–2.47, respectively) (Supplementary Table S2, Supplemental digital content 1, *http://links.lww.com/EJGH/A288*). No association was detected for the remaining single-nucleotide polymorphisms (SNPs) (Supplementary Table S2, Supplemental digital content 1, *http://links.lww.com/EJGH/A288*).

When we tested survival in the set of CRC patients ($n = 232$ for OS_{all} and $n = 186$ for OS_{pM0} and EFS_{pM0}), who received a 5-FU-based adjuvant regimen, two polymorphisms, rs289747 and rs12445252, associated with survival; however, the effect size of rs289747 was similar to the one observed in the whole cohort. The other polymorphism, rs12445252, reported a decreased OS_{all}, OS_{pM0} and EFS_{pM0}, supported by the K–M analysis, which was restricted to 5-FU treated patients (Fig. 1). The association was observed according to the dosage of the minor allele T ($n = 111$ for OS_{all} and $n = 92$ for OS_{pM0} and EFS_{pM0}) (HR: 1.99, 95% CI: 1.36–2.92, HR: 2.37, 95% CI: 1.55–3.63, HR: 1.64, 95% CI: 1.14–2.35, respectively) (Supplementary Table S3, Supplemental digital content 1, *http://links.lww.com/EJGH/A288*). Patient characteristics, such as age at diagnosis and sex, did not affect survival in patients who underwent 5-FU chemotherapy ($P > 0.11$). No association was detected for the remaining SNPs (Supplementary Table S3, Supplemental digital content 1, *http://links.lww.com/EJGH/A288*).

OS_{all} and OS_{pM0} analysis for rs27194 ($P = 0.003$ and $= 0.005$, respectively) and rs12445252 ($P = 0.0004$ and $= 0.0001$, respectively), as well as $OS_{pM}0$ analysis for

Copyright © 2018 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Table 1. Clinical characteristics of colorectal cancer patients

	Overall survival of all patients					Overall survival of 5-fluorouracil-treated patients				
Parameters	No. at risk	Deaths	HR	95% CI	*P*	No. at risk	Deaths	HR	95% CI	*P*
Sex										
Female	225	90	1.00	–		93	31	1.00	–	
Male	364	192	1.52	1.18–1.95	**0.0012**	139	39	0.85	0.53–1.37	0.51
Total	**589**	**282**				**232**	**70**			
Pathological tumor stage										
T1 + T2	128	33	1.00	–		33	9	1.00	–	
T3 + T4	441	230	2.54	1.76–3.65	**< 0.0001**	197	61	1.27	0.63–2.56	0.50
Total	**569**	**263**				**230**	**70**			
Pathological lymph nodes										
N −	299	93	1.00	–		96	20	1.00		
N +	256	160	2.77	2.14–3.58	**< 0.0001**	128	48	2.08	1.24–3.51	**0.006**
Total	**555**	**253**				**224**	**68**			
Pathological metastases										
M −	427	147	1.00	–		186	48			
M +	162	135	4.29	3.37–5.46	**< 0.0001**	46	22	4.13	2.47–6.90	**< 0.0001**
Total	**589**	**282**				**232**	**70**			
TNM stage										
Stage I	93	20	1.00	–		12	4	1.00	–	
Stage II	172	49	1.36	0.81–2.29	0.24	72	10	0.28	0.09–0.88	**0.03**
Stage III	162	78	2.72	1.66–4.44	**< 0.0001**	102	34	0.76	0.27–2.14	0.60
Stage IV	162	135	7.51	4.67–12.1	**< 0.0001**	46	22	2.35	0.81–6.82	0.12
Total	**589**	**282**				**232**	**70**			
Localization										
Colon	403	187	1.00	–		167	43	1.00	–	
Rectum	186	95	1.03	0.80–1.32	0.83	65	27	1.57	0.97–2.55	0.07
Total	**589**	**282**				**232**	**70**			
Age at diagnosis										
< 65	283	129	1.00	–		140	41	1.00	–	
≥ 65	306	153	1.02	1.01–1.03	**0.007**	92	29	1.02	1.00–1.05	**0.05**
Total	**589**	**282**				**232**	**70**			
Mean (range)	64.21 (27–90)					61.25 (29–86)				
Median	65					62				
SD	10.73					10.67				
Chemotherapy										
5-Fluorouracil	–	–	–	–	–	166	50	–	–	–
Folfox	–	–	–	–	–	66	20	–	–	–
Total	–	–	–	–	–	**232**	**70**	–	–	–

P < 0.05 are indicated in bold.
CI, confidence interval; HR, hazard ratio; SD, standard deviation.

rs289747 ($P = 0.003$), remained statistically significant even after Bonferroni correction ($P = 0.005$).

Discussion

In this study carried out in a cohort of 589 CRC patients, we reported a significant association between two *NLRC5* polymorphisms, rs27194 and rs289747, and CRC survival. In addition, among the CRC patients, who underwent the 5-FU-based therapy, we showed a significant association between rs12445252 and survival.

rs27194 is a 3′UTR variant, which might be responsible for aberrant *NLRC5* expression in CRC by affecting several miRNA binding sites (*http://vm24141.virt.gwdg.de/cgi-bin/microsniper/process.cgi*), including miR-942. miR-942 is known to activate the Wnt/β-catenin pathway [13], which is constitutively activated in many cancers, including CRC [14]. Furthermore, rs27194 is predicted to affect the binding site affinity of PRRX2 and TCF4, involved in the transforming growth factor-β pathway [15] and Wnt/β-catenin pathway [14], respectively, two strongly deregulated in CRC [14,16].

rs289747 is an intronic variant, mapping to promoter histone marks and DNase hypersensitivity sites in several tissues, including blood and the gastrointestinal tract. On the basis of in-silico predictions, it is presumed to affect an OCT1 binding site, increasing its binding affinity. Overexpression of *OCT1* has been reported in many cancers, including CRC [17]. In addition, the IFN*γ* promoter has been reported to contain an OCT1 binding site [18]. As a consequence, the secretion of *IFNγ* by OCT proteins might be increased, leading to a dysregulation of the downstream pathway genes' expression, such as *NLRC5* [1].

The polymorphism associated with survival after 5-FU treatment, rs12445252, is an intronic eQTL variant, which negatively influenced the expression of *NLRC5* (effect size: 0.25, *P* value: 8.9e – 9, *https://www.gtexportal.org/home/*). The mechanisms underlying the 5-FU capacity to selectively target MDSCs remain unclear. However, depletion of MDSCs results in an enhanced T-cell-dependent antitumor immune activity [19]. Increased levels of IFNγ secretion stimulate the expression of downstream genes involved in antitumor immunity, among which is *NLRC5* [1]. Therefore, a decreased expression of *NLRC5* could affect the chemotherapeutic efficacy of 5-FU.

Studies on expression levels of *NLRC5* in CRC and normal tissue are sparse and contradictory: one study did not show any difference between the CRC and healthy tissue [20], whereas another study showed an increased expression of *NLRC5* in CRC compared with normal tissue, probably because of the high inflammatory state in

Copyright © 2018 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

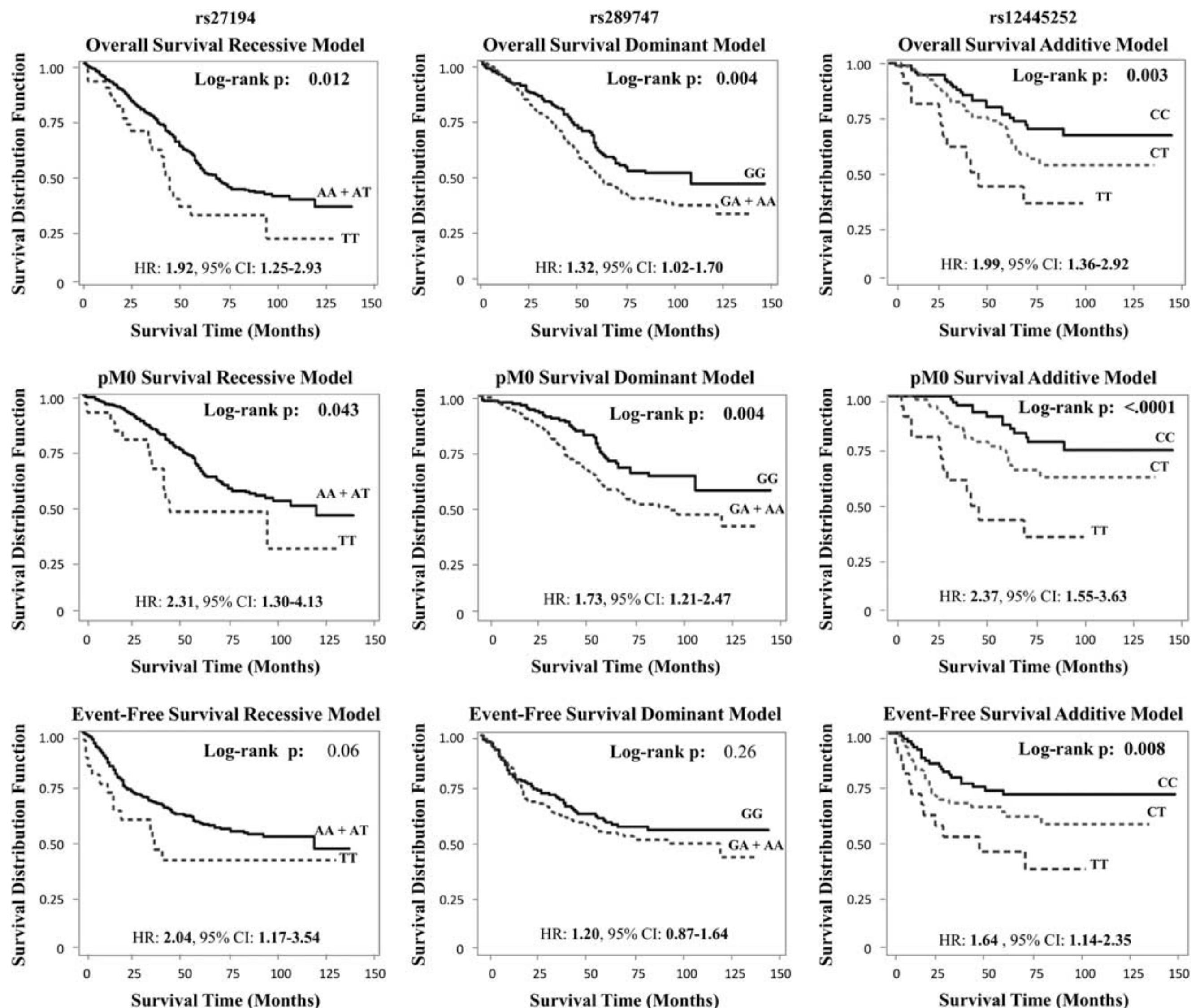


Fig. 1. Kaplan–Meier analysis of survival according to genotypes of rs27194 and rs289747 in the whole study population ($n=589$ for OS_{all} and $n=427$ for OS_{pM0} and EFS_{pM0}) and rs12445252 in 5-fluorouracil-treated patients ($n=232$ for OS_{all} and $n=186$ for OS_{pM0} and EFS_{pM0}). Hazard ratios (HRs) and 95% confidence intervals (CIs) for each presented model are also shown. EFS, event-free survival; OS, overall survival.

this cancer type [3]. Moreover, our own unpublished data showed varying degrees of *NLRC5* expression in healthy gut tissues, with the highest expression in colon adenocarcinoma; *NLRC5* was also inducible by IFNγ in HCT116 cells. A few studies have addressed the role of *NLRC5* and/or *MHC class I* gene expression in survival of cancer patients [3,21,22]. All these findings have indicated high expression of *NLRC5* and/or *MHC class I* genes as good prognostic markers, also in CRC. The study by Yoshihama *et al*. [3] specifically focused on transcriptional regulation of *NLRC5* expression. They showed that both genetic and epigenetic mechanisms within tumor cells may have an impact on NLRC5 activity and, consequently, on MHC class I-dependent immune responses [3]. Our study on germline variants and their influence on CRC survival potentially through transcriptional regulation adds a new layer on the complex function of NLRC5 in the innate immune system. The clinical applicability of these SNPs as prognostic markers warrants further validation in independent cohorts, as well as experimental validation to confirm the in-silico-predicted effects of these SNPs on *NLRC5* expression. Once validation has been performed, it is eventually possible to design specific therapeutic strategies for patients with a predicted worse survival outcome.

Conclusion

Our results suggest that polymorphisms in immune surveillance genes, such as *NLRC5*, may serve as candidate prognostic markers of the clinical outcome of CRC. Furthermore, our results indicate a potential prognostic role of *NLRC5* rs12445252 in the survival of CRC patients in response to 5-FU treatment.

Acknowledgements

C.C., A.W., A.F.: conceived and designed the study. C.C.: performed the experiments. C.C., M.S.: analyzed the data.

Copyright © 2018 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

C.C., A.F.: wrote the paper. C.C., K.J., V.V., P.V., M.L., V.L., O.V.: sample collection and management. C.C., K.J., V.V., L.V., M.L., V.L., O.V.: database management. C.C., K.J., V.V., P.V., M.L., A.W., A.F.: revised the paper critically. C.C., M.S., K.J., V.V., M.L., V.L., O.V., L.V., N.A., K.H., P.V., A.W., A.F.: final approval of the paper to be published.

