The Pharmacogenomics Journal (2016) 16, 173–179

www.nature.com/tpj

ORIGINAL ARTICLE

A subset of genetic susceptibility variants for colorectal cancer also has prognostic value

S Noci[1], M Dugo[2], F Bertola[3], F Melotti[4,6], A Vannelli[5,7], TA Dragani[1] and A Galvan[1]

We investigated the possible influence of 86 single-nucleotide polymorphisms (SNPs), known to associate with the risk of colorectal cancer (CRC), on overall survival and time to recurrence (TTR) in 733 Italian CRC patients followed up for up to 84 months after surgery. In the Cox multivariate analysis, adjusted for gender, age, pathological stage and adjuvant chemotherapy (yes/no), the risk of death significantly increased by rare allele count ($P < 0.05$) for rs1801133 (*MTHFR*), rs4939827 (*SMAD7*), rs2306283 (*SLCO1B1*) and rs12898159 (*BMP4*), whereas for rs736775 (*GPX3*) the opposite was observed. Two additional SNPs associated with TTR, namely rs16892766 (downstream of *EIF3H*) and rs10749971 (*COLCA2*). Our findings show that some genetic variants previously found to associate with CRC risk are also associated with survival after treatment. The identification of alleles defining subgroups of patients with worse clinical outcome may have application in developing pharmacogenetic strategies aimed at personalizing CRC treatment.

The Pharmacogenomics Journal (2016) **16,** 173–179; doi:10.1038/tpj.2015.35; published online 12 May 2015

INTRODUCTION

Colorectal cancer (CRC) is the third most frequent neoplasm worldwide, with over one million new incident cases diagnosed in 2008.[1] Environmental factors, in particular alcohol consumption and smoking, are associated with an increased risk of sporadic CRC.[2] However, a genetic component is also believed to be relevant. A large twin study, published in 2000, estimated that 35% of colorectal cancer risk may be explained by heritable factors.[3] Since then, numerous genome-wide association studies have identified single-nucleotide polymorphisms (SNPs) associated with the risk of CRC. Currently, at least 20 genomic regions (and various candidate genes) are implicated in disease risk, including 8q23.3 (*EIF3H*), 8q24.21 (*CASC8*), 10p14, 11q23.1 (*COLCA1*, *COLCA2*), 14q22.2 (*BMP4*), 15q13.3 (*GREM1*), 16q22.1 (*CDH1*), 18q21.1 (*SMAD7*), 19q13.11 (*RHPN2*) and 20p12.3 (*BMP2*).[4–12]

Some of the SNPs associated with CRC susceptibility have also been tested in different cohorts for a possible impact on CRC outcome, but the prognostic significance of these variants remains controversial. For example, polymorphisms in *CASC8*, mapping in the 8q24.21 region, were associated with an increased risk of death in one study,[13] which contradicted earlier reports of no association.[14,15] rs10795668, an SNP in the gene-free 10p14 region, did not associate with CRC prognosis in a study of patients with Scottish ancestry,[15] whereas it associated with both a better clinical outcome in a Spanish study[16] and a reduced risk of recurrence in a Chinese study.[17]

For patients with stages III and IV CRC and for those with high-risk stage II CRC, adjuvant chemotherapy—of which 5-fluorouracil (5-FU) is the treatment standard—is generally recommended, as it has been demonstrated to improve both disease-free survival and overall survival.[18] However, as a substantial percentage of treated CRC patients develop a recurrence or metastasis within 5 years of diagnosis,[19] it is important to understand why some patients relapse, whereas others do not.

Genetics is helping to elucidate these differences in treatment response. In particular, polymorphisms in *TYMS* and *MTHFR* genes involved in 5-FU action have been linked to the variability in patients' response to this drug. Polymorphisms in *TYMS*, which encodes thymidylate synthase, were predictive of survival in 89 patients who received 5-FU,[20] although these results were not confirmed by a more recent study.[21] Polymorphisms in *MTHFR*, which encodes methylenetetrahydrofolate reductase, were associated with toxicity (for example, diarrhea and mucositis) from 5-FU-based chemoradiation therapy but not with efficacy,[21] although no statistically significant association was found between markers of toxicity outcomes and the C667T *MTHFR* polymorphism in a randomized trial of chemotherapy for advanced colorectal cancer.[22] To further investigate the genetic determinants of treatment response and to clarify some of the conflicting data on the prognostic role of CRC risk-associated SNPs, we genotyped 96 SNPs in 770 Italian patients operated for colorectal adenocarcinoma. The association of genotype with both overall survival and time to recurrence (TTR) was evaluated, taking into account the administration of adjuvant chemotherapy and the clinicopathological parameters that may have influenced tumor progression and outcome.

METHODS

Patients

Since May 2008, we collected baseline and follow-up clinicopathological data and blood samples from patients treated surgically for colorectal

[1]Department of Predictive and Preventive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [2]Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [3]Consortium for Human Molecular Genetics, University of Milano-Bicocca, Monza, Italy; [4]Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy and [5]Division of General Surgery B, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. Correspondence: Dr TA Dragani, Department of Predictive and Preventive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Via G.A. Amadeo 42, Milan 20133, Italy.
E-mail: tommaso.dragani@istitutotumori.mi.it
[6]Present address: Pathology Service, Azienda Ospedaliera di Desenzano del Garda, Manerbio, Italy.
[7]Present address: Division of Oncologic and Gastrointestinal Surgery, Valduce Hospital, Como, Italy.
Received 29 September 2014; revised 3 February 2015; accepted 24 March 2015; published online 12 May 2015

adenocarcinoma at the Fondazione IRCCS Istituto Nazionale dei Tumori in Milan, Italy. In December 2012, when this study began, material from 770 Italian patients was available. The study protocol was approved by the Independent Ethical Committee of this institute, and written informed consent was provided by all patients for the use of their samples for research purposes.

