Int. J. Cancer: **124,** 1887–1891 (2009)

Common germline variation in mismatch repair genes and survival after a diagnosis of colorectal cancer

Thibaud Koessler[1]*, Elizabeth M. Azzato[1,2], Barbara Perkins[1], Robert J. Macinnis[1,3], David Greenberg[4], Douglas F. Easton[5] and Paul D.P. Pharoah[1]

[1]*Department of Oncology, University of Cambridge, Cambridge, United Kingdom*
[2]*Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD*
[3]*Centre for Molecular Environmental Genetic and Analytic Epidemiology, The University of Melbourne, Melbourne, Australia*
[4]*Eastern Cancer Registration and Information Centre, Cambridge, United Kingdom*
[5]*Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom*

The mismatch repair (MMR) genes are involved in the maintenance of genomic integrity. Recently, we showed that common variants in these genes are unlikely to contribute significantly to colorectal cancer risk. The aim of this study was to investigate the role of common variants in the mismatch repair pathway as prognostic markers in colorectal cancer patients. We genotyped 2,060 patients for 68 SNPs in 7 mismatch repair genes (*MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH6*, *PMS1* and *PMS2*), using a single nucleotide polymorphism (SNP) tagging approach. Genotypes at the tag SNPs and multi-SNP haplotypes were tested for association with overall survival (OS) and disease specific survival (DSS) using a Cox regression model. Eight SNPs and 10 haplotypes were significant at a nominal $p < 0.05$ in the univariate analyses. Stepwise analysis showed that haplotype effects were mainly due to associated SNPs carried by these haplotypes. After adjustment for sex, age at diagnosis and stage when using overall survival and stage only when using disease specific survival, prognostic values were unattenuated. The most significant SNP associated with disease specific survival after adjustment was rs863221, located in *MSH3* (HR: 0.59, 95% confidence interval (CI) 0.42–0.82, *p*-value: 0.001). In conclusion, we find some evidence that common variants in mismatch repair genes may contribute to survival of patients with colorectal cancer.
© *2008 Wiley-Liss, Inc.*

Key words: colorectal cancer; single nucleotide polymorphism; haplotype; survival

Colorectal cancer (CRC) is the third most common cancer in England after breast and lung cancers. Each year, 36,000 new cases are diagnosed and 16,000 deaths from colorectal cancer are reported in the UK.[1] Several factors are known to influence survival after diagnosis, but the only routinely used prognostic marker is clinical stage which amalgamates depth of tumour invasion, nodal status and distant metastasis. Other factors thought to influence prognosis include lifestyle,[2,3] systemic inflammatory response to the tumour[4] and the immunologic microenvironment of the tumour.[5,6] Somatic genetic changes in the tumour, such as microsatellite instability, have also been found to influence clinical course of disease, although the results of different studies of specific markers have been contradictory.[7,8] More recently, germline genetic variation/molecular markers have been associated with prognosis, including variants in the thymidilate synthase gene (particularly for patients receiving adjuvant therapy with 5-fluorouracil),[9,10] and VEGF polymorphisms.[11]

The DNA mismatch repair (MMR) pathway has been extensively studied in colorectal cancer. The 7 genes (*MSH2*, *MSH3*, *MSH6*, *MLH1*, *MLH3*, *PMS1* and *PMS2*) in this pathway are involved in the maintenance of the genome integrity as well as the reparation of damages. Rare, deleterious mutations in these genes lead to hereditary non-polyposis coli cancer (HNPCC), but this autosomal dominant syndrome causes less than 5% of all colorectal cancers. Recent results from genome wide association studies have found several polymorphisms that are associated with colorectal cancer risk,[12–15] but common variants in mismatch repair genes do not seem to play a significant role.[16–18]

Nevertheless, it is plausible that germline variation in MMR genes influence prognosis, as treatments such as radiotherapy and chemotherapy damage DNA of cancer cells leading to apoptosis. In 2006, Barnetson *et al.* found no difference in survival between carriers and non-carriers of pathological mutations in *MLH1*, *MSH2* and *MSH6*.[19] However, there are no published data assessing the effect of common variants in MMR on survival after diagnosis. The aim of our study was to investigate the role of common genetic variation in the mismatch repair pathway as prognostic markers in colorectal cancer.

Material and methods

Study subjects

Two thousand and sixty cases of invasive colorectal cancer or anal adenocarcinoma from SEARCH were included in these analyses. The SEARCH colorectal study is an ongoing study of colorectal cancer in the region served by the Eastern Cancer Registration and Information Centre (ECRIC, formerly East Anglian Cancer Registry). Eligible cases are all those registered by ECRIC since 2000 who were aged 18–69 at diagnosis. All participants in the study gave informed consent, provided a blood sample and completed a comprehensive epidemiologic questionnaire. The patients are flagged by the cancer registry for death notification. Cause of death was obtained from the death certificate. This study is approved by the Eastern Multi-Centre Research Ethics Committee (Eastern MREC).