Conflicts of interest

Dr Pavel Vodicka and Dr Veronika Vymetalkova are currently receiving a grant (AZV 15-27580A and AZV 15-26535A) from Internal Grant Agency of the Ministry of Health of the Czech Republic. Katerina Jiraskova is currently receiving a grant (GAUK 112515) from the Grant Agency of the Charles University. Professor Dr Alexander Weber was supported by the University of Tübingen, the Federal State Baden-Württemberg Junior Professor Program. For the remaining authors there are no conflicts of interest.

References

1 Meissner TB, Li A, Biswas A, Lee KH, Liu YJ, Bayir E, *et al*. NLR family member NLRC5 is a transcriptional regulator of MHC class I genes. *Proc Natl Acad Sci USA* 2010; 107:13794–13799.

2 Ludigs K, Seguín-Estévez Q, Lemeille S, Ferrero I, Rota G, Chelbi S, *et al*. NLRC5 exclusively transactivates mhc class I and related genes through a distinctive SXY module. *PLoS Genet* 2015; 11:e1005088.

3 Yoshihama S, Roszik J, Downs I, Meissner TB, Vijayan S, Chapuy B, *et al*. NLRC5/MHC class I transactivator is a target for immune evasion in cancer. *Proc Natl Acad Sci USA* 2016; 113:5999.

4 Rose MG, Farrell MP, Schmitz JC. Thymidylate synthase: a critical target for cancer chemotherapy. *Clin Colorectal Cancer* 2002; 1:220–229.

5 Umansky V, Blattner C, Gebhardt C, Utikal J. The role of myeloid-derived suppressor cells (MDSC) in cancer progression. *Vaccines (Basel)* 2016; 4:E36.

6 Apetoh L, Ghiringhelli F. Enhancing the anticancer effects of 5-fluorouracil: current challenges and future perspectives. *Biomed J* 2015; 38:111.

7 Draghiciu O, Lubbers J, Nijman HW, Daemen T. Myeloid derived suppressor cells-an overview of combat strategies to increase immunotherapy efficacy. *Oncoimmunology* 2015; 4:e954829.

8 Apetoh L, Végran F, Ladoire S, Ghiringhelli F. Restoration of antitumor immunity through selective inhibition of myeloid derived suppressor cells by anticancer therapies. *Curr Mol Med* 2011; 11:365–372.

9 Vincent J, Mignot G, Chalmin F, Ladoire S, Bruchard M, Chevriaux A, *et al*. 5-Fluorouracil selectively kills tumor-associated myeloid-derived suppressor cells resulting in enhanced T cell-dependent antitumor immunity. *Cancer Res* 2010; 70:3052–3061.

10 Martin F, Apetoh L, Ghiringhelli F. Role of myeloid-derived suppressor cells in tumor immunotherapy. *Immunotherapy* 2012; 4:43–57.

11 Sun HL, Zhou X, Xue YF, Wang K, Shen YF, Mao JJ, *et al*. Increased frequency and clinical significance of myeloid-derived suppressor cells in human colorectal carcinoma. *World J Gastroenterol* 2012; 18:3303–3309.

12 Vymetalkova V, Pardini B, Rosa F, Jiraskova K, Di Gaetano C, Bendova P, *et al*. Polymorphisms in microRNA binding sites of mucin genes as predictors of clinical outcome in colorectal cancer patients. *Carcinogenesis* 2017; 38:28–39.

13 Ge C, Wu S, Wang W, Liu Z, Zhang J, Wang Z, *et al*. miR-942 promotes cancer stem cell-like traits in esophageal squamous cell carcinoma through activation of Wnt/β-catenin signalling pathway. *Oncotarget* 2015; 6:10964–10977.

14 Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene* 2017; 36:1461–1473.

15 Juang YL, Jeng YM, Chen CL, Lien HC. PRRX2 as a novel TGF-β-induced factor enhances invasion and migration in mammary epithelial cell and correlates with poor prognosis in breast cancer. *Mol Carcinog* 2016; 55:2247–2259.

16 Xu Y, Pasche B. TGF-beta signaling alterations and susceptibility to colorectal cancer. *Hum Mol Genet* 2007; 16 (Spec No 1):R14–R20.

17 Wang YP, Song GH, Chen J, Xiao C, Li C, Zhong L, *et al*. Elevated OCT1 participates in colon tumorigenesis and independently predicts poor prognoses of colorectal cancer patients. *Tumour Biol* 2016; 37:3247–3255.

18 Mueller K, Quandt J, Marienfeld RB, Weihrich P, Fiedler K, Claussnitzer M, *et al*. Octamer-dependent transcription in T cells is mediated by NFAT and NF-κB. *Nucleic Acids Res* 2013; 41:2138–2154.

19 Ghiringhelli F, Bruchard M, Apetoh L. Immune effects of 5-fluorouracil: ambivalence matters. *Oncoimmunology* 2013; 2:e23139.

20 Liu R, Truax AD, Chen L, Hu P, Li Z, Chen J, *et al*. Expression profile of innate immune receptors, NLRs and AIM2, in human colorectal cancer: correlation with cancer stages and inflammasome components. *Oncotarget* 2015; 6:33456–33469.

21 Watson NFS, Ramage JM, Madjd Z, Spendlove I, Ellis IO, Scholefield JH, *et al*. Immunosurveillance is active in colorectal cancer as downregulation but not complete loss of MHC class I expression correlates with a poor prognosis. *Int J Cancer* 2006; 118:6–10.

22 Simpson JAD, Al-Attar A, Watson NFS, Scholefield JH, Ilyas M, Durrant LG. Intratumoral T cell infiltration, MHC class I and STAT1 as biomarkers of good prognosis in colorectal cancer. *Gut* 2010; 59:926–933.

Copyright © 2018 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

<u>Manuscript VII</u>

Jiraskova K, Hughes DJ, Brezina S, Gumpenberger T, Veskrnova V, Buchler T, Schneiderova M, Levy M, Liska V, Vodenkova S, Di Gaetano C, Naccarati A, Pardini B, Vymetalkova V, Gsur A, Vodicka P.

Functional polymorphisms in DNA repair genes are associated with sporadic colorectal cancer susceptibility and clinical outcome

Int J Mol Sci. (2018) 20(1). pii: E97

International Journal of
Molecular Sciences

MDPI

Article

Functional Polymorphisms in DNA Repair Genes Are Associated with Sporadic Colorectal Cancer Susceptibility and Clinical Outcome

Katerina Jiraskova [1,2,*], David J. Hughes [3], Stefanie Brezina [4], Tanja Gumpenberger [4], Veronika Veskrnova [5], Tomas Buchler [5], Michaela Schneiderova [6], Miroslav Levy [7], Vaclav Liska [8,9], Sona Vodenkova [1,2,10], Cornelia Di Gaetano [11,12], Alessio Naccarati [2,11], Barbara Pardini [11,12], Veronika Vymetalkova [1,2,8], Andrea Gsur [4] and Pavel Vodicka [1,2,8,*]

1. Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Albertov 4, 128 00 Prague, Czech Republic; sona.vodenkova@iem.cas.cz (S.V.); veronika.vymetalkova@iem.cas.cz (V.V.)
2. Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Videnska 1083, 142 00 Prague, Czech Republic; alessio.naccarati@iigm.it
3. Cancer Biology and Therapeutics Group, UCD Conway Institute, University College Dublin, Dublin 4, Ireland; david.hughes@ucd.ie
4. Institute of Cancer Research, Department of Medicine I, Medical University of Vienna, Borschkegasse 8a, A-1090 Vienna, Austria; stefanie.brezina@meduniwien.ac.at (S.B.); tanja.gumpenberger@meduniwien.ac.at (T.G.); andrea.gsur@meduniwien.ac.at (A.G.)
5. Department of Oncology, First Faculty of Medicine, Charles University and Thomayer Hospital, Videnska 800, 140 59 Prague, Czech Republic; veskrnova.ver@gmail.com (V.V.); buchlert@yahoo.com (T.B.);
6. Department of Surgery, General University Hospital in Prague, U Nemocnice 499/2, 128 08 Prague, Czech Republic; schneider77@seznam.cz
7. Department of Surgery, First Faculty of Medicine, Charles University and Thomayer Hospital, Thomayerova 815/5, 140 00 Prague, Czech Republic; miroslav.levy@lf1.cuni.cz
8. Biomedical Centre, Faculty of Medicine in Pilsen, Charles University in Prague, 323 00 Pilsen, Czech Republic; vena.liska@skaut.cz
9. Department of Surgery, Medical School in Pilsen, Charles University, Alej svobody 80, 304 600 Pilsen, Czech Republic
10. Department of Medical Genetics, Third Faculty of Medicine, Charles University, Ruska 2411/87, 100 00 Prague, Czech Republic
11. Molecular and Genetic Epidemiology; Genomic Variation in Human Populations and Complex Diseases, IIGM Italian Institute for Genomic Medicine, Via Nizza 52, 10126 Turin, Italy; cdigaeta@unito.it (C.D.G.); barbara.pardini@iigm.it (B.P.)
12. Department of Medical Sciences, University of Turin, Corso Dogliotti 14, 10126 Turin, Italy

* Correspondence: katerina.jiraskova@iem.cas.cz (K.J.); pavel.vodicka@iem.cas.cz (P.V.); Tel.: +420-241-062-251 (K.J.); +420-241-062-694 (P.V.)

Received: 27 November 2018; Accepted: 21 December 2018; Published: 27 December 2018

Abstract: DNA repair processes are involved in both the onset and treatment efficacy of colorectal cancer (CRC). A change of a single nucleotide causing an amino acid substitution in the corresponding protein may alter the efficiency of DNA repair, thus modifying the CRC susceptibility and clinical outcome. We performed a candidate gene approach in order to analyze the association of non-synonymous single nucleotide polymorphisms (nsSNPs) in the genes covering the main DNA repair pathways with CRC risk and clinical outcome modifications. Our candidate polymorphisms were selected according to the foremost genomic and functional prediction databases. Sixteen nsSNPs in 12 DNA repair genes were evaluated in cohorts from the Czech Republic and Austria. Apart from the tumor-node-metastasis (TNM) stage, which occurred as the main prognostic factor in all of the performed analyses, we observed several significant associations of different nsSNPs with survival and clinical outcomes in both cohorts. However, only some of the genes (*REV3L*, *POLQ*, and *NEIL3*) were prominently defined as prediction factors in the classification and regression tree

analysis; therefore, the study suggests their association for patient survival. In summary, we provide observational and bioinformatics evidence that even subtle alterations in specific proteins of the DNA repair pathways may contribute to CRC susceptibility and clinical outcome.

Keywords: DNA repair genes; functional single nucleotide polymorphism; colorectal cancer susceptibility; survival analysis

---

1. Introduction

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer death worldwide [1]. In Europe, the highest incidence rates are reported in Eastern and Central European countries, such as the Czech Republic and Austria [2]. CRC represents a multifactorial disease associated with several genetic and environmental factors [3].

The prognosis for patients with CRC is heavily dependent on stage at diagnosis; the five-year survival rate is up to 90% for stage I, but only <15% for stage IV [4]. Over half of the cases are diagnosed at an advanced stage of disease (III and IV), with treatment usually involving complete primary tumor resection and appropriate chemotherapy. While the treatment can reduce the risk of relapse and increase patients' survival, it can also cause severe side effects and impair quality of life [5]. The differences in medication response are considerably affected by individual inherited genetic susceptibility. Current approaches to choose and implement chemotherapy regimens for CRC patients are primarily determined by tumor staging and histopathological examination. Developing prognostic and predictive biomarkers based on a personal genetic background would greatly aid the selection of an optimal treatment by oncologists, so as to improve clinical outcome for each patient.

Genetics plays a key role in predisposition to CRC, its initiation, and progression [6]. Several studies provided evidence that single nucleotide polymorphisms (SNPs) in DNA repair genes could alter DNA repair function, modulate its capacity, and thus induce genetic instability or unregulated cell growth and cancer [7–9]. In the last decade, while association studies (including genome-wide) have identified multiple SNPs involved in CRC susceptibility, none have been validated as biomarkers for clinical use [10–14]. Furthermore, most of the anticancer agents are targeted to induce DNA damage, which overwhelms the cellular DNA repair capacity and thus leads to apoptosis. The most affected are the rapidly dividing cells, such as cancer cells. Treatment efficacy is therefore influenced by the DNA repair capacity of cancer cells, and the differences in treatment response might be affected by the inherited variations of genes encoding DNA repair enzymes [15].

In this study, we hypothesized that SNPs causing amino acid substitution (non-synonymous SNPs—nsSNPs) in DNA repair genes that are known to be involved in maintaining genome stability (cancer prevention) and in chemotherapy response (cancer treatment), may influence CRC susceptibility and modulate the clinical outcome after cancer diagnosis. We evaluated the association of 16 nsSNPs in 12 DNA repair genes with CRC risk, post-diagnosis survival, and therapy outcomes in a discovery set of 1832 patients and 1172 controls from the Czech Republic and in an independent replication set comprising 950 patients and 820 controls from Austria.

2. Results

2.1. SNP Selection

In total, sixteen nsSNPs in 12 genes passed the selection criteria and were successfully genotyped and analyzed in the Czech cohort (Table 1). The same nsSNPs were analyzed in the replication Austrian cohort, except for two nsSNPs (*FAAP24* rs3816032 and *MUS81* rs545500), where the genome-wide association study (GWAS) data were not available.

Table 1. Selected single nucleotide polymorphisms (SNPs) in DNA repair genes, with a minor allele frequency of [a] > 0.10 in the European population.