Histopathological analysis of resected tumor specimens confirmed the diagnosis of colorectal adenocarcinoma in all cases. After surgery, some patients received 5-FU-based adjuvant chemotherapy, chosen on the basis of their clinicopathological characteristics. The regimens used included 5-FU alone, 5-FU in the XELOX, FOLFOX or FOLFIRI regimens, and 5-FU in combination with a monoclonal antibody (bevacizumab or cetuximab).

SNP selection and genotyping

We searched PubMed (www.ncbi.nlm.nih.gov/pubmed) for studies in English describing SNPs associated with CRC risk or prognosis. All SNPs discovered in genome-wide association studies, case–control association studies and CRC candidate-gene analyses were preliminarily selected, as were SNPs identified by pharmacogenomic studies in patients receiving fluoropyrimidine-based adjuvant chemotherapy. A list of 96 CRC-relevant SNPs was submitted for evaluation on Illumina's Assay Design Tool (Illumina Assay Design Tool for Array Probe Design. http://support.illumina.com/array/array_software/assay_design_tool.html, which evaluates loci for expected genotyping success rate (assigning a 'final SNP score') and flags loci with an error message if an assay cannot be designed for technical reasons. Scores range from 0 to 1.1, with values < 0.4 indicating a low success rate, values from 0.4 to 0.6 indicating a moderate success rate and values from 0.6 to 1.1 indicating a high success rate.[23] The 96 SNPs evaluated had final SNP scores ranging from 0.41 to 1.1 and no error messages (Supplementary Table 1), and thus they were all included in the design of custom GoldenGate Genotyping assays with VeraCode technology (Illumina, San Diego, CA, USA).

Genomic DNA was extracted from frozen blood samples using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA). Samples of DNA were quantified using the NanoDrop 2000c UV–vis spectrophotometer (Thermo Scientific, Waltham, MA, USA) and diluted to 50 ng μl^{-1} in water. Subsequently, 15 µl of each sample was placed into a well of a 96-well plate. Two samples on each plate were assayed in duplicate to demonstrate data reproducibility, and one well received only water as negative control.

High-throughput SNP screening was performed using the custom-designed GoldenGate Genotyping assays with VeraCode technology on an Illumina BeadXpress Reader platform according to the manufacturer's protocol (Illumina). Data generated were analyzed using the Illumina's GenomeStudio data analysis software (v2011.1, Illumina), which performs automated genotype clustering and calling. SNPs with an Illumina GenCall_10 score $\geqslant 0.5$ and a call rate $\geqslant 80\%$ were automatically included; for one SNP (rs10929302) with a GenCall score of 0.32, a visual exam of the cluster plot supported its inclusion. In addition, we excluded from the study SNPs that had minor allele frequency (MAF) < 0.025.

Statistical analyses

Two end points were considered in the survival analyses. The time to recurrence (TTR) was defined as the time from the date of surgery to the first local or distant recurrence or to the diagnosis of a second primary colon cancer. The calculation of TTR did not consider patients who were alive and free of recurrences, who died without having had a recurrence or who had distant metastasis at the time of diagnosis. Overall survival was defined as the time from surgery to death from any cause or to the day of last follow-up, which was truncated at 84 months or in June 2013, whichever came first.

Univariate Cox analyses were carried out to examine the relationships between outcome (TTR and overall survival) and clinicopathological characteristics, except for adjuvant chemotherapy, which was tested by multivariate analysis stratified for age at surgery ($\leqslant$64 vs $>$64 years), gender and pathological TNM stage (I and II vs III and IV; TNM, T category describes the primary tumor site, N category describes the regional lymph node involvement, M category describes the presence or otherwise of distant metastatic spread; http://www.uicc.org/resources/tnm). The Kaplan–Meier method was used to compare overall survival in patients stratified by TNM stage into two groups.

Genotype frequencies were tested for Hardy–Weinberg equilibrium using the χ^2 test. The associations of genotypes with overall survival and TTR were evaluated in a Cox proportional hazards model adjusted for age at surgery ($\leqslant$64 vs $>$64 years), gender, TNM stage (I and II vs III and IV) and adjuvant chemotherapy; the choice of these covariates was based on their correlation with overall survival or TTR and, for variables with a substantial amount of missing data, on the results of pairwise comparisons with TNM stage obtained using the χ^2 test. In the multivariate analyses, each SNP was initially evaluated for association with overall survival or TTR using the trend test: a numerical value (0, 1 and 2) was assigned to each patient's genotype according to the number of rare alleles. SNPs that significantly associated with survival or TTR by the trend test ($P < 0.05$) were further analyzed to test for the additive or dominant model of penetrance.[24] For this purpose, homozygosity of the common allele was set as the reference value. In the additive model, we calculated hazard ratios (HRs) for the three genotype groups using the reference group as denominator (that is, patients who were homozygous at the common allele had HR = 1). In the dominant model, the numerator of the HR referred to patients who were either homozygous for the rare allele or heterozygous, with the reference group in the denominator.