Tag SNP selection

A set of 75 SNPs was selected to tag the known common variation in 7 genes in the MMR pathway (*MSH2*, *MSH3*, *MSH6*, *MLH1*, *MLH3*, *PMS1* and *PMS2*) as previously described.[18] Assays were successfully designed for 68, and these tagged 323 of the 391 common variants in these genes (Table I) with $r^2 > 0.8$ (mean $r^2 = 0.94$), where r is the pairwise correlation between SNPs. Eleven SNP were tagged by a specific, multi-marker haplotype of 2 or 3 tSNPs (Supp. Info. Table I).

Additional Supporting Information may be found in the online version of this article.

Abbreviations: CRC, colorectal cancer; DF, degree of freedom; DSS, disease specific survival; HWE, Hardy-Weinberg equilibrium; LCL, lower confidence limit; MMR, mismatch repair; OS, overall survival; SNP, single nucleotide polymorphism; UCL, upper confidence limit.

Grant sponsors: Cancer Research, United Kingdom; Foundation Dr. Henri Dubois-Ferriere Dinu Lipatti; National Cancer Institute.

***Correspondence to:** Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, United Kingdom. Fax: +44-1223-740-147. E-mail: Thibaud@srl.cam.ac.uk

Received 11 August 2008; Revised 25 September 2008; Accepted after revision 14 October 2008
DOI 10.1002/ijc.24120
Published online 7 November 2008 in Wiley InterScience (www.interscience.wiley.com).

TABLE I – EFFICIENCY OF TAG SNP SELECTION FOR EACH GENE

Gene	Chromosome position	Region covered (kb)	Common variants	No of tSNPs	No tSNPs genotyped	SNPs tagged by specific haplotype	Proportion variants tagged with $r^2 > 0.8$
MSH6	2p16	33.81	23	12	11	None	96%
MSH2	2p22-p21	90.1	50	14	12	None	96%
PMS1	2q31.1	103.25	39	9	9	1	97%
MLH1	3p21.3	67.36	29	6	4	1	89%
MSH3	5q11-q12	232.3	226	25	24	9	94%
PMS2	7p22.2	45.84	10	7	6	None	90%
MLH3	14q24.3	44.83	14	2	2	None	100%

TABLE II – CHARACTERISTICS OF COLORECTAL CANCER PATIENTS

	No of patients (N = 2,060)	
Cases characteristics	No.	%
Age at diagnosis		
<50	223	11
50–60	695	34
>60–<70	1,142	55
Grade		
Total known	1,826	89
Unknown	234	11
I	149	8
II	1,422	78
III	255	14
Stage		
Total known	2,021	98
Unknown	39	2
I	434	22
II	897	44
III	613	30
IV	77	4
Localisation		
Colon	1,164	58
Recto sigmoid	181	9
Rectum	713	35
Anus	2	0.1
Treatments		
Chemotherapy		
Yes	693	34
No	1,216	66
Radiotherapy		
Yes	241	12
No	1,668	88
Surgery		
Yes	1,887	99
No	22	1
5-year survival (all cause mortality)		
Rate	83	
95% CI	78–86	
5-year survival (cause specific mortality)		
Rate	86	
95% CI	82–89	

Genotyping

All samples were genotyped using the Taqman 7900HT Sequence Detection System according to the manufacturer's instructions. Each assay was carried out using 10 ng genomic DNA in a 5 μL reaction using Taqman Universal PCR Master Mix (Applied Biosystems, Warrington United Kingdom), forward and reverse primers and FAM- and VIC-labelled probes designed by Applied Biosystems (ABI Assay-by-Design). Primer and probe sequences and assay conditions used for each polymorphism analysed are available from the corresponding author on request. All assays were carried out in 384-well arrays with 12 duplicate samples in each plate for quality control. Where discordant genotypes were observed in duplicates, the genotyping was repeated. Genotypes were determined using Allelic Discrimination Sequence Detection software (Applied Biosystems). DNA samples that did not give a clear genotype result at the first attempt were not repeated. Hence, there are variations in the number of samples successfully genotyped for each polymorphism. Call rates ranged from 98.4% to 99.8% for all the SNPs and overall concordance between duplicate samples was 100%. DNA was extracted from blood samples by Whatman International (Ely, UK) using a chloroform/phenol method.