Genomic Annotation							**Functional Genomics**					**Comparative Genomics**	
Gene ID	DNA Repair Pathway	UniProtKB	SNP ID	Base Change	Amino Acid Change	MAF in EUR [a]	LD with Other SNPs Associated with CRC	LD within the Same Gene	F-SNP Prediction Result (on Protein Coding)	ELASPIC (ΔΔG)	DUET (ΔΔG)	Element GERP RS Score >800	SIPHYs
EME1	DSB	Q96AY2	rs12450550	T > C	Ile350Thr	0.24	no	no	deleterious	Destabilizing (Core 1.646)	Destabilizing (−3.002 Kcal/mol)		
FAAP24	DSB	Q9BTP7	rs3816032 [b]	T > C	Ile192Thr	0.11	no	no	deleterious	Destabilizing (Core 1.133)	Destabilizing (−1.653 Kcal/mol)		X
FANCI	DSB	Q9NVI1	rs2283432	C > G	Cys742Ser	0.38	no	no	deleterious	NA	NA	X	X
MUS81	DSB	Q96NY9	rs545500 [b]	C > G	Arg180Pro	0.33	no	no	deleterious	NA	NA		X
NEIL3	BER	Q8TAT5	rs7689099	C > G	Pro117Arg	0.12	no	no	deleterious	NA	NA	X	X
POLE	BER, DSB, NER	Q07864	rs5744934	A > G	Asn1396Ser	0.13	no	no	deleterious	NA	NA		X
POLN	DSB	Q7Z5Q5	rs2353552	C > A	Gln121His	0.13	no	no	deleterious	NA	NA		
			rs9328764	A > G	Arg425Cys	0.12	no	no	deleterious	NA	Destabilizing (−1.765 Kcal/mol)		X
POLQ	DSB	O75417	rs1381057	C > T	Gln2513Arg	0.33	no	no	deleterious	Destabilizing (Core 1.648)	NA		X
			rs3218649	C > G	Thr982Arg	0.39	no	no	deleterious	NA	NA	X	
			rs3218651	T > C	His1201Arg	0.15	no	no	damaging	NA	NA	X	
RAD51D	DSB	O75771	rs4796033	C > T	Arg165Gln	0.13	no	no	deleterious	Destabilizing (Core 1.843)	NA		X
REV1	DSB	Q9UBZ9	rs3087386	G > A	Phe257Ser	0.43	no	no	deleterious	NA	NA	X	
			rs3087399	A > G	Asn373Ser	0.12	no	no	deleterious	NA	Destabilizing (−0.596 Kcal/mol)	X	X
REV3L	DSB	O60673	rs3204953	G > A	Val2986Ile	0.17	no	no	deleterious	Destabilizing (Core 1.965)	NA	X	X
RPA1	BER, DSB, NER	P27694	rs5030755	A > G	Thr351Ala	0.10	no	no	deleterious	NA	Destabilizing (−1.037 Kcal/mol)		X

SNP—single nucleotide polymorphism; MAF—minor allele frequency; EUR—European population; LD—linkage disequilibrium; CRC—colorectal cancer; GERP—Genomic Evolutionary Rate Profiling; DSB—double strand break repair pathway; BER—base excision repair pathway; NER—nucleotide excision repair pathway; NA—not applicable; X—evolutionary conserved position. [a] https://www.ensembl.org/index.html. [b] Data for SNPs were not available in the Austrian cohort.

2.2. Case-Control Study

The characteristics of the study participants are shown in Table 2. Compared with controls, CRC cases in the Czech cohort had a slightly higher prevalence of male individuals, and were more likely to be older, to smoke, to have diabetes mellitus, and a positive family history of CRC. In the Austrian cohort, CRC cases were more often males and smokers.

For all of the SNPs, the distribution of the genotypes within the studied genes in controls was in agreement with the Hardy–Weinberg equilibrium. The SNPs significantly associated with CRC risk are presented in Table 3.

Czech cohort. The carriers of the *TC* genotype in *EME1* rs12450550 had an associated increased CRC risk (*TC* vs. *TT*; odds ratio (OR) 1.19; 95% confidence interval (CI) 1.00–1.40; $p = 0.05$), with the same tendency observed for the presence of the variant C allele in the dominant model (*TC+CC* vs. *TT*; OR 1.19; 95% CI 1.02–1.40; $p = 0.03$). However, this association should be considered cautiously due to the low frequency of the variant *CC* genotype in this study group. The variant *AA* genotype in *REV3L* rs3204953 was associated with an increased risk of CRC in the codominant and recessive model (*AA* vs. *GG*; OR 2.32; 95% CI 1.27–4.25; $p = 0.006$; and *AA* vs. *GG+GA*; OR 2.28; 95% CI 1.24–4.17; $p = 0.008$). After stratification according to the tumor site, the association was similar in colon cancer patients (*AA* vs. *GG*; OR 2.59; 95% CI 1.36–4.91; $p = 0.004$; and *AA* vs. *GG+GA*; OR 2.52; 95% CI 1.33–4.77; $p = 0.005$). By considering multiple testing correction using the Benjamini–Hochberg false discovery rate (FDR) measure, only the association for rs3204953 in *REV3L* remained significant ($q = 0.01$).

Austrian cohort. The association of the SNP genotypes with CRC risk was observed only for the rectal cancer patients. Carriers of the *CT* genotype in *POLQ* rs1381057 were at an increased risk of the disease (*CT* vs. *CC*; OR 1.32; 95% CI 1.01–1.74; $p = 0.04$), with the same tendency observed in the presence of the variant T allele in the dominant model (*CT+TT* vs. *CC*; OR 1.34; 95% CI 1.04–1.74; $p = 0.03$). The variant *GG* genotype in *REV1* rs3087399 was associated with a decreased risk of rectal cancer in the codominant and recessive model (*GG* vs. *AA*; OR 0.13; 95% CI 0.02–0.96; $p = 0.05$; and *GG* vs. *AA+AG*; OR 0.12; 95% CI 0.02–0.90; $p = 0.04$). None of the associations remained significant after the Benjamini–Hochberg correction.

2.3. Survival Analyses

In total, 1832 Czech and 950 Austrian CRC cases were included in the survival analyses. In the univariate assessment, several covariates were associated with survival, including established prognostic factors such as male sex, higher age, smoking habit, and cancer stage, which were associated with decreased patients' survival and increased risk of recurrence (Table 4).

Czech cohort. Overall, no SNPs were associated with either the overall survival (OS) or event free survival (EFS). However, after stratification according to tumor localization, nominally significant associations were detected for six SNPs in the univariate assessment (Table S1). In colon cancer patients, two SNPs were associated with increased EFS (rs3816032 and rs2283432; $p = 0.02$ for either variant genotype). In rectal cancer patients, one SNP was associated with an increased EFS (rs7689099; $p = 0.02$), and three with decreased OS or EFS (rs545500, rs3218649, and rs3087386; $p = 0.02$, 0.02, and 0.03, respectively).

Austrian cohort. Four SNPs were associated with either OS or EFS (Table S1). Three SNPs were observed in association with increased OS or EFS in CRC patients (rs12450550, rs2283432, and rs3204953; $p = 0.03$, $p = 0.02$, and $p = 0.02$, respectively). Rs3087386 was found to be significantly associated with decreased OS and EFS in colon cancer patients ($p = 0.02$).

Table 2. Characteristics of the study populations.

Variables		Czech Republic					Austria				
		Controls No. (%)	Cases No. (%)	OR	95% CI	p-Value	Controls No. (%)	Cases No. (%)	OR	95% CI	p-Value
Sex	Female	478 (40.8)	696 (38.1)	Ref			353 (43.1)	389 (40.9)	Ref		
	Male	694 (59.2)	1133 (61.9)	1.20	1.02–1.40	0.03	467 (56.9)	561 (59.1)	1.39	1.09–1.78	0.01
Age (years)	<50	269 (22.9)	140 (9.2)	Ref			92 (11.2)	129 (13.5)	Ref		
	(50, 60]	546 (46.6)	433 (28.4)	1.52	1.20–1.94	0.0006	147 (17.9)	208 (21.9)	2.18	0.72–1.43	0.92
	(60, 70]	183 (15.6)	639 (42.0)	6.71	5.16–8.72	<0.0001	282 (34.4)	323 (34.0)	0.82	0.60–1.13	0.22
	>70	174 (14.9)	310 (20.4)	3.42	2.60–4.51	<0.0001	299 (36.5)	291 (30.6)	0.70	0.51–0.96	0.03
BMI	(18.5, 25]	93 (8.0)	358 (23.5)	Ref			189 (23.7)	296 (35.7)	Ref		
	<18.5	334 (28.5)	370 (24.3)	3.22	2.45–4.24	<0.0001	2 (0.3)	17 (2.0)	5.43	1.24–23.76	0.02
	(25, 30]	529 (45.1)	508 (33.4)	0.84	0.69–1.02	0.08	364 (45.6)	336 (40.5)	0.59	0.47–0.75	<0.0001
	>30	213 (18.4)	286 (18.8)	1.13	0.89–1.43	0.31	243 (30.4)	181 (21.8)	0.48	0.37–0.62	<0.0001
Smoking habit	No	638 (57.6)	769 (53.1)	Ref			447 (55.7)	251 (48.8)	Ref		
	Yes [a]	470 (42.4)	679 (46.9)	1.33	1.13–1.56	<0.001	356 (44.3)	263 (51.2)	1.30	0.97–1.73	0.08
DM	No	555 (85.5)	1076 (80.4)	Ref			370 (82.8)	817 (86.0)	Ref		
	Yes	94 (14.5)	263 (19.6)	1.41	1.09–1.84	0.01	77 (17.2)	133 (14.0)	0.62	0.42–0.92	0.02
Family history of CRC	No	942 (89.3)	1103 (84.1)	Ref			NDA	NDA			
	Yes	113 (10.7)	209 (15.9)	1.65	1.28–2.12	<0.001	NDA	NDA			
Diagnosis	Colon		1192 (65.8)					586 (62.6)			
	Rektum		621 (34.2)					350 (37.4)			
tnm stage	I		277 (16.8)					188 (21.2)			
	II		498 (30.2)					227 (25.5)			
	III		491 (29.8)					354 (39.8)			
	IV		384 (23.3)					120 (13.5)			
Chemotherapy	None		795 (49.9)					389 (43.0)			
	5-FU		494 (31.0)					253 (28.0)			
	5-FU combined with oxaliplatin		303 (19.1)					262 (29.0)			

OR—odds ratio; CI—confidence interval; BMI—body mass index; DM—diabetes mellitus; CRC—colorectal cancer; TNM—tumor-node-metastasis; 5-FU—5-fluorouracil; NDA—no data available. Significant results are in bold. Numbers may not add up to 100% of available subjects because of missing data. [a] Ex-smokers included in smokers.

Table 3. Associations of SNPs in DNA repair genes with the risk of CRC and its major sub-sites (colon/rectum).

Gene SNP	Genotype	Controls [a]	All CRC Patients Cases [a]	OR [b]	95% CI	*p*-Value	Colon Cancer Patients Cases [a]	OR [b]	95% CI	*p*-Value	Rectal Cancer Patients Cases [a]	OR [b]	95% CI	*p*-Value	HWE [c] X^2, *p*-Value
Czech Republic															
EME1 rs12450550	*TT*	678	815	Ref			526	Ref			284	Ref			1.07, 0.58
	TC	410	570	1.19	1.00–1.40	0.05	363	1.17	0.97–1.40	0.11	198	1.20	0.96–1.50	0.11	
	CC	73	108	1.24	0.90–1.70	0.20	64	1.15	0.80–1.65	0.46	41	1.38	0.91–2.09	0.13	
	TC+CC	483	678	1.19	1.02–1.40	0.03	427	1.16	0.97–1.39	0.10	239	1.23	0.99–1.52	0.06	
	TT+TC	1088	1385	Ref			889	Ref			482	Ref			
	CC	73	108	1.16	0.84–1.59	0.36	64	1.08	0.76–1.55	0.66	41	1.28	0.86–1.93	0.23	
REV3L rs3204953	*GG*	839	1049	Ref			666	Ref			371	Ref			4.68, 0.10
	GA	304	405	1.09	0.91–1.30	0.37	261	1.12	0.91–1.37	0.27	139	1.06	0.83–1.34	0.66	
	AA	15	43	2.32	1.27–4.25	0.006 *	30	2.59	1.36–4.91	0.004 *	13	1.97	0.92–4.22	0.08	
	GA+AA	319	448	1.14	0.96–1.36	0.13	291	1.19	0.98–1.44	0.08	152	1.10	0.87–1.39	0.42	
	GG+GA	1143	1454	Ref			927	Ref			510	Ref			
	AA	15	43	2.28	1.24–4.17	0.008 *	30	2.52	1.33–4.77	0.005 *	13	1.95	0.91–4.18	0.09	
Austria															
POLQ rs1381057	*CC*	372	413	Ref			267	Ref			134	Ref			1.49, 0.47
	CT	349	423	1.09	0.90–1.34	0.38	250	1.00	0.80–1.25	1.00	166	1.32	1.01–1.74	0.04	
	TT	99	114	1.05	0.77–1.42	0.76	65	0.93	0.65–1.32	0.68	49	1.40	0.94–2.08	0.10	
	CT+TT	448	537	1.08	0.90–1.31	0.40	315	0.98	0.80–1.22	0.89	215	1.34	1.04–1.74	0.03	
	CC+CT	721	836	Ref			517	Ref			300	Ref			
	TT	99	114	1.00	0.75–1.34	0.98	65	0.93	0.66–1.29	0.66	49	1.21	0.84–1.75	0.32	
REV1 rs3087399	*AA*	593	673	Ref			414	Ref			243	Ref			0.02, 0.99
	AG	208	259	1.10	0.89–1.36	0.39	151	1.04	0.81–1.32	0.78	105	1.25	0.95–1.66	0.11	
	GG	19	18	0.83	0.43–1.60	0.58	17	1.27	0.65–2.48	0.48	1	0.13	0.02–0.96	0.05	
	AG+GG	227	277	1.08	0.87–1.32	0.50	168	1.06	0.83–1.34	0.65	106	1.16	0.88–1.53	0.30	
	AA+AG	801	932	Ref			565	Ref			348	Ref			
	GG	19	18	0.81	0.42–1.56	0.53	17	1.26	0.65–2.44	0.50	1	0.12	0.02–090	0.04	

OR—odds ratio; CI—confidence interval. Nominally significant results are in bold. Results that passed the Benjamini–Hochberg test for multiple comparisons are marked with an asterisk. [a] Numbers may not add up to 100% of subjects due to genotyping failure. All of the samples that did not give a reliable result in the first round of genotyping were retested in up to two additional rounds. Samples failing these procedures were omitted from the analysis. [b] Logistic regression analysis values are adjusted for age. [c] X^2 and *p*-values for the deviation of the observed and of the numbers expected from the Hardy–Weinberg equilibrium (HWE) in the controls.