The Kaplan–Meier method was used to obtain survival and recurrence-free curves according to genotype for selected SNPs. All statistical tests were two-sided and were carried out using the freely available R packages (http://www.r-project.org/).

RESULTS

Patients' clinicopathological characteristics and outcome

In this study, we investigated the influence of genetics on survival and TTR after surgical resection of CRC, by genotyping 770 Italian patients for 96 SNPs associated with CRC risk. Genotyping, performed using custom-designed GoldenGate assays, provided reliable data ($\geqslant$80% SNP call rate) for 733 patients, who were included in the analysis (Table 1). The patients' median age at surgery was 64 years, but there was a wide age range; male gender predominated (58.4%). The patients' tumors originated in the colon in ~48% of cases and in the rectum in 52%. Metastasis was present at surgery (TNM stage IV) in 90 cases (12.3%). Adjuvant chemotherapy had been administered to 382 patients but not to 275 patients (data are missing for 76 cases). For the other clinicopathological characteristics listed in Table 1, the high rate of missing data is attributable to the retrospective nature of this study. The patients were followed up for a median period of 38.2 months. At the end of the follow-up period (84 months after surgery or June 2013, whichever came first), 620 patients (84.6%) were alive.

To identify clinicopathological characteristics that associated with the risk of developing a recurrence (expressed as TTR) and with overall survival, exploratory univariate Cox analyses were performed (Table 2). With regard to TTR, borderline statistical associations were observed for age at surgery, gender and histological grade 3 ($P < 0.05$), and no association was found for smoking habit or tumor site. Instead, statistically significant associations were found for TNM stage (HRs of 1.9 and 3.8 for stages II and III, respectively), pT and pN categories (HRs $\geqslant$ 1.7), perineural invasion (HR = 2.1) and tumor deposits (HR = 6.0), all with $P < 1.0 \times 10^{-2}$. Overall survival did not associate with age at surgery, gender or smoking habit, whereas it weakly associated with tumor site, histological grade 3 and tumor deposits; strong associations ($P < 1.0 \times 10^{-3}$) were instead found for TNM stage (HRs $\geqslant$ 4.7), for individual TNM categories (HRs $\geqslant$ 3.5) and for perineural invasion (HR = 3.3). It is noteworthy that multivariate analysis of adjuvant chemotherapy, carried out by stratifying for age at surgery, gender and TNM stage, associated with overall survival (HR = 0.44; 95% CI, 0.23–0.83; $P = 0.012$) but not with TTR, suggesting that in our series adjuvant chemotherapy prolonged survival without reducing events of recurrence.

To confirm the observed impact of pathological TNM stage on overall survival, we combined the lower stages (I and II) and higher stages (III and IV) into two groups and performed a Kaplan–Meier analysis. A huge difference in survival was seen

Table 1. Clinicopathological characteristics of 733 patients with colorectal cancer

Characteristic	*Patients*[a]
Age at surgery (years), median (range)	64 (24–93)
Gender	
Male	428 (58.4)
Female	305 (41.6)
Smoking habit[b]	
Never	255
Ever	201
Tumor site[b]	
Colon	348
Rectum	369
Pathological T category	
T1	86 (11.7)
T2	162 (22.1)
T3	378 (51.6)
T4	107 (14.6)
Pathological N category	
N0	421 (57.4)
N1	206 (28.1)
N2	106 (14.5)
Pathological M category	
M0	643 (87.7)
M1	90 (12.3)
TNM stage[c]	
I	200 (27.3)
II	208 (28.4)
III	233 (31.8)
IV	90 (12.3)
Histologic grade[b]	
1	19
2	435
3	157
Perineural invasion[b]	
No	276
Yes	256
Tumor deposits[b]	
No	48
Yes	85
Adjuvant chemotherapy[b]	
No	275
Yes	382
Follow-up status[d]	
Follow-up period (months), median (range)	38.2 (13.1–248)
Patients alive at last follow-up	620 (84.6)

Abbreviation: TNM, T category describes the primary tumor site, N category describes the regional lymph node involvement, M category describes the presence or otherwise of distant metastatic spread; http://www.uicc.org/resources/tnm. [a]Values are *n* (%) unless otherwise stated. [b]Total number of cases is <733 owing to missing data. [c]For two patients, TNM stage was not available. [d]Follow-up period truncated at 84 months or in June 2013.

between the two groups ($P=2\times10^{-19}$, log-rank test), indicating that the collapse of the four TNM stages into two produced a useful covariate for multivariate analyses of SNP genotype with survival.

SNP genotyping results

During the analysis of raw genotype data, five SNPs (rs3918290, rs1052133, rs4779584, rs961253 and rs2567608) were eliminated because of poor clustering. For the remaining 91 SNPs, the MAF was >0.025 in all cases except for three SNPs (rs1800932, rs2234671 and rs12299484) with MAF = 0 and two (rs3729740 and rs78378222) with MAF near zero (Supplementary Table 1). These five SNPs were excluded for not being informative, leaving 86 SNPs for further analysis. Genotype frequencies for these SNPs were tested for Hardy–Weinberg equilibrium, and no deviation was observed ($P>0.05$ for all SNPs, χ^2 test; data not shown).