Survival analysis

Each SNP was assessed for deviation of the genotype frequencies from those expected under Hardy-Weinberg equilibrium (HWE) using a χ^2 test [1 degree of freedom (df)]. The primary test for association between SNPs or multi-SNP haplotypes and the outcome (all cause mortality or death from colorectal cancer) was assessed using Cox regression with number of rare-alleles carried as the independent variable (trend test under the log-additive genetic model). Patients were recruited at a variable time after diagnosis so analyses were conducted allowing for left truncated data. Time at risk began on date of diagnosis, but time under observation began at the date of blood draw and ended at the date of death from any cause, or, if death did not occur, on September 30, 2007. This generates an unbiased estimate of the hazard ratio provided the proportional hazard assumption is correct.[20] Even if this assumption is violated the procedure provides a valid test of association.[21] Follow-up for all patients was censored at 15 years after diagnosis. SNPs with p-value < 0.05 using the test for trend were then also evaluated under dominant and recessive genetic models and the best fitting model determined using the log likelihoods.

Other potential prognostic factors evaluated were: age at diagnosis (18–69 years old), sex (male or female), tumour site (colon or other localisations), clinical stage (TNM I, II, III and IV), histopathological grade (well differentiated, moderately differentiated and poorly differentiated) (Table II). Age and stage distribution for all cases diagnosed in the study area can be seen in Supporting Information Table II.

SNPs and multi-SNP haplotypes that were significant in the univariate analyses were re-evaluated in multivariate models with overall survival (OS) and disease specific survival (DSS). The multivariate model included all prognostic factors with p-value $<$ 0.05 in univariate analyses. Multivariate analyses were performed using Cox proportional hazard model. Proportional hazards assumptions were verified with a non-zero slope test of the Schoenfeld residuals.[22] Covariates that violated the proportional hazard assumption were treated as time varying covariates allowing the coefficients ($\log_e$ hazard ratio) to vary with the log of time (See Appendix for formula). When a gene or haplotype block contained multiple significant SNPs or haplotypes, we performed a backward stepwise procedure to determine the best fitting model. Sub-group hazard ratios were calculated for each category of stage, grade, location, chemotherapy and radiotherapy for SNPs and haplotypes significant after the multivariate analysis. Heterogeneity between categories was assessed using likelihood ratio test. All analyses were performed with Intercooled Stata version 8 (STATA Corp, College Station, TX).

Haplotype analysis

In addition to the specific SNP-tagging multi-marker haplotypes, we carried out a general test for association for all common haplotypes within haplotype blocks. These haplotypes may improve the tagging of SNPs that were otherwise poorly tagged

TABLE III – SNPs, SNP TAGGING HAPLOTYPES, AND HAPLOTYPES ASSOCIATED WITH OUTCOME IN UNIVARIATE ANALYSIS ($p < 0.05$)

Gene	Block	SNP	Best fitting genetic model	Overall survival			Disease specific survival		
				HR	95% CI	*p*-value	HR	95% CI	*p*-value
MSH2		rs4638843	Dominant	1.41	1.08–1.83	0.01	1.48	1.04–2.11	0.03
MSH3	MSH3 B1	rs40139	Co-Dominant	1.03	0.87–1.23	0.69	1.35	1.07–1.69	0.01
MSH3	MSH3 B1	rs863221	Dominant	0.86	0.67–1.10	0.22	0.57	0.41–0.78	0.001
MSH3	MSH3 B1	rs836805	Dominant	0.88	0.69–1.12	0.29	0.62	0.44–0.86	0.005
MSH3	MSH3 B2	rs26779	Dominant	0.89	0.69–1.13	0.35	0.60	0.43–0.83	0.002
MSH3	MSH3 B2	rs33015	Recessive	1.59	1.14–2.20	0.005	2.11	1.41–3.16	0.0007
MSH3	MSH3 B3	rs27385	Co-Dominant	0.66	0.45–0.95	0.03	0.86	0.55–1.33	0.5
MSH6	MSH6 B1	rs3136245	Recessive	1.89	1.19–2.99	0.006	2.56	1.50–4.38	0.001
Gene	Block	SNP tagging haplotype	Model	HR	95% CI	*p*-value	HR	95% CI	*p*-value
MSH3		h001	Co-Dominant	1.03	0.87–1.22	0.75	1.32	1.05–1.66	0.02
Gene	Block	Haplotype	Model	HR	95% CI	*p*-value	HR	95% CI	*p*-value
MSH2		h000000001100	Co-Dominant	1.29	1.02–1.64	0.04	1.25	0.91–1.74	0.17
MSH3	MSH3 B1	h001000000	Co-Dominant	1.03	0.84–1.25	0.79	1.35	1.05–1.74	0.02
MSH3	MSH3 B2	h0000100	Co-Dominant	1.33	0.95–1.87	0.11	2.04	1.39–2.99	0.0009
MSH3	MSH3 B3	h00100000	Co-Dominant	0.67	0.46–0.97	0.04	0.87	0.56–1.37	0.55
MSH3	MSH3 B3	h10000001	Co-Dominant	1.30	0.94–1.78	0.12	1.61	1.09–2.37	0.02
MSH6	MSH6 B1	h10000	Co-Dominant	1.47	1.07–2.01	0.03	1.59	1.07–2.38	0.02
MLH1		h0000	Co-Dominant	0.83	0.58–1.21	0.34	0.43	0.22–0.85	0.02
MLH3		h00	Co-Dominant	1.53	1.14–2.04	0.01	1.35	0.89–2.04	0.15
PMS1		h011000111	Co-Dominant	1.29	0.88–1.89	0.19	1.63	1.03–2.58	0.04
PMS2		h110100	Co-Dominant	0.75	0.56–0.98	0.04	0.75	0.52–1.08	0.12