Table 4. Clinical characteristics significantly affecting overall survival (OS) and event free survival (EFS) in CRC patients with complete follow up.

Variables		Czech Republic					Austria				
		N [a]	OS		EFS		N [a]	OS		EFS	
			HR (95% CI)	*p*-Value	HR (95% CI)	*p*-Value		HR (95% CI)	*p*-Value	HR (95% CI)	*p*-Value
Sex	Female	696	Ref		Ref		389	Ref		Ref	
	Male	1133	1.47 (1.20–1.80)	0.0002	1.29 (1.09–1.52)	0.003	561	1.37 (1.03–1.83)	0.03	1.43 (1.11–1.83)	0.005
Age (years)	<50	149	Ref		Ref		129	Ref		Ref	
	(50, 60]	433	0.96 (0.62–1.50)	0.87	1.06 (0.76–1.49)	0.72	208	1.44 (0.77–2.69)	0.26	1.62 (0.99–2.65)	0.05
	(60, 70]	639	1.08 (0.71–1.65)	0.72	0.90 (0.65–1.25)	0.54	323	2.14 (1.21–3.79)	0.01	1.68 (1.05–2.68)	0.03
	>70	610	1.47 (0.97–2.24)	0.07	1.05 (0.76–1.46)	0.77	291	3.11 (1.77–5.47)	<0.0001	2.53 (1.60–4.00)	<0.0001
BMI	(18.5, 25]	434	Ref		Ref		296	Ref		Ref	
	<18.5	456	0.99 (0.77–1.27)	0.92	1.06 (0.86–1.32)	0.58	17	1.12 (0.41–3.07)	0.83	1.31 (0.57–3.00)	0.52
	(25, 30]	626	0.83 (066–1.06)	0.13	0.94 (0.77–1.15)	0.54	336	0.79 (0.55–1.12)	0.18	0.85 (0.63–1.15)	0.29
	>30	315	0.58 (0.43–0.80)	0.0008	0.83 (0.65–1.06)	0.13	181	1.13 (0.77–1.65)	0.54	1.04 (0.74–1.46)	0.83
Smoking habit	No	967	Ref		Ref		251	Ref		Ref	
	Yes [b]	777	1.266 (1.049–1.529)	0.01	1.27 (1.08–1.48)	0.003	263	0.93 (0.65–1.32)	0.67	1.02 (0.75–1.39)	0.91
Stage	I	277	Ref		Ref		188	Ref		Ref	
	II	498	1.75 (1.10–2.80)	0.02	1.99 (1.41–2.82)	0.0001	227	1.00 (0.57–1.76)	1.00	0.90 (0.56–1.45)	0.67
	III	491	3.46 (2.22–5.39)	<0.0001	3.45 (2.47–4.83)	<0.0001	354	1.51 (0.92–2.45)	0.10	1.55 (1.03–2.32)	0.03
	IV	384	8.91 (5.78–13.74)	<0.0001	6.00 (4.30–8.38)	<0.0001	120	7.98 (4.95–12.88)	<0.0001	9.33 (6.21–14.02)	<0.0001
5FU-based chemotherapy	No	765	Ref		Ref		389	Ref		Ref	
	Yes	797	1.022 (0.84–1.24)	0.82	1.387 (1.18–1.63)	<0.0001	515	1.33 (0.99–1.78)	0.06	1.79 (1.38–2.32)	<0.0001

HR—hazard ratio; CI—confidence interval; BMI—body mass index. Significant results are in bold. [a] Numbers may not add up to 100% of the available subjects because of missing information. [b] Ex-smokers included in smokers.

2.4. Survival and Therapy

To examine the association of SNPs with the therapy outcome, we further stratified patients according to the treatment received into the following three separate groups: (1) CRC patients receiving no treatment or (2) patients receiving a 5-Fluorouracil (5-FU) regimen without or (3) in combination with oxaliplatin. The group of patients treated with a combination of 5-FU and oxaliplatin was investigated separately, because the latter drug induces a different type of DNA damage compared to 5-FU alone, and thus different DNA repair pathways and genes may be involved [16,17]. The univariate model for survival and therapy showed several genotypes nominally significantly associated with OS or EFS (detailed description in supplementary text and Tables S2, S3, and S4).

2.5. Classification and Regression Tree Survival Analysis

In order to assess the prognostic utility of the investigated DNA repair gene polymorphisms, the interactive effects of genotypes and clinico-pathological parameters in association with five-year OS and EFS were explored using a classification and regression tree (CART) analysis. Only patients with complete data for all of the parameters described in the Material and Methods were included in the analysis (n = 1105 (60%) for the Czech CRCs, and n = 841 (88%) for the Austrian CRCs). The results indicated that the tumor–node–metastasis (TNM) stage was chosen as the initial optimal split factor for predicting both OS and EFS in both of the cohorts (Figures 1–4).

2.5.1. Overall Survival

Czech cohort. The five-year OS analysis resulted in four terminal nodes. Variables determining the structure of the tree included TNM stage, age, sex, chemotherapy, and five SNPs—rs3087386, rs3218649, rs3218651, rs545500, and rs5744934. Among the stage I CRC patients, the subsequent split showed interactions between age and sex. In stage II, the carriers of *GC+CC* genotypes in *POLQ* rs3218649 were associated with a better prognosis. However, the *GG* genotype in females showed almost similar OS prognosis and an even better prognosis when in combination with *CC+CT* genotypes in *REV1* rs3087386 (*CC+CT* 96.4% vs. *TT* 65.2%). In stage III, the subsequent split was age, which was seen to interact with *GG+GC* genotypes in *MUS81* rs545500 and *AG+GG* genotypes in *POLE* rs5744934 (*AG+GG* 94.9% vs. *AA* 68.1%). The *AA* genotype of *POLE* gene further interacted with *POLQ* rs3218649 (*GG+GC* genotypes 75.1% vs. *CC* genotype 32.5%). In stage IV, chemotherapy was the next most significant factor and the level of OS increased when in combination with the rs3218651 variant in *POLQ* gene (*AG+GG* 65.6% vs. *AA* 43.2%). The structure of the tree and corresponding survival curves from terminal nodes are presented in Figure 1.

Austrian cohort. The final tree structure contained six terminal nodes and included nine variables (age, TNM stage, and seven SNPs—rs1381057, rs2283432, rs3204953, rs3218651, rs4796033, rs5030755, and rs5744934). Among CRC patients at stage I, age as the subsequent split showed interactions with the *AA* genotype in *POLE* rs5744934 and *CC+CG* genotypes in *FANCI* rs2283432. The interaction was concluded by *CT+TT* genotypes in *POLQ* rs1381057 (*CT+TT* 97.6% vs. *CC* 84.8%). In stage II, age was the next most significant factor, and the level of OS increased when in combination with the *CC* genotype in *RAD51D* rs4796033 and the *AA* genotype in *POLQ* rs3218649 (*AA* 100% vs. *AG+GG* 92.9%). Carriers of the *CT+TT* genotype in *RAD51D* rs4796033 showed a better prognosis in combination with the *CC* genotype in *POLQ* rs1381057 (*CC* 100% vs. *CT+TT* 77.8%). In stage III, age was the most significant factor for OS. In stage IV, age was further associated with three SNPs (*RPA1* rs5030755 combined with *REV3L* rs3204953 as a terminal node: *GG* 44.4% vs. *GA+AA* 27.3%; and *REV3L* rs3204953 combined with *POLQ* rs3218651 as a terminal node: *AA* 24.2% vs. *AG+GG* 0%). The structure of the tree and corresponding survival curves from terminal nodes are presented in Figure 2.

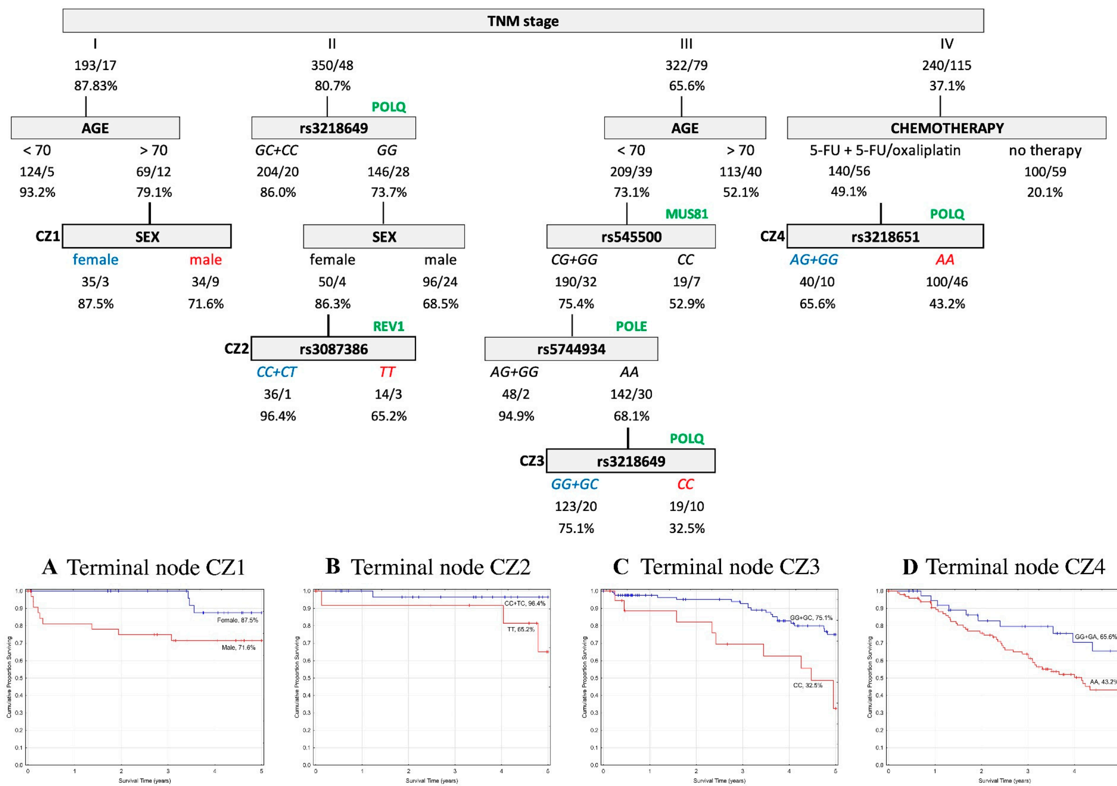


Figure 1. Overall survival (OS) classification and regression tree analysis of colorectal cancer patients from the Czech Republic. Numbers under each node indicate the total number of cases in the subcategory/number of events and percentage of patients with five-year OS. Terminal nodes are bordered in bold, and the corresponding Kaplan–Meier curves are shown underneath. (**A**) Terminal node CZ1; (**B**) Terminal node CZ2; (**C**) Terminal node CZ3; (**D**) Terminal node CZ4.