Seven CRC risk-associated SNPs also associate with survival or TTR

Multivariate regression analysis of the 86 SNPs with overall survival or TTR included age at surgery, gender, pathological TNM stage and adjuvant chemotherapy as covariates. As pathological stage incorporates tumor size and lymph node involvement, which associated with survival in the univariate analyses, only TNM stage (collapsed into two groups) was incorporated in the multivariate model. Moreover, considering the high percentage of missing data for perineural invasion and tumor deposits (which also associated with survival in univariate analyses), we examined their association with pathological stage and found $P<10^{-16}$ for both (χ^2 test); therefore, it was not essential to include them in the model.

In the multivariate analysis, five SNPs associated with overall survival in a trend test with increasing numbers of the rare allele (Table 3). These SNPs all mapped within or in close proximity to a known gene: for four SNPs (in *MTHFR*, *SMAD7*, *SLCO1B1* and *BMP4*), the rare allele was associated with lower survival (HR > 1), whereas for rs736775 in *GPX3* the opposite was observed. When an additive model was tested, patients who were homozygous for the rare allele were at a significantly greater risk of death than those homozygous for the common allele (reference) for rs1801133 (*MTHFR*; HR = 2.50), rs4939827 (*SMAD7*; HR = 2.30), rs2306283 (*SLCO1B1*; HR = 2.18) and rs12898159 (*BMP4*; HR = 1.82); for rs736775 (GPX3), homozygosity at the rare allele (AA) did not significantly influence survival, possibly owing to the relatively small size of this group ($n=99$). The presence of only one copy of the rare allele (heterozygosity) was sufficient to significantly increase the probability of death over that in the reference group for rs1801133 (AG vs GG, HR = 1.85) and rs2306283 (AG vs AA, HR = 1.96), whereas for rs736775 it favored survival (GA vs GG, HR = 0.55). In a dominant model, the presence of the rare allele was significantly associated with the risk of death in the cases of rs1801133 (*MTHFR*; HR = 2.02) and rs2306283 (*SLCO1B1*; HR = 2.01) and with survival for rs736775 (*GPX3*; HR = 0.57). These results indicate that for SNPs rs1801133 and rs2306283 their rare alleles may display an additive mode of penetrance, that is, the penetrance of the heterozygote genotype is between the penetrances of the two homozygote genotypes; for SNPs rs4939827 and rs12898159 their rare alleles showed a recessive mode of penetrance, whereas for rs736775 a compound heterozygosity effect seems to have a role.

With regard to TTR, a trend analysis on the effect of the number of copies of the rare allele identified two SNPs associated with recurrence (Table 4). For rs16892766, which maps downstream of *EIF3H*, the HR of 1.79 indicated that the rare allele increased the risk of relapse. In the dominant model, patients carrying the rare C allele had a higher risk of developing recurrences than patients homozygous for the common A allele (HR = 2.02). In the additive model, homozygosity for the rare allele (CC) did not alter the risk of recurrence, most likely because of the very small number of patients ($n=3$). For rs10749971, mapping downstream of *COLCA2*, the HR of 0.72 indicated that the rare allele was associated with a lower risk of relapse.

Kaplan–Meier survival analysis was carried out for the SNPs that associated with overall survival (rs1801133, rs4939827 and rs2306283) or with TTR (rs16892766) with a $P_{trend}<0.01$. This

Table 2. Univariate analyses of the association of clinicopathological characteristics with outcome in 733 patients with colorectal cancer

Factor	*Patients (n)*	*Time to recurrence*[a]		*Overall survival*[a]	
		HR (95% CI)	P-*value*	*HR (95% CI)*	P-*value*
Age, years					
≤64	377	1 (ref.)		1 (ref.)	
>64	356	0.7 (0.5–0.9)	0.018	1.2 (0.8–1.8)	0.273
Gender					
Male	428	1.4 (1.0–1.9)	0.027	1.1 (0.8–1.6)	0.562
Female	305	1 (ref)		1 (ref)	
Smoking habit					
Never	255	0.8 (0.6–1.2)	0.255	0.8 (0.5–1.3)	0.324
Ever	200	1 (ref)		1 (ref)	
Tumor site					
Colon	348	1 (ref)		1 (ref)	
Rectum	369	0.9 (0.7–1.2)	0.510	0.7 (0.5–1.0)	0.042
pT category					
T1	86	1 (ref)		1 (ref)	
T2	162	1.7 (0.8–3.7)	0.190	3.5 (0.4–28.6)	0.239
T3	378	3.7 (1.8–7.6)	3.50E – 04	15.9 (2.2–114.3)	6.10E – 03
T4	107	9.3 (4.4–19.7)	3.70E – 09	52.6 (7.2–382.3)	9.00E – 05
pN category					
N0	421	1 (ref)		1 (ref)	
N+	312	4.1 (3.0–5.6)	2.00E – 16	4.6 (3.0–7.0)	2.00E – 12
pM category					
M0	643	NA		1 (ref)	
M1	90	NA		13.8 (9.4–20.1)	2.00E – 16
TNM stage					
I	200	1 (ref)		1 (ref)	
II	208	1.9 (1.1–3.4)	3.00E – 02	4.7 (1.6–14.1)	5.20E – 03
III	233	3.8 (2.3–6.4)	3.90E – 07	8.3 (2.9–23.3)	6.70E – 05
IV	90	NA		64.4 (23.3–177.9)	1.00E – 15
Histologic grade					
1	19	1 (ref)		1 (ref)	
2	435	2.1 (0.7–6.6)	0.218	3.6 (0.5–26.4)	0.204
3	157	4.4 (1.4–14.2)	0.013	10.5 (1.4–77.4)	0.021
Perineural invasion					
No	276	1 (ref)		1 (ref)	
Yes	256	2.1 (1.4–3.0)	7.00E-05	3.3 (2.1–5.4)	4.40E – 07
Tumor deposits					
No	48	1 (ref)		1 (ref)	
Yes	85	6.0 (2.1–17.0)	6.30E – 04	3.5 (1.2–10.0)	0.020
Adjuvant chemotherapy[b]					
No	275	1 (ref)		1 (ref)	
Yes	382	0.99 (0.54–1.82)	0.979	0.44 (0.23–0.83)	0.012