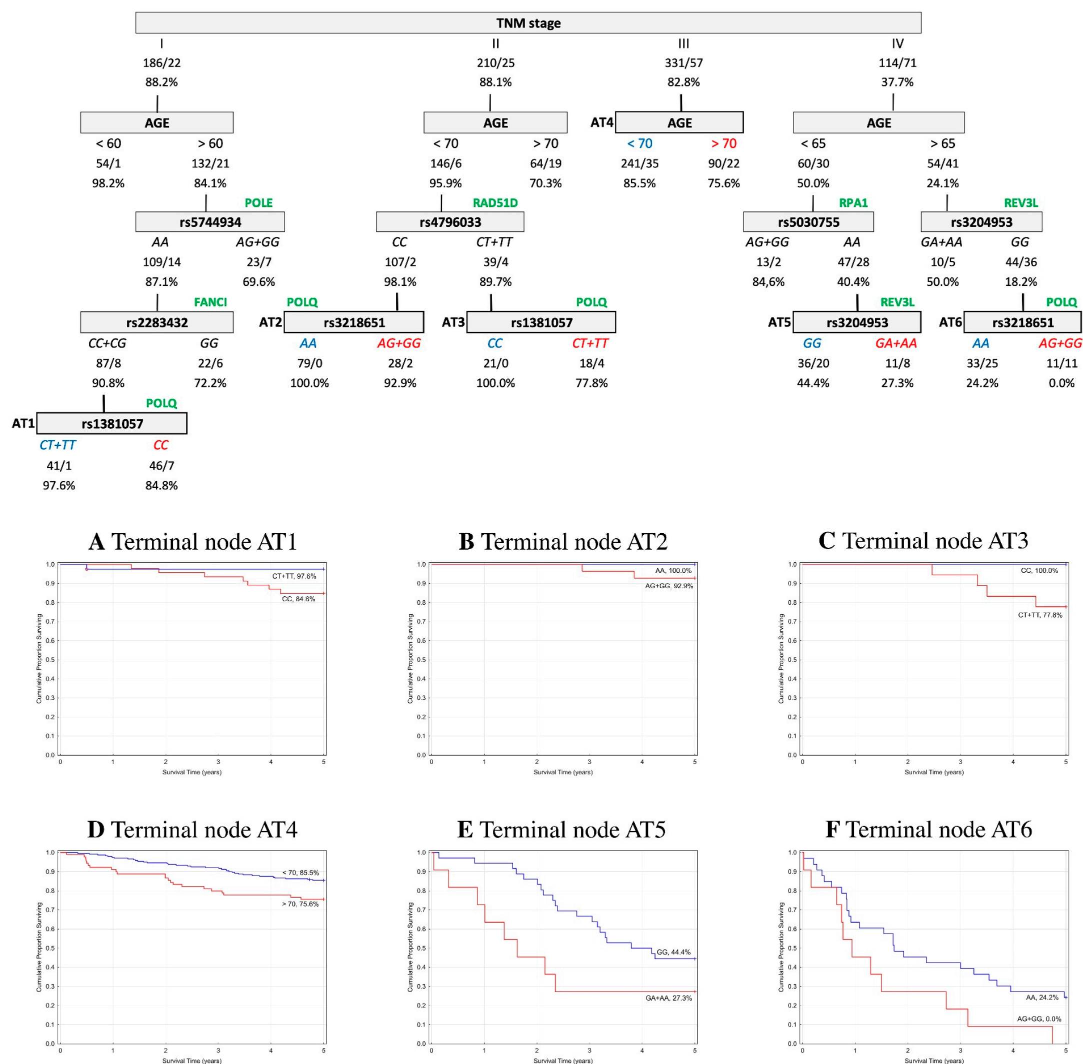


Figure 2. Overall survival (OS) classification and regression tree analysis of colorectal cancer patients from Austria. Numbers under each node indicate the total number of cases in the subcategory/number of events, and the percentage of patients with five-year OS. Terminal nodes are bordered in bold, and the corresponding Kaplan–Meier curves are shown underneath. (**A**) Terminal node AT1; (**B**) Terminal node AT2; (**C**) Terminal node AT3; (**D**) Terminal node AT4; (**E**) Terminal node AT5; (**F**) Terminal node AT6.

2.5.2. Event-Free Survival

Czech cohort. Regarding the five-year EFS, the final tree structure contained five terminal nodes and included 11 variables (age, TNM stage, chemotherapy, and eight SNPs—rs12450550, rs1381057, rs3087399, rs3204953, rs5030755, rs545500, rs5744934, and rs7689099). Among CRC patients at stage I, the subsequent split was for *EME1* rs12450550 (*TT+TC* 83.5% vs. *CC* 52.1%). In stage II, chemotherapy was the first split, when patients with no treatment and those with 5-FU-based therapy without oxaliplatin showed almost the same prognosis level. Patients without treatment had a better prognosis when associated with *CG+GG* genotypes in *MUS81* rs545500 in combination with the *GG* genotype in *NEIL3* rs7689099 (*GG* 75.5% vs. *CC+CG* 56.9%). Patients treated only with 5-FU had a better prognosis when in association with the *GG* genotype in *REV3L* rs3204953 (*GG* 78.6% vs. *GA+AA* 52.3%). On the other hand, the negative association for rs3204953 with the prognosis level was further worsened by the *AA* genotype in *POLE* rs5744934 and *AG+GG* genotype in *REV1* rs3087399 (*AA* 50.0% vs. *AG+GG* 20.5%). In stage III, the subsequent split *AG+GG* genotype in *RPA1* rs5030755 was seen to interact with

patients under 70 years of age and the *CC+CT* genotype in *POLQ* rs1381057 (*CC+CT* 78.9% vs. *TT* 44.4%). Patients with the *AA* genotype for rs5030755 were further associated with a worse prognosis level in combination with the wild type allele C in *NEIL3* (*GG* 46.7% vs. *CC+CG* 32.3%). The structure of the tree and corresponding survival curves from terminal nodes are presented in Figure 3.

Austrian cohort. The final tree structure contained four terminal nodes determined by five variables (age, TNM stage, and three SNPs—rs3087386, rs3204953, and rs4796033). In stage I, a better EFS prognosis was shown within patients under 60 years of age. In stage II, the subsequent split was age, which was seen to interact with the *CC* genotype in *RAD51D* rs4796033 and *CT+TT* genotypes in *REV1* rs3087386 (*CT+TT* 98.7% vs. *CC* 84.9%). Furthermore, *GA+AA* genotypes in *REV3L* rs3204953 showed a better EFS prognosis in stage II patients over the age of 70 (*GA+AA* 88.9% vs. *GG* 60.9%). In stage III, the subsequent split showed an interaction with age only. The structure of the tree and corresponding survival curves from terminal nodes are presented in Figure 4.

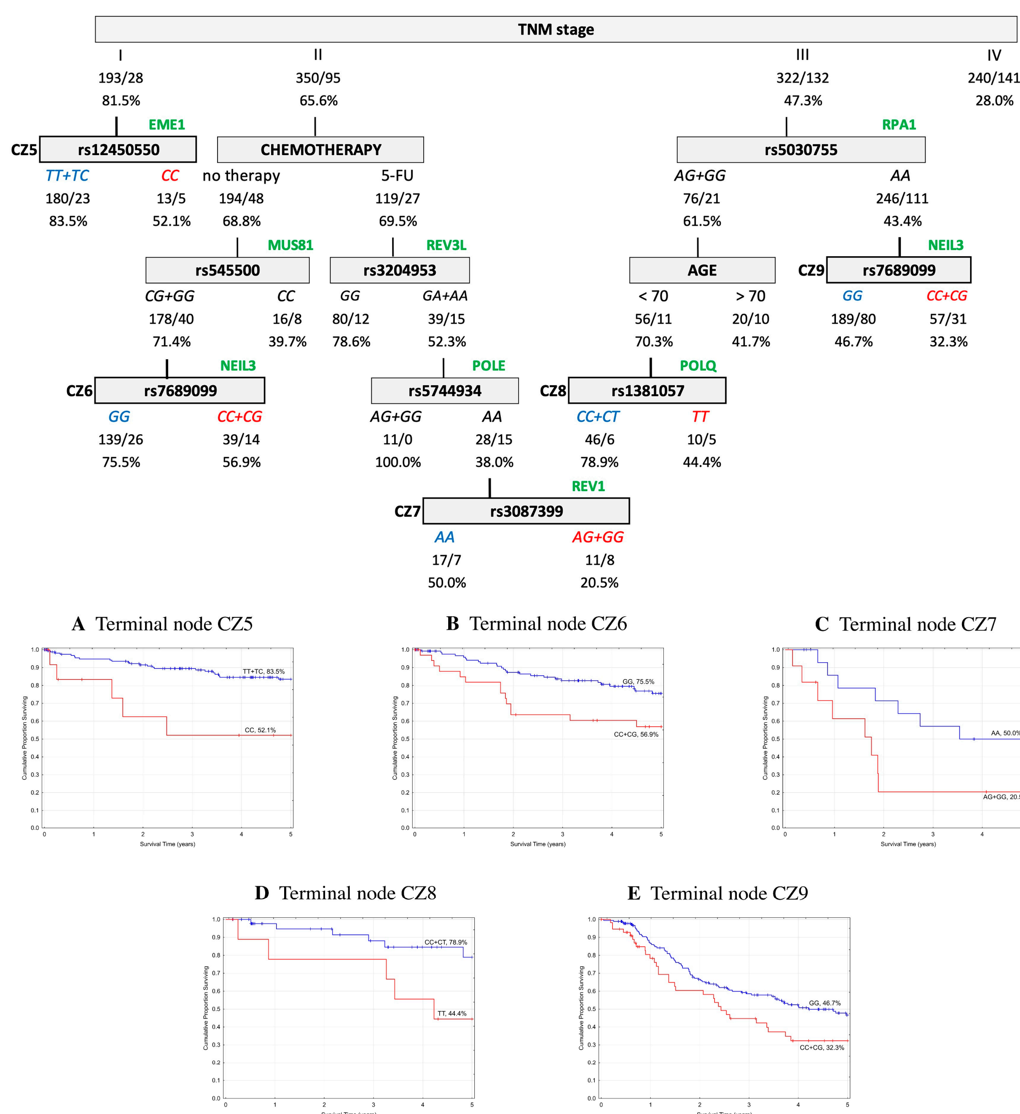


Figure 3. Event free survival (EFS) classification and regression tree of colorectal cancer patients from the Czech Republic. Numbers under each node indicate the total number of cases in the subcategory/number of events, and the percentage of patients with five-year EFS. Terminal nodes are bordered in bold, and the corresponding Kaplan–Meier curves are shown underneath. (**A**) Terminal node CZ5; (**B**) Terminal node CZ6; (**C**) Terminal node CZ7; (**D**) Terminal node CZ8; (**E**) Terminal node CZ9.

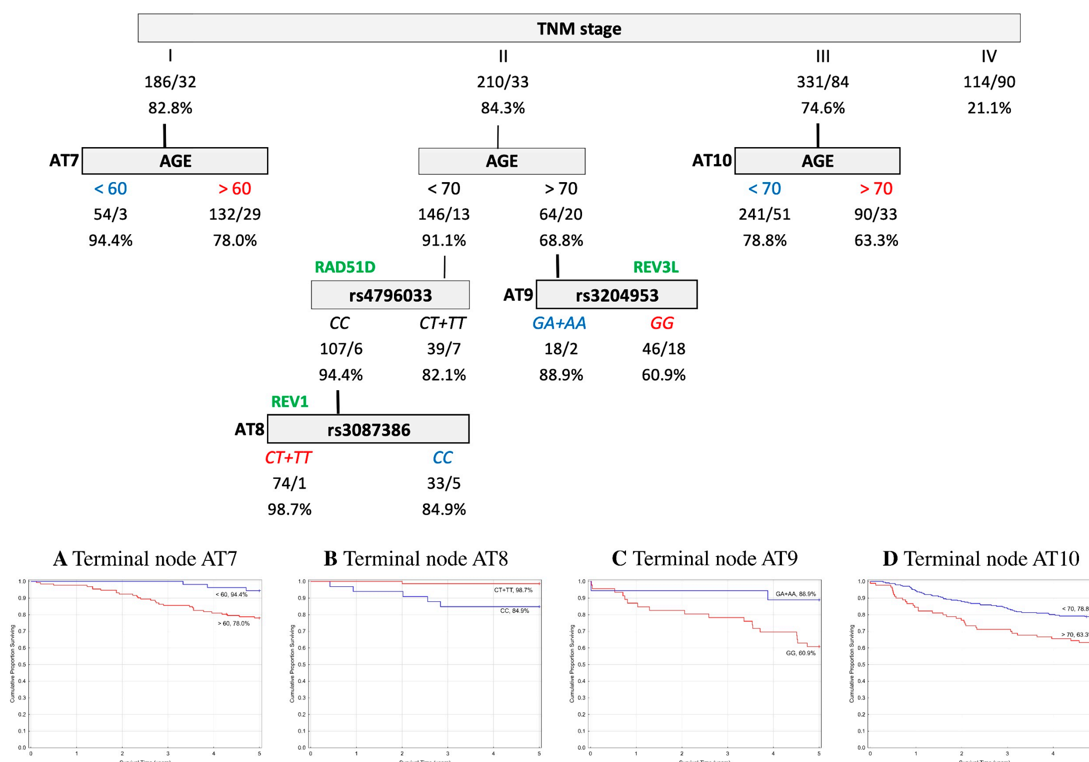


Figure 4. Event free survival (EFS) classification and regression tree of colorectal cancer patients from Austria. Numbers under each node indicate the total number of cases in the subcategory/number of events and the percentage of patients with five-year EFS. Terminal nodes are bordered in bold, and the corresponding Kaplan–Meier curves are shown underneath. (**A**) Terminal node AT7; (**B**) Terminal node AT8; (**C**) Terminal node AT9; (**D**) Terminal node AT10.

3. Discussion

DNA repair has an essential role in maintaining genome integrity and preventing carcinogenesis. Amino acid alterations by nsSNPs in DNA repair genes can cause changes to the function or level of the coded proteins, resulting in abrogated DNA repair, which in combination with continuous endogenous DNA damage over time could lead to genomic damage and carcinogenesis [7,18]. In the present study, we sought to identify associations between 16 potentially functional genetic polymorphisms in 12 DNA repair genes with CRC risk, patients' survival, and response to chemotherapy in Czech and Austrian cohorts. To our knowledge, no similar studies have previously examined these selected SNPs in relation to CRC susceptibility and clinical outcomes after diagnosis.