Abbreviations: CI, confidence interval; HR, hazard ratio; NA, not applicable; TNM, T category describes the primary tumor site, N category describes the regional lymph node involvement, M category describes the presence or otherwise of distant metastatic spread; http://www.uicc.org/resources/tnm. [a]Follow-up at 84 months. [b]Analysis stratified for median age at surgery, gender and TNM stage (stages I and II vs III and IV). NA, Not applicable because patients with metastases at surgery were excluded from the calculation of time to recurrence.

analysis showed that the patterns of the curves differed among the four SNPs (Figure 1). For example, carriers of at least one copy of the rare allele of rs1801133 had shorter survival than patients who were homozygous for the common allele (Figure 1a), whereas at rs4939827 homozygosity—but not heterozygosity—for the rare allele conferred shorter survival compared with patients homozygous for the common allele (Figure 1b). In the case of rs16892766 associated with TTR, where only three patients were homozygous for the rare allele, the Kaplan–Meier curves clearly showed the association of heterozygosity with a greater rate of recurrence (Figure 1d).

DISCUSSION

This study examined the prognostic role of genetic variants known to influence the risk of CRC, using a uniform series of Italian patients treated at a single institution over a period of < 5 years. Seven SNPs were found to have prognostic value, including five

Table 3. Genetic variants associating with overall survival in 733 patients with colorectal cancer at multivariate Cox analysis

SNP	*Region*	*Gene*	*Genotype*	N[a]	*HR*[b]	*95% CI*	P-*value*[c]
rs1801133	1p36.22	MTHFR	0, 1, 2	728	1.56	1.16–2.09	0.003
			GG	210	1.00		
			AG	368	1.85	1.07–3.22	0.028
			AA	150	2.50	1.35–4.65	0.004
			AG+AA	518	2.02	1.19–3.43	0.009
rs4939827	18q21.1	SMAD7	0, 1, 2	726	1.52	1.13–2.05	0.006
			AA	248	1.00		
			AG	358	1.20	0.73–1.98	0.477
			GG	120	2.30	1.31–4.04	0.004
			AG+GG	478	1.46	0.92–2.33	0.112
rs2306283	12p12.1	SLCO1B1	0, 1, 2	716	1.48	1.11–1.97	0.008
			AA	229	1.00		
			AG	362	1.96	1.17–3.28	0.011
			GG	125	2.18	1.17–4.07	0.014
			AG+GG	487	2.01	1.22–3.31	0.006
rs736775	5q33.1	GPX3	0, 1, 2	719	0.71	0.51–0.98	0.039
			GG	285	1.00		
			GA	335	0.55	0.35–0.86	0.009
			AA	99	0.66	0.34–1.27	0.211
			GA+AA	434	0.57	0.38–0.87	0.009
rs12898159	14q22.2	BMP4	0, 1, 2	717	1.35	1.01–1.80	0.046
			GG	207	1.00		
			GA	362	1.39	0.82–2.33	0.220
			AA	148	1.82	1.00–3.29	0.047
			GA+AA	510	1.51	0.92–2.47	0.110

Abbreviations: CI, confidence interval; HR, hazard ratio; SNP, single-nucleotide polymorphism. [a]For each SNP, the genotypes of a few patients were missing owing to genotyping failure. [b]Adjusted for age at surgery (median), gender, pathological TNM stage (I and II or III and IV), and adjuvant chemotherapy (yes/no). Follow-up data were censored at 84 months. [c]*P*-values derived from trend test on 0, 1 or 2 copies of the rare allele and on comparisons of genotypes containing the rare allele with that of homozygosity of the common genotype.