In the discovery set from the Czech Republic, the results showed an association between the variant *AA* genotype in *REV3L* rs3204953 and an increased risk of CRC. *REV3L* encodes a catalytic subunit of an error-prone DNA polymerase ζ, whose involvement in both double strand break (DSB) repair and translesion synthesis (TLS) pathways may explain why it is the only known specialized DNA polymerase reducing spontaneous tumor development [19,20]. DSBs, i.e. breaks in both DNA strands, are one of the most cytotoxic lesions for genetic integrity, and if not adequately repaired, DSB can result in mutagenic events or cell death [21]. TLS is a DNA damage tolerance process that allows cells to continue replication past DNA templates containing bulky lesions without resulting in stalled replication forks and therefore preventing DNA strand breaks.

Disrupted REV3L in cancer cell lines showed the importance of accurately regulated *REV3L* expression, when its inhibition induced DNA damage and growth arrest in cancer cells, whereas overexpression led to increased spontaneous mutation rates [22]. Expression levels of this polymerase have also been linked to sensitivity to chemotherapy. While defects in the protein resulted in an increased sensitivity to therapy in multiple tumor cell lines, its overexpression induced increased

therapy resistance [23–25]. Furthermore, a decreased expression of *REV3L* has also been reported in tumor compared with the adjacent non-malignant tissue in colon cancer [26,27].

An association of rs3204953 was observed with a higher risk of breast cancer in a Swedish cohort, however, the results were not replicated in a Polish cohort [28]. Other genetic variants in *REV3L* have been found to be associated with both disease development risk and patients' survival for different tumor types, such as breast cancer, stomach cancer, and CRC [28–30]. None of the other associated SNPs were found in linkage disequilibrium with rs3204953 examined here.

In addition to the in silico predictions of the F-SNP database of the deleterious nature of the rs3204953 SNP for *REV3L* protein function, we also used web-servers ELASPIC and DUET to assess the energetic impact of the amino acid change. In ELASPIC, the valine to isoleucine substitution was predicted to decrease the protein stability, resulting in a protein favoring an unfolded state (as the Gibbs free energy of folding for the domain affected by the SNP is changed by $\Delta\Delta G = 1.97$).

Regarding the clinical outcome, results of the Czech five-year EFS CART analysis showed that rs3204953 in *REV3L* was chosen as the optimal split for the CRC stage II patients receiving 5-FU-based chemotherapy. This finding indicates its possible use in personalized treatment strategies by identifying CRC stage II patients who are likely to benefit from adjuvant therapy.

Despite the promising results in the Czech population, an association of *REV3L* SNP with CRC risk could not be confirmed in the Austrian replication set. However, *REV3L* emerged several times as the optimal split in the Austrian CART analyses as well. Thus, according to all of the available data, we suggest that the *REV3L* gene may impact CRC susceptibility, survival, and therapy outcomes and warrants further investigation.

In survival CART analyses, TNM stage and age were shown as the most significant prognostic factors in both of the study cohorts. Apart from these clinico-pathological factors, we observed significant associations of several nsSNPs with patients' survival and clinical outcomes. However, a few of these were shown as significant more than once in the CART analysis, suggesting their potentially greater relevance on patients' survival. For example, *POLQ* gene polymorphisms appeared four times as the optimal split factor in the Czech CART analyses (rs1381057, rs3218649 twice, and rs3218651) and four times in the Austrian CART analyses (rs1381057 twice and rs3218651 twice). Polymerase θ encoded by *POLQ* is an error-prone polymerase with a similar role to polymerase ζ, and is involved in the base excision repair (BER) and DSB repair [31]. In addition to DNA repair, this polymerase also plays a crucial role in TLS [32].

The expression of polymerase θ is tightly regulated. A complementary body of literature reported an upregulation of *POLQ* in different tumor tissues (breast cancer, non-small cell lung cancer, oral squamous cell carcinoma, stomach cancer, and CRC), and this overexpression acted as a strong prognostic factor [33–36].

Strikingly, at least nine polymorphisms out of 23 known SNPs in the human *POLQ* gene are predicted to alter protein function [32]. Several *POLQ* SNPs have also been associated with a risk of different tumors, such as breast cancer, esophageal cancer, and Non-Hodgkin's Lymphoma [28,37–40]. While only some of the breast cancer studies included rs3218649, no significant association was detected [28,38,40] and none of the other associated SNPs were found in the linkage disequilibrium with our selected SNPs.

The abovementioned studies highlighted the significance of adequate *POLQ* functioning and regulation for tumor suppression. Furthermore, the protein stability prediction for rs1381057 by ELASPIC estimated a change of the Gibbs free energy to $\Delta\Delta G = 1.65$, suggesting that the substitution of glutamine to arginine decreases the altered protein stability. Unfortunately, we could not perform a protein stability prediction of rs3218649 and rs3218651 by ELASPIC, as these SNPs do not fall within the domain boundaries required by the software.

Another SNP, rs7689099 in *NEIL3* gene, emerged twice in the Czech five-year EFS CART analysis as the optimal split factor after the TNM stratification, suggesting its significance in patients´ survival. The *NEIL3* encodes a DNA glycosylase, playing an important role in the first step of the BER

pathway [41]. The process of eliminating damaged nucleotides by BER is crucial to evade mutations at these sites, which is likely to aid tumor suppression [42].

The upregulation of *NEIL3* appears to be involved in the maintenance of cancer cell growth or the progression of malignancy. Significantly elevated expression levels in tumors, compared to corresponding non-malignant tissues, were reported in 20 cancer sites, including CRC [43,44]. The overexpression was further observed in association with the progression of primary melanoma to distant metastasis [45].

Sequence variability in different DNA glycosylases have been proposed as susceptibility factors for different malignancies [46]. Specifically, *NEIL3* SNPs were associated with the risk of glioma, prostate, and thyroid cancer [47–49], with rs7689099 being associated with a reduced risk of differentiated thyroid carcinoma and prostate cancer [47,49]. None of the other associated SNPs were found in linkage disequilibrium with rs7689099.

As rs7689099 in *NEIL3* gene does not fall within the domain boundaries, we could not use ELASPIC protein stability prediction. Again, the association of *NEIL3* SNP with the survival of CRC patients was not replicated in the Austrian sample set. However, considering the available data, we suggest that the variation of the *NEIL3* gene also has relevance for CRC susceptibility, survival, and therapy outcome.

In agreement with the in silico predictions about the functionality of the SNPs, we observed several significant associations of different genetic variants with survival and clinical outcomes of CRC patients both from the Czech Republic and Austria. However, we were not able to confirm the particular associations of individual SNPs between the discovery and the replication set. One might argue that the failure to replicate the association results might be due to differential gene–environmental interactions, and the differences in the clinical composition between the case-control populations of the discovery and the replication set (Table 2). Furthermore, it is also possible that other factors might have biased the results, for example earlier CRC detection in Austrian patients thanks to a better general awareness of the disease and a high standard medical care. This assumption is supported by the results in five-year OS CART analyses for stage III patients, where we observed a substantial difference between survival in the Czech and Austrian patients (65.6% vs. 82.8%). Our conclusion was based on the fact that CRC stage III is further divided into three more separate categories (IIIA, IIIB, and IIIC) according to the extent to which cancer has spread (i.e., number of lymph nodes affected). The survival rate then significantly decreases with the disease advancement. For example, in colon cancer patients, the survival rate for stage IIIA is about 90%, for stage IIIB 72%, and for stage IIIC only about 53% [50]. The strengths of the present work include the recruitment of a considerable number of cases and controls at the same centers, homogeneous for their ancestry (all Caucasian from the Czech Republic and Austria), and clinically well-defined (follow-up data collected by the same physicians), thus minimizing any possible population bias.

In conclusion, this is the first study to evaluate the association of genetic variants in DNA repair genes, selected by likely functional relevance with CRC. We identified several nsSNPs potentially affecting either CRC susceptibility or patients' survival. Our data provide observational evidence of the potential role of nsSNPs in CRC pathogenesis, and suggest that even subtle alterations in the specific proteins that function in DNA repair pathways may lead to inaccurate DNA repair, and thus contribute to carcinogenesis.

Due to the lack of replication of significant associations, further studies on independent populations are warranted. This is underlined by the involvement of the same DNA repair genes in both Czech and Austrian CRC populations. Moreover, it is important to functionally characterize these candidate genetic variants, and to find biological mechanisms underlying the associations in order to assess these nsSNPs as prognostic and/or predictive biomarkers in CRC. Potential clinical uses are to help define individual CRC risk and tailor disease management based on the unique molecular profile of each patient.

4. Material and Methods

4.1. SNP Selection and In Silico Analysis of Functional Relevance and Conservation

From the complete list of DNA repair genes available online (http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html, March 2014 version), all of the genes involved in repairing DNA damage caused by 5-FU or oxaliplatin were retrieved, as these are common chemotherapeutic treatment regimens for CRC.

In total, 106 genes of BER, nucleotide excision repair (NER), and DSB (including interstrand cross-links repair (ICL), fanconi anemia (FA), and TLS pathways) were searched for nsSNPs in the freely available F-SNP database [51]. The database also provides integrated information about possible effects of the base change on the coded protein, and thus helps to identify nsSNPs with a potential pathological effect on human health. The F-SNP data are obtained from several genomic databases, like SIFT, PolyPhen2, SNPeffect, and SNPs3D. The variants predicted as deleterious or damaging were further studied.

Selected relevant nsSNPs were then filtered for a MAF >10% in European populations to provide sufficient study power with the size of our case-control study, in order to uncover moderate genetic effects. The information was primarily derived from the Ensembl 2015 database—1000 Genomes Project Phase 3, EUR population (https://www.ensembl.org/index.html). Whenever this was not possible, other reference populations were considered (i.e., HAPMAP CEU population).

The SNPs with the required MAF were tested for linkage disequilibrium (LD) with the data from HapMap (v. 3, release R2 in the CEU population, ftp://ftp.ncbi.nlm.nih.gov/hapmap/). The 38 identified nsSNPs were further searched within the Genetic Association Database (http://geneticassociationdb.nih.gov, accessed on 9 January 2014). From these, sixteen nsSNPs were already investigated elsewhere in relation to CRC, and therefore were excluded from this study.

The 22 remaining nsSNPs were tested by comparative genomics to evaluate the probability that the nucleotide is located in an evolutionary conserved position or within a constrained element, using the Genomic Evolutionary Rate Profiling GERP++ RS (Rejected Substitutions) score. An element with a GERP++ RS score >800 defines ultra-conserved regions among mammals. SiPhy evaluates the conservation of the motif around the SNPs.

After this selection, sixteen nsSNPs in 12 DNA repair genes complied with the required selection criteria. The workflow for the selection is depicted in Figure 5.

To evaluate the stability of the final protein affected by nsSNP, we further utilized web-server tools ELASPIC and DUET to assess the energetic impact of the amino acid change (http://elaspic.kimlab.org/ and http://biosig.unimelb.edu.au/duet/stability). The main output is the predicted variation in the Gibbs free energy ($\Delta\Delta G$) of folding and/or binding for every domain affected by the SNP.

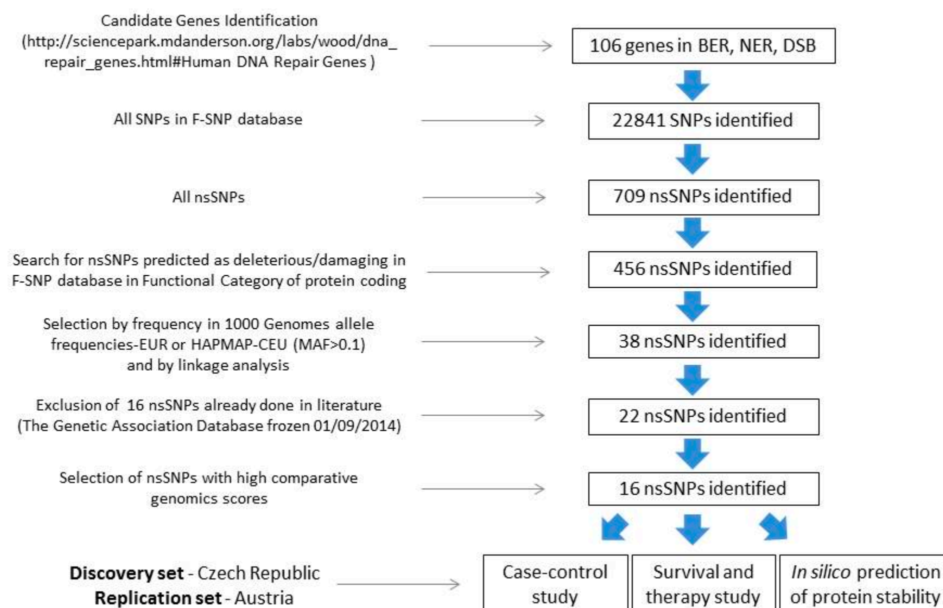


Figure 5. Workflow strategy for the selection and analyses of functional non-synonymous polymorphisms in DNA repair genes. BER—base excision repair; NER—nucleotide excision repair; DSB—double strand break repair; SNP—single nucleotide polymorphism; nsSNP—non-synonymous SNP; MAF—minor allele frequency.