Table 4. Genetic variants found to associate with time to recurrence in 733 patients with colorectal cancer at multivariate Cox analysis

SNP	*Region*	*Gene*	*Genotype*	N[a]	*HR*[b]	*95% CI*	P-*value*[c]
rs16892766	8q23.3	EIF3H	0, 1, 2	631	1.79	1.17–2.72	0.007
			AA	534	1.00		
			AC	94	2.06	1.27–3.34	0.003
			CC	3	1.39	0.19–10.17	0.744
			AC+CC	97	2.02	1.26–3.25	0.004
rs10749971	11q23.1	COLCA2	0, 1, 2	632	0.72	0.53–0.97	0.034
			AA	247	1.00		
			AG	292	0.66	0.43–1.00	0.052
			GG	93	0.57	0.29–1.13	0.107
			AG+GG	385	0.64	0.43–0.95	0.027

Abbreviations: CI, confidence interval; HR, hazard ratio; SNP, single-nucleotide polymorphism. [a]For each SNP, the genotypes of a few patients were missing because of genotyping failure. [b]Adjusted for age at surgery (median), gender, pathological TNM stage (I and II or III and IV) and adjuvant chemotherapy (yes/no). Follow-up data were censored at 84 months. [c]*P*-values derived from trend test on 0, 1 or 2 copies of the rare allele and on comparisons of genotypes containing the rare allele with that of homozygosity of the common genotype.

associated with overall survival (rs1801133, rs4939827, rs2306283, rs736775 and rs12898159) and two with TTR (rs16892766 and rs10749971). Each of these SNPs is located within or near a known gene. Two of the SNPs associated with overall survival are missense mutations in genes whose products are involved in chemotherapy: rs1801133 maps to *MTHFR*, which is involved in 5-FU metabolism, whereas rs2306283 maps to *SLCO1B1*, which encodes a transporter of irinotecan, a component of the FOLFIRI regimen. Two other SNPs that associated with overall survival (rs4939827 and rs12898159) map to genes of the TGF-β pathway (*SMAD7* and *BMP4*, respectively).

These results require validation in independent cohorts of patients, as well as follow-up studies to determine the molecular mechanisms by which these SNPs influence both survival and prognosis. Moreover, because the regression analyses were adjusted for disease stage collapsed into only two classes (stages I and II vs stages III and IV), the SNPs identified here do not permit one to discern survival differences between stages II and III. Continued research in larger data sets, leading to the possible identification of other SNPs that differentiate patients with stage II vs III disease, will be an important extension of the present findings that could influence CRC treatment decisions.

The pivotal and well-established treatment of CRC is chemotherapy based on 5-FU, a pyrimidine analog that inhibits thymidylate synthase through the formation of an inhibitory ternary complex, consisting of its active metabolite 5-fluoro-2-deoxyuridine-5-monophosphate, the enzyme and its cofactor 5,10-methylenetetrahydrofolate (MTHF). Inhibition of thymidylate

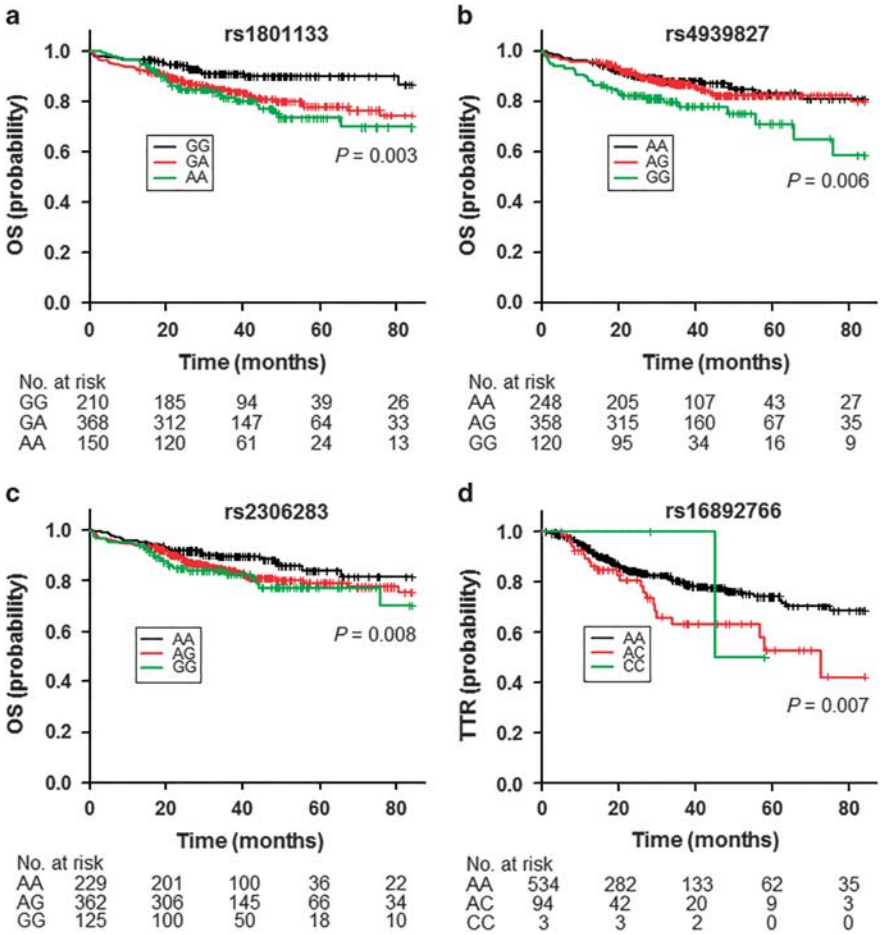

Figure 1. Kaplan–Meier curves representing the overall survival (**a**–**c**) and the time to recurrence (**d**) of 733 Italian patients with colorectal adenocarcinoma according to their genotypes at (**a**) rs1801133, (**b**) rs4939827, (**c**) rs2306283 and (**d**) rs16892766. The homozygosity carrier status of the rare allele is shown with green lines, the homozygosity carrier status of the common allele is shown in black and the heterozygosity is shown in red. Crosses denote censored samples. Below the figures are reported the number of patients at risk at the specified times of follow-up.