4.2. Study Populations and Data Collection

Patients included in the study were newly diagnosed histologically confirmed individuals with sporadic CRC. The exclusion criteria were as follows: (1) hereditary CRC forms (Lynch syndrome and familial adenomatous polyposis) and (2) a personal history of previous malignant disease. Personal data, such as date of birth, sex, lifestyle habits, body mass index (BMI), diabetes mellitus, and family/personal history of cancer, were obtained using a structured questionnaire in order to determine potential risk factors for CRC. For all subjects, clinical data including tumor-related parameters, such as the location of the tumor, International Union Against Cancer (UICC) TNM stage system, degree of tumor differentiation, and adjuvant chemotherapy treatment details, were collected, along with information about distant metastasis, relapse, and date of death.

Patients were divided into three subgroups according to the therapy received. The first group of patients did not receive any adjuvant chemotherapy after surgery. The second group of patients received a 5-FU-based adjuvant regimen as a postoperative therapy (based either on a Mayo, a simplified DeGramont, or a Xeloda regimen). The third group of subjects received adjuvant 5-FU treatment combined with oxaliplatin (based either on a FOLFOX or a XELOX regimen).

The study was approved by the local ethics committee of each participating hospital, and written informed consent to participate in the study and to approve the use of their biological samples for genetic analyses was obtained from all patients, according to the 1964 Helsinki declaration.

4.2.1. Discovery Set—Czech Republic

Patients (n = 1832) were recruited at several oncological and gastroenterological departments of different hospitals all over the Czech Republic from September 2003 to January 2014. The last update of the patients' follow-up for this study was in December 2015. Characteristics of the study participants are shown in Table 2 (partially described in the literature [52,53].

The control group consisted of 659 healthy blood donors and 513 colonoscopy-negative controls, which were collected during the same time period as the cases. Healthy blood donor volunteers were recruited at the Faculty Hospital Kralovske Vinohrady in Prague and the Vojkov hospital. The group of colonoscopy-negative controls consisted of subjects admitted to the hospital gastroenterology

departments who had negative colonoscopy results for malignancy or idiopathic bowel diseases. The reasons for undergoing the colonoscopy were as follows: (i) positive fecal occult blood test, (ii) hemorrhoids, (iii) abdominal pain of unknown origin, and (iv) macroscopic bleeding. All individuals were subjected to standard examinations so as to verify the health status for blood donation, and were cancer-free at the time of the sampling.

DNA was extracted from the peripheral blood lymphocytes using standard procedures. When blood was not available (for 690 cases), healthy colon/rectal tissue was used to obtain DNA by using the DNeasy Blood and Tissue Kit (Qiagen, Courtaboeuf, France). Genotyping was performed at LGC Genomics (Hoddesdon, Herts, UK), using the KASP™, a competitive allele-specific PCR genotyping system. For quality control purposes, duplicate samples (5% of the total numbers of samples) were repeated for each SNP. Two no-template controls were included in each plate. The genotype correlation between the duplicate samples was >98%. Two CRC cases were eliminated due to low genotyping rates.

4.2.2. Replication Set—Austria

In the ongoing Colorectal Cancer Study of Austria (CORSA), over 13,000 participants comprising CRC cases (stages I–IV); adenomas; and population-based, colonoscopy-negative controls have been recruited since 2003, in cooperation with the province-wide screening program "Burgenland Prevention Trial of Colorectal Disease with Immunological Testing" (B-PREDICT). All inhabitants of the Austrian province Burgenland aged between 40 and 80 years are invited annually to participate in fecal immunochemical testing (FIT). FIT-positive tested individuals are offered a complete colonoscopy and are asked to participate in CORSA at the time of colonoscopy. Only the individuals with histologically confirmed sporadic CRC were included in this study.

Further CRC cases were recruited at multiple centers in Vienna, including the Medical University of Vienna (Department of Surgery), the Sozialmedizinisches Zentrum Süd, the Hospital Rudolfstiftung, and the Medical University of Graz (Department of Internal Medicine). The replication set comprised 950 CRC patients and 820 colonoscopy-negative controls from CORSA. The last update of the patients' follow-up was performed in August 2018. Baseline characteristics of this cohort are presented in Table 2 , and the study has previously been described [54].

The genomic DNA isolation from peripheral blood was performed using the QIAamp DNA Blood Midi Kit, according to the manufacturer's recommendations (Qiagen, Valencia, CA, USA), and was stored at −80 °C. Genotyping was performed using the population-optimized Axiom Genome-Wide CEU 1 Array (Affymetrix, Santa Clara, CA, USA). The arrays were processed at the Institute of Human Genetics, Helmholtz Center Munich, Germany, and genotype assignment was performed as described in Hofer et al. [54]. Data for two SNPs (*FAAP24* rs3816032 and *MUS81* rs545500) were not covered on the array, and therefore could not be included in further analyses.

4.3. Statistical Analysis

In controls, the genotype frequencies for each polymorphism were tested for deviation from the Hardy–Weinberg equilibrium, using a Pearson χ^2-test (1 degree of freedom) with a type-I error threshold set at $\alpha = 0.05$.

The association between nsSNPs and CRC risk was determined by logistic regression, and was calculated by estimating the ORs, and their 95% CIs were adjusted for age. The ancestral allele (evolutionary primal) was used as a reference. For all nsSNPs, co-dominant, dominant, and recessive models were calculated.

In this study, the outcome variables measured were OS and EFS. OS was defined as the time from the surgery to the date of death, or the date of last follow up (for the Czech cohort it was December 2015, for the Austrian cohort it was August 2018). EFS was defined as the time from surgery to the occurrence of distant metastasis, local recurrence, or death, whichever came first. The survival curves for OS and EFS were derived by the Kaplan–Meier log-rank test. The relative risk of death and recurrence was estimated as a hazard ratio (HR) with 95% CIs, using Cox regression (no covariates adjustment was applied).

A multivariate analysis, referred to as a CART [55], was used to assess the prognostic value of interactions between the standard clinico-pathological variables and the genetic variants in relation to their impact on five-year survival in CRC patients. The analysis constructs a set of decision rules that stratify the homogenous risk groups of the responsive variable. Splits for each variable were examined, and the variable (predictor) that provides the best or "optimal" split was selected. Each subgroup was further divided in the same manner. In the Czech sample set, CART was implemented using nine common clinical and pathological variables, including age, sex, smoking habit (non-smokers vs. smokers vs. ex-smokers), diabetes mellitus, positive family history of CRC, diagnosis (colon vs. rectal cancer), TNM stage, grade, and therapy (no therapy vs. 5-FU-based without oxaliplatin vs. 5-FU in combination oxaliplatin), and all examined nsSNPs. In the Austrian sample set, because the information for five of the variables (smoking habit, positive family history of CRC, grade, and nsSNPs rs3816032 and rs545500) were only available for a small number of patients, only six common clinico-pathological variables and 14 nsSNPs were implemented for the CART analysis of this cohort.

Statistical analyses were performed using SAS software (SAS Institute, Cary, NC, USA). Graphs were performed using SW STATISTICA (StatSoft, Inc., Tulsa, OK, USA). Multiple testing corrections were performed using the Benjamini–Hochberg FDR [56].

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/1/97/s1.

Author Contributions: K.J., V.V. (V. Vymetalkova), and P.V. conceived the study. V.V. (V. Veskrnova), T.B., M.S., M.L., and V.L. performed hospital-based sample and data collection for the discovery set. K.J. performed data curation for the discovery set. S.B. and T.G. performed data curation for the replication set. K.J. and C.D.G. performed a computational prediction of the functional SNPs. K.J. and S.V. performed data analyses and interpreted the results with assistance from V.V. (V. Vymetalkova), A.N., and B.P. K.J. wrote the manuscript. P.V., D.J.H., and A.G. revised the manuscript. All of the authors read and approved the final manuscript for publication.

Funding: This study was supported by the Grant Agency of Charles University (GAUK112515) (KJ), the Health Research Board of Ireland project grant HRA_PHS/2015/1142 (DJH), the Centrum of Clinical and Experimental liver surgery (UNCE/MED/006), the National Sustainability Program I (NPU I) Nr. LO1503 provided by the Ministry of Education Youth and Sports of the Czech Republic (VL), the Grant Agency of the Czech Republic (GACR 17-16857S) (AN), the Fondazione Umberto Veronesi "Post-doctoral fellowship Year 2014–2017" (BP), the Internal Grant Agency of the Ministry of Health of the Czech Republic (AZV MZ 15-26535A and AZV 17-30920A) (VV), the FFG BRIDGE (grant 829675), and the Herzfelder´sche Familienstiftung (AG) and the Grant Agency of the Czech Republic (GACR 15-14789S) (PV).

Acknowledgments: This article is based upon work from COST Action CA17118, supported by COST (European Cooperation in Science and Technology, www.cost.eu).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

5-FU	5-Fluorouracil
BER	Base excision repair
BMI	Body mass index
CART	Classification and regression tree analysis
CI	Confidence intervals
CRC	Colorectal cancer
DSB	Double strand break repair
EFS	Event-free survival
FA	Fanconi anemia
FDR	False discovery rate
GERP	Genomic evolutionary rate profiling
GWAS	Genome-wide association study
HRs	Hazard ratios
ICL	Interstrand cross-links repair
LD	Linkage disequilibrium
MAF	Minor allele frequency

NER	Nucleotide excision repair
nsSNP	Non-synonymous single nucleotide polymorphism
ORs	Odds ratios
OS	Overall survival
RS	Rejected substitutions
TLS	Translesion synthesis
TNM	Tumor–node–metastasis stage system
UICC	International union against cancer

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**. [CrossRef] [PubMed]
2. Ferlay, J.; Steliarova-Foucher, E.; Lortet-Tieulent, J.; Rosso, S.; Coebergh, J.W.; Comber, H.; Forman, D.; Bray, F. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *Eur. J. Cancer* **2013**, *49*, 1374–1403. [CrossRef] [PubMed]
3. Aran, V.; Victorino, A.P.; Thuler, L.C.; Ferreira, C.G. Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality. *Clin. Colorectal Cancer* **2016**, *15*, 195–203. [CrossRef] [PubMed]
4. Binefa, G.; Rodriguez-Moranta, F.; Teule, A.; Medina-Hayas, M. Colorectal cancer: From prevention to personalized medicine. *World J. Gastroenterol.* **2014**, *20*, 6786–6808. [CrossRef] [PubMed]
5. Gustavsson, B.; Carlsson, G.; Machover, D.; Petrelli, N.; Roth, A.; Schmoll, H.J.; Tveit, K.M.; Gibson, F. A review of the evolution of systemic chemotherapy in the management of colorectal cancer. *Clin. Colorectal Cancer* **2015**, *14*, 1–10. [CrossRef] [PubMed]
6. Brenner, H.; Kloor, M.; Pox, C.P. Colorectal cancer. *Lancet* **2014**, *383*, 1490–1502. [CrossRef]
7. Vodicka, P.; Stetina, R.; Polakova, V.; Tulupova, E.; Naccarati, A.; Vodickova, L.; Kumar, R.; Hanova, M.; Pardini, B.; Slyskova, J.; et al. Association of DNA repair polymorphisms with DNA repair functional outcomes in healthy human subjects. *Carcinogenesis* **2007**, *28*, 657–664. [CrossRef]
8. He, J.; Shi, T.Y.; Zhu, M.L.; Wang, M.Y.; Li, Q.X.; Wei, Q.Y. Associations of Lys939Gln and Ala499Val polymorphisms of the XPC gene with cancer susceptibility: A meta-analysis. *Int. J. Cancer* **2013**, *133*, 1765–1775. [CrossRef] [PubMed]
9. Slyskova, J.; Naccarati, A.; Pardini, B.; Polakova, V.; Vodickova, L.; Smerhovsky, Z.; Levy, M.; Lipska, L.; Liska, V.; Vodicka, P. Differences in nucleotide excision repair capacity between newly diagnosed colorectal cancer patients and healthy controls. *Mutagenesis* **2012**, *27*, 225–232. [CrossRef] [PubMed]
10. Peters, U.; Jiao, S.; Schumacher, F.R.; Hutter, C.M.; Aragaki, A.K.; Baron, J.A.; Berndt, S.I.; Bezieau, S.; Brenner, H.; Butterbach, K.; et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology* **2013**, *144*, 799–807.e24. [CrossRef]
11. Whiffin, N.; Hosking, F.J.; Farrington, S.M.; Palles, C.; Dobbins, S.E.; Zgaga, L.; Lloyd, A.; Kinnersley, B.; Gorman, M.; Tenesa, A.; et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum. Mol. Genet.* **2014**, *23*, 4729–4737. [CrossRef] [PubMed]
12. Shen, C.; Yan, T.; Wang, Z.; Su, H.C.; Zhu, X.; Tian, X.; Fang, J.Y.; Chen, H.; Hong, J. Variant of SNP rs1317082 at CCSlnc362 (RP11-362K14.5) creates a binding site for miR-4658 and diminishes the susceptibility to CRC. *Cell Death Dis.* **2018**, *9*, 1177. [CrossRef] [PubMed]
13. Tanskanen, T.; van den Berg, L.; Valimaki, N.; Aavikko, M.; Ness-Jensen, E.; Hveem, K.; Wettergren, Y.; Bexe Lindskog, E.; Tonisson, N.; Metspalu, A.; et al. Genome-wide association study and meta-analysis in Northern European populations replicate multiple colorectal cancer risk loci. *Int. J. Cancer* **2018**, *142*, 540–546. [CrossRef] [PubMed]
14. Huyghe, J.R.; Bien, S.A.; Harrison, T.A.; Kang, H.M.; Chen, S.; Schmit, S.L.; Conti, D.V.; Qu, C.; Jeon, J.; Edlund, C.K.; et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat. Genet.* **2018**. [CrossRef] [PubMed]
15. Slyskova, J.; Cordero, F.; Pardini, B.; Korenkova, V.; Vymetalkova, V.; Bielik, L.; Vodickova, L.; Pitule, P.; Liska, V.; Matejka, V.M.; et al. Post-treatment recovery of suboptimal DNA repair capacity and gene expression levels in colorectal cancer patients. *Mol. Carcinog.* **2015**, *54*, 769–778. [CrossRef] [PubMed]