synthase results in depleted intracellular thymidylate levels, ultimately suppressing DNA synthesis. Despite this enzyme being the target of 5-FU action, no association was found in this study between the two tested variants in its gene *TYMS* and patients' survival, confirming the results of Thomas *et al.* (2011).[21] 5-FU's inhibition of thymidylate synthase requires the presence of MTHF, whose levels depend on the activity of methylenetetrahydrofolate reductase, which uses MTHF as substrate. In the gene that encodes this enzyme, *MTHFR*, two SNPs have previously been associated with CRC risk[21] and, in this study, one of them (rs1801133; Ala222Val) associated with overall survival. This SNP has been widely studied for its influence on the efficacy and toxicity of fluoropyrimidine agents (reviewed in[25]). The minor allele of rs1801133 encodes a missense valine mutation that reduces, without abolishing, the enzymatic activity of the common alanine allele.[26] The resulting increased intracellular concentrations of MTHF may augment the cytotoxic activity of 5-FU by enhancing the formation and stability of the ternary inhibitory complex.

For patients with metastatic CRC, the standard first-line treatment consists of folinic acid (leucovorin), 5-FU and irinotecan (FOLFIRI regimen) combined with bevacizumab and cetuximab. Irinotecan uptake depends on the activity of the solute carrier organic anion transporter family member 1B1 (encoded by *SLCO1B1*). Some polymorphisms in this gene impair transporter function.[27] The present study investigated only one SNP in *SLCO1B1* (rs2306283), finding it associated with overall survival. This SNP has previously been reported to be associated with neutropenia in patients treated with irinotecan.[28] Our results suggest that germline polymorphisms of genes involved in 5-FU and irinotecan pathways may affect CRC patients' survival owing to modulation of toxicity or efficacy of adjuvant treatments.

Two additional SNPs in this study that associated with overall survival are linked to the transforming growth factor-β (TGF-β) pathway, which has an important role in cell proliferation, differentiation and migration. rs4939827 maps to the intron between exons 3 and 4 in SMAD family member 7 (*SMAD7*). In the present study, the minor allele of rs4939827 was associated with reduced overall survival, confirming the results of a recent meta-analysis of five cohort studies;[29] another SNP (rs4464148), mapping to the same intron of *SMAD7*, did not, however, associate with survival. The other SNP linked to the TGF-β pathway is rs12898159, one of five investigated SNPs mapping to *BMP4*, which encodes bone morphogenetic protein 4. rs12898159 maps downstream of *BMP4*, 1163 bp from the 3′-UTR; it also maps 89 bp upstream of *MIR5580*, a recently identified microRNA. A reporter gene study that compared the alleles of this SNP found no difference in enhancer activity in CRC cell lines,[12] and thus its function remains unknown. Our study is not the first to associate SNPs in genes of the TGF-β pathway with colorectal cancer: previous studies found that SNPs in or near *SMAD7* and *BMP4*, as well as in *GREM1*, *BMP2* and *RHPN2*, associated with CRC susceptibility (reviewed in[30]) and with colon cancer development and progression (reviewed in[31]).

When we considered CRC progression, we found two SNPs that associated with TTR. rs16892766, which maps 23686 bp downstream of the eukaryotic translation initiation factor 3, subunit H (*EIF3H*), was first identified as a CRC susceptibility SNP in a genome-wide association study.[7] However, this polymorphism did not associate with *EIF3H* mRNA expression in a series of colorectal adenomas and carcinomas.[32] Nevertheless, an *in vitro* functional study carried out in breast cancer cell lines demonstrated the involvement of *EIF3H* in cell viability, cell cycle progression and colony formation.[33] The other SNP associated with TTR, rs10749971, maps downstream of *COLCA2* and upstream of *COLCA1* (which encode colorectal cancer-associated 2 and 1, respectively). These genes lie in the 11q23.1 region, which has been implicated in CRC risk by a study and a meta-analysis.[6,9] The minor (G) allele of rs10749971 was previously reported to associate with a reduced risk of recurrence in patients with stage III CRC receiving 5-FU-based adjuvant chemotherapy,[13] and our results confirm that this allele favors prognosis.

In conclusion, the identification of a set of CRC-risk alleles that are also involved in prognosis suggests that the genes with these polymorphisms—and their encoded proteins—modulate both patients' survival and the progression of their disease. In addition, the identification of alleles that define subgroups of patients with worse clinical outcome may have application in the development of pharmacogenetic strategies aimed at personalizing CRC treatment.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank Dr Valerie Matarese for scientific editing and writing support. This work was supported in part by a grant from the Italian Association for Cancer Research (AIRC, grant no. 12162). The funder had no role in the design and conduct of the study, in the collection, analysis or interpretation of the data, or in the preparation, review or approval of the manuscript.

REFERENCES

1 Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; **127**: 2893–2917.
2 Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. *Int J Cancer* 2009; **125**: 171–180.
3 Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000; **343**: 78–85.
4 Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008; **40**: 1426–1435.
5 Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 2010; **42**: 973–977.
6 Lascorz J, Forsti A, Chen B, Buch S, Steinke V, Rahner N *et al.* Genome-wide association study for colorectal cancer identifies risk polymorphisms in German familial cases and implicates MAPK signalling pathways in disease susceptibility. *Carcinogenesis* 2010; **31**: 1612–1619.
7 Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 2008; **40**: 623–630.
8 Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, Howarth K *et al.* Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet* 2011; **7**: e1002105.
9 Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet* 2012; **131**: 217–234.
10 Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 2008; **40**: 631–637.
11 Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007; **39**: 989–994.
12 Lubbe SJ, Pittman AM, Olver B, Lloyd A, Vijayakrishnan J, Naranjo S *et al.* The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. *Oncogene* 2012; **31**: 3777–3784.
13 Dai J, Gu J, Huang M, Eng C, Kopetz ES, Ellis LM *et al.* GWAS-identified colorectal cancer susceptibility loci associated with clinical outcomes. *Carcinogenesis* 2012; **33**: 1327–1331.
14 Cicek MS, Slager SL, Achenbach SJ, French AJ, Blair HE, Fink SR *et al.* Functional and clinical significance of variants localized to 8q24 in colon cancer. *Cancer Epidemiol Biomark Prev* 2009; **18**: 2492–2500.
15 Tenesa A, Theodoratou E, Din FV, Farrington SM, Cetnarskyj R, Barnetson RA *et al.* Ten common genetic variants associated with colorectal cancer risk are not associated with survival after diagnosis. *Clin Cancer Res* 2010; **16**: 3754–3759.
16 Abuli A, Lozano JJ, Rodriguez-Soler M, Jover R, Bessa X, Munoz J *et al.* Genetic susceptibility variants associated with colorectal cancer prognosis. *Carcinogenesis* 2013; **34**: 2286–2291.
17 Xing J, Myers RE, He X, Qu F, Zhou F, Ma X *et al.* GWAS-identified colorectal cancer susceptibility locus associates with disease prognosis. *Eur J Cancer* 2011; **47**: 1699–1707.
18 Sargent D, Sobrero A, Grothey A, O'Connell MJ, Buyse M, Andre T *et al.* Evidence for cure by adjuvant therapy in colon cancer: observations based on individual patient data from 20 898 patients on 18 randomized trials. *J Clin Oncol* 2009; **27**: 872–877.
19 Andre T, Boni C, Navarro M, Tabernero J, Hickish T, Topham C *et al.* Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. *J Clin Oncol* 2009; **27**: 3109–3116.
20 Marcuello E, Altes A, del Rio E, Cesar A, Menoyo A, Baiget M. Single nucleotide polymorphism in the 5' tandem repeat sequences of thymidylate synthase gene predicts for response to fluorouracil-based chemotherapy in advanced colorectal cancer patients. *Int J Cancer* 2004; **112**: 733–737.
21 Thomas F, Motsinger-Reif AA, Hoskins JM, Dvorak A, Roy S, Alyasiri A *et al.* Methylenetetrahydrofolate reductase genetic polymorphisms and toxicity to 5-FU-based chemoradiation in rectal cancer. *Br J Cancer* 2011; **105**: 1654–1662.
22 Braun MS, Richman SD, Thompson L, Daly CL, Meade AM, Adlard JW *et al.* Association of molecular markers with toxicity outcomes in a randomized trial of chemotherapy for advanced colorectal cancer: the FOCUS trial. *J Clin Oncol* 2009; **27**: 5519–5528.
23 Pelucchi S, Mariani R, Calza S, Fracanzani AL, Modignani GL, Bertola F *et al.* CYBRD1 as a modifier gene that modulates iron phenotype in HFE p.C282Y homozygous patients. *Haematologica* 2012; **97**: 1818–1825.
24 Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc* 2011; **6**: 121–133.
25 De Mattia E, Toffoli G. C677T and A1298C MTHFR polymorphisms, a challenge for antifolate and fluoropyrimidine-based therapy personalisation. *Eur J Cancer* 2009; **45**: 1333–1351.
26 Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 1995; **10**: 111–113.
27 Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. *Pharmacol Rev* 2011; **63**: 157–181.
28 Innocenti F, Kroetz DL, Schuetz E, Dolan ME, Ramirez J, Relling M *et al.* Comprehensive pharmacogenetic analysis of irinotecan neutropenia and pharmacokinetics. *J Clin Oncol* 2009; **27**: 2604–2614.
29 Phipps AI, Newcomb PA, Garcia-Albeniz X, Hutter CM, White E, Fuchs CS *et al.* Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. *Gastroenterology* 2012; **143**: 51–54.
30 Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet* 2009; **10**: 353–358.
31 Xu Y, Pasche B. TGF-beta signaling alterations and susceptibility to colorectal cancer. *Hum Mol Genet* 2007; **16**: R14–R20.
32 Pittman AM, Naranjo S, Jalava SE, Twiss P, Ma Y, Olver B *et al.* Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. *PLoS Genet* 2010; **6**: e1001126.
33 Mahmood SF, Gruel N, Chapeaublanc E, Lescure A, Jones T, Reyal F *et al.* A siRNA screen identifies RAD21, EIF3H, CHRAC1 and TANC2 as driver genes within the 8q23, 8q24.3 and 17q23 amplicons in breast cancer with effects on cell growth, survival and transformation. *Carcinogenesis* 2014; **35**: 670–682.

Supplementary Information accompanies the paper on the The Pharmacogenomics Journal website (http://www.nature.com/tpj)