16. Wyatt, M.D.; Wilson, D.M., 3rd. Participation of DNA repair in the response to 5-fluorouracil. *Cell. Mol. Life Sci.* **2009**, *66*, 788–799. [CrossRef]
17. Martin, L.P.; Hamilton, T.C.; Schilder, R.J. Platinum resistance: The role of DNA repair pathways. *Clin. Cancer Res.* **2008**, *14*, 1291–1295. [CrossRef]
18. De Mattia, E.; Cecchin, E.; Toffoli, G. Pharmacogenomics of intrinsic and acquired pharmacoresistance in colorectal cancer: Toward targeted personalized therapy. *Drug Resist. Updat.* **2015**, *20*, 39–70. [CrossRef]
19. Vaisman, A.; Woodgate, R. Translesion DNA polymerases in eukaryotes: What makes them tick? *Crit. Rev. Biochem. Mol. Biol.* **2017**, *52*, 274–303. [CrossRef]
20. Wittschieben, J.P.; Patil, V.; Glushets, V.; Robinson, L.J.; Kusewitt, D.F.; Wood, R.D. Loss of DNA polymerase zeta enhances spontaneous tumorigenesis. *Cancer Res.* **2010**, *70*, 2770–2778. [CrossRef]
21. Aparicio, T.; Baer, R.; Gautier, J. DNA double-strand break repair pathway choice and cancer. *DNA Repair (Amst.)* **2014**, *19*, 169–175. [CrossRef]
22. Knobel, P.A.; Marti, T.M. Translesion DNA synthesis in the context of cancer research. *Cancer Cell Int.* **2011**, *11*, 39. [CrossRef] [PubMed]
23. Galluzzi, L.; Senovilla, L.; Vitale, I.; Michels, J.; Martins, I.; Kepp, O.; Castedo, M.; Kroemer, G. Molecular mechanisms of cisplatin resistance. *Oncogene* **2012**, *31*, 1869–1883. [CrossRef] [PubMed]
24. Wang, H.; Zhang, S.Y.; Wang, S.; Lu, J.; Wu, W.; Weng, L.; Chen, D.; Zhang, Y.; Lu, Z.; Yang, J.; et al. REV3L confers chemoresistance to cisplatin in human gliomas: The potential of its RNAi for synergistic therapy. *Neuro Oncol.* **2009**, *11*, 790–802. [CrossRef]
25. Roos, W.P.; Tsaalbi-Shtylik, A.; Tsaryk, R.; Guvercin, F.; de Wind, N.; Kaina, B. The translesion polymerase Rev3L in the tolerance of alkylating anticancer drugs. *Mol. Pharmacol.* **2009**, *76*, 927–934. [CrossRef] [PubMed]
26. Stallons, L.J.; McGregor, W.G. Translesion synthesis polymerases in the prevention and promotion of carcinogenesis. *J. Nucleic Acids* **2010**, *2010*, 643857. [CrossRef]
27. Brondello, J.M.; Pillaire, M.J.; Rodriguez, C.; Gourraud, P.A.; Selves, J.; Cazaux, C.; Piette, J. Novel evidences for a tumor suppressor role of Rev3, the catalytic subunit of Pol zeta. *Oncogene* **2008**, *27*, 6093–6101. [CrossRef] [PubMed]
28. Varadi, V.; Bevier, M.; Grzybowska, E.; Johansson, R.; Enquist, K.; Henriksson, R.; Butkiewicz, D.; Pamula-Pilat, J.; Tecza, K.; Hemminki, K.; et al. Genetic variation in genes encoding for polymerase zeta subunits associates with breast cancer risk, tumour characteristics and survival. *Breast Cancer Res. Treat.* **2011**, *129*, 235–245. [CrossRef]
29. Pan, J.; Chi, P.; Lu, X.; Xu, Z. Genetic polymorphisms in translesion synthesis genes are associated with colorectal cancer risk and metastasis in Han Chinese. *Gene* **2012**, *504*, 151–155. [CrossRef] [PubMed]
30. Hussain, S.K.; Mu, L.N.; Cai, L.; Chang, S.C.; Park, S.L.; Oh, S.S.; Wang, Y.; Goldstein, B.Y.; Ding, B.G.; Jiang, Q.; et al. Genetic variation in immune regulation and DNA repair pathways and stomach cancer in China. *Cancer Epidemiol. Biomark. Prev.* **2009**, *18*, 2304–2309. [CrossRef] [PubMed]
31. Yousefzadeh, M.J.; Wood, R.D. DNA polymerase POLQ and cellular defense against DNA damage. *DNA Repair (Amst.)* **2013**, *12*, 1–9. [CrossRef] [PubMed]
32. Beagan, K.; McVey, M. Linking DNA polymerase theta structure and function in health and disease. *Cell. Mol. Life Sci.* **2016**, *73*, 603–615. [CrossRef] [PubMed]
33. Kawamura, K.; Bahar, R.; Seimiya, M.; Chiyo, M.; Wada, A.; Okada, S.; Hatano, M.; Tokuhisa, T.; Kimura, H.; Watanabe, S.; et al. DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. *Int. J. Cancer* **2004**, *109*, 9–16. [CrossRef] [PubMed]
34. Lemee, F.; Bergoglio, V.; Fernandez-Vidal, A.; Machado-Silva, A.; Pillaire, M.J.; Bieth, A.; Gentil, C.; Baker, L.; Martin, A.L.; Leduc, C.; et al. DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13390–13395. [CrossRef] [PubMed]
35. Allera-Moreau, C.; Rouquette, I.; Lepage, B.; Oumouhou, N.; Walschaerts, M.; Leconte, E.; Schilling, V.; Gordien, K.; Brouchet, L.; Delisle, M.B.; et al. DNA replication stress response involving PLK1, CDC6, POLQ, RAD51 and CLASPIN upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients. *Oncogenesis* **2012**, *1*, e30. [CrossRef] [PubMed]
36. Pillaire, M.J.; Selves, J.; Gordien, K.; Gourraud, P.A.; Gentil, C.; Danjoux, M.; Do, C.; Negre, V.; Bieth, A.; Guimbaud, R.; et al. A 'DNA replication' signature of progression and negative outcome in colorectal cancer. *Oncogene* **2010**, *29*, 876–887. [CrossRef] [PubMed]

37. Li, W.Q.; Hu, N.; Hyland, P.L.; Gao, Y.; Wang, Z.M.; Yu, K.; Su, H.; Wang, C.Y.; Wang, L.M.; Chanock, S.J.; et al. Genetic variants in DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population. *Carcinogenesis* **2013**, *34*, 1536–1542. [CrossRef]
38. Brandalize, A.P.; Schuler-Faccini, L.; Hoffmann, J.S.; Caleffi, M.; Cazaux, C.; Ashton-Prolla, P. A DNA repair variant in POLQ (c.-1060A > G) is associated to hereditary breast cancer patients: A case-control study. *BMC Cancer* **2014**, *14*, 850. [CrossRef]
39. Rendleman, J.; Antipin, Y.; Reva, B.; Adaniel, C.; Przybylo, J.A.; Dutra-Clarke, A.; Hansen, N.; Heguy, A.; Huberman, K.; Borsu, L.; et al. Genetic variation in DNA repair pathways and risk of non-Hodgkin's lymphoma. *PLoS ONE* **2014**, *9*, e101685. [CrossRef]
40. Family, L.; Bensen, J.T.; Troester, M.A.; Wu, M.C.; Anders, C.K.; Olshan, A.F. Single-nucleotide polymorphisms in DNA bypass polymerase genes and association with breast cancer and breast cancer subtypes among African Americans and Whites. *Breast Cancer Res. Treat.* **2015**, *149*, 181–190. [CrossRef]
41. Krokeide, S.Z.; Laerdahl, J.K.; Salah, M.; Luna, L.; Cederkvist, F.H.; Fleming, A.M.; Burrows, C.J.; Dalhus, B.; Bjoras, M. Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin. *DNA Repair (Amst.)* **2013**, *12*, 1159–1164. [CrossRef] [PubMed]
42. Wallace, S.S.; Murphy, D.L.; Sweasy, J.B. Base excision repair and cancer. *Cancer Lett.* **2012**, *327*, 73–89. [CrossRef] [PubMed]
43. Hildrestrand, G.A.; Neurauter, C.G.; Diep, D.B.; Castellanos, C.G.; Krauss, S.; Bjoras, M.; Luna, L. Expression patterns of Neil3 during embryonic brain development and neoplasia. *BMC Neurosci.* **2009**, *10*, 45. [CrossRef] [PubMed]
44. Shinmura, K.; Kato, H.; Kawanishi, Y.; Igarashi, H.; Goto, M.; Tao, H.; Inoue, Y.; Nakamura, S.; Misawa, K.; Mineta, H.; et al. Abnormal Expressions of DNA Glycosylase Genes NEIL1, NEIL2, and NEIL3 Are Associated with Somatic Mutation Loads in Human Cancer. *Oxid. Med. Cell. Longev.* **2016**, *2016*, 1546392. [CrossRef] [PubMed]
45. Kauffmann, A.; Rosselli, F.; Lazar, V.; Winnepenninckx, V.; Mansuet-Lupo, A.; Dessen, P.; van den Oord, J.J.; Spatz, A.; Sarasin, A. High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene* **2008**, *27*, 565–573. [CrossRef] [PubMed]
46. D'Errico, M.; Parlanti, E.; Pascucci, B.; Fortini, P.; Baccarini, S.; Simonelli, V.; Dogliotti, E. Single nucleotide polymorphisms in DNA glycosylases: From function to disease. *Free Radic. Biol. Med.* **2017**, *107*, 278–291. [CrossRef]
47. Barry, K.H.; Koutros, S.; Berndt, S.I.; Andreotti, G.; Hoppin, J.A.; Sandler, D.P.; Burdette, L.A.; Yeager, M.; Freeman, L.E.; Lubin, J.H.; et al. Genetic variation in base excision repair pathway genes, pesticide exposure, and prostate cancer risk. *Environ. Health Perspect.* **2011**, *119*, 1726–1732. [CrossRef]
48. Bethke, L.; Webb, E.; Murray, A.; Schoemaker, M.; Johansen, C.; Christensen, H.C.; Muir, K.; McKinney, P.; Hepworth, S.; Dimitropoulou, P.; et al. Comprehensive analysis of the role of DNA repair gene polymorphisms on risk of glioma. *Hum. Mol. Genet.* **2008**, *17*, 800–805. [CrossRef]
49. Cipollini, M.; Figlioli, G.; Maccari, G.; Garritano, S.; De Santi, C.; Melaiu, O.; Barone, E.; Bambi, F.; Ermini, S.; Pellegrini, G.; et al. Polymorphisms within base and nucleotide excision repair pathways and risk of differentiated thyroid carcinoma. *DNA Repair (Amst.)* **2016**, *41*, 27–31. [CrossRef]
50. Allione, A.; Pardini, B.; Viberti, C.; Oderda, M.; Allasia, M.; Gontero, P.; Vineis, P.; Sacerdote, C.; Matullo, G. The prognostic value of basal DNA damage level in peripheral blood lymphocytes of patients affected by bladder cancer. *Urol. Oncol.* **2018**, *36*, 241.e15–241.e23. [CrossRef]
51. Lee, P.H.; Shatkay, H. F-SNP: Computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res.* **2008**, *36*, D820–D824. [CrossRef] [PubMed]
52. Vymetalkova, V.; Pardini, B.; Rosa, F.; Jiraskova, K.; Di Gaetano, C.; Bendova, P.; Levy, M.; Veskrnova, V.; Buchler, T.; Vodickova, L.; et al. Polymorphisms in microRNA binding sites of mucin genes as predictors of clinical outcome in colorectal cancer patients. *Carcinogenesis* **2017**, *38*, 28–39. [CrossRef] [PubMed]
53. Schneiderova, M.; Naccarati, A.; Pardini, B.; Rosa, F.; Gaetano, C.D.; Jiraskova, K.; Opattova, A.; Levy, M.; Veskrna, K.; Veskrnova, V.; et al. MicroRNA-binding site polymorphisms in genes involved in colorectal cancer etiopathogenesis and their impact on disease prognosis. *Mutagenesis* **2017**, *32*, 533–542. [CrossRef] [PubMed]

54. Hofer, P.; Hagmann, M.; Brezina, S.; Dolejsi, E.; Mach, K.; Leeb, G.; Baierl, A.; Buch, S.; Sutterluty-Fall, H.; Karner-Hanusch, J.; et al. Bayesian and frequentist analysis of an Austrian genome-wide association study of colorectal cancer and advanced adenomas. *Oncotarget* **2017**, *8*, 98623–98634. [CrossRef] [PubMed]
55. Lemon, S.C.; Roy, J.; Clark, M.A.; Friedmann, P.D.; Rakowski, W. Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Ann. Behav. Med.* **2003**, *26*, 172–181. [CrossRef] [PubMed]
56. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc. B* **1995**, *57*, 289–300. [CrossRef]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).