HR, Hazard ratio; MSH3 B1, MSH3 block 1; MSH3 B2, MSH3 block 2; MSH3 B3, MSH3 block 3; MSH6 B1, MSH6 block 1.

and may even tag some rare variants in the genes. Furthermore, it is plausible that there are true cis-effects between SNPs such that a specific haplotype has a causal association.

Haplotype blocks were defined such that the common haplotypes (>5% frequency) accounted for at least 80% of the haplotype diversity. *MLH1*, *MLH3*, *PMS1*, *PMS2* and *MSH2* comprised a single haplotype block, *MSH6* had 2 blocks (MSH6 B1 and MSH6 B2) and *MSH3* was divided into 3 blocks (MSH3 B1, MSH3 B2 and MSH3 B3). For both specific haplotype marker tests and the general comparison of haplotype frequencies by haplotype block, haplotype frequencies and subject-specific expected haplotype indicators were calculated using an in-house programme that implements an expectation substitution approach to account for haplotype uncertainty given unphased genotype data. Subjects missing >50% genotype data in each block were excluded from haplotype analysis. Rare haplotypes (<2% frequency) were pooled. Specific SNP-tagging haplotypes (Supp. Info. Table I) and global haplotypes (Supp. Info. Table III) were also assessed in univariate analyses. Supporting Information Table IV presents the SNPs relative position for the 10 haplotypes found significant in the univariate analyses.

Admixture maximum likelihood test

We used the admixture maximum likelihood (AML) test[23] as a single experiment-wise test. In brief, the AML method formulates the alternative hypothesis in terms of the proportion of SNPs (α) that are associated with disease and the effect size when an association exists (η). The parameters α and η can be estimated by maximum likelihood and the test statistic derived as a likelihood ratio test. The significance of this statistic was assessed using a permutation test.

Results

Characteristics of colorectal cancer patients

Clinical and pathological data for the 2,060 patients (1,187 males (58%) and 873 women (42%)) are presented in the Table II. During the 8,572 person-years at risk, 265 people died (all cause mortality) with 147 deaths from colorectal cancer (disease specific mortality). The 5-year survival rate was 83% (95% CI 86–78%) for all cause mortality and 86% (95% CI 82–89%) for disease specific mortality. The median time between diagnosis and entry in the study was 3.7 years (range, 0.0–14.9). The median age at recruitment was 61 years old (range, 19–69 years). The tumour characteristics as well as treatments of patients in the study are described in Table II.

Survival analysis

Tables III–IV show only results for significant associations—the results for all analyses can be found in Supporting Information Tables I–VII. Genotype frequencies for all 68 tSNPs were consistent with Hardy-Weinberg equilibrium. The results of the univariate Cox regression analyses are listed in Table III. None of the SNPs in *MHL1*, *MLH3*, *PMS*1 and *PMS2* was associated with either OS or DSS. Eight SNPs (Table III), on 3 different genes (*MSH2*, *MSH3* and *MSH6*) were associated with OS and/or DSS. Rs4638843 (*MSH2*) is associated with OS and DSS under a dominant model. Rs3136245 in *MSH6* block one (MSH6 B1) was associated with OS and DSS under a recessive model. *MSH3* has 6 SNPs associated with outcome: 3 in one haplotype block (rs40139, rs863221 and rs836805), 2 in a second block (rs26779 and rs33015) and 1 in another block (rs27385). Eleven SNPs were tagged by specific multi-marker haplotypes but only one of these—a triplet of SNPs (multi-marker haplotype h001: rs245408, rs184967 and rs40139) tagging rs32950 and rs2035256—was significantly associated with DSS (HR 1.32, 95% CI 1.05–1.66, *P*-value = 0.02).

Seventy-five haplotypes in 10 haplotype blocks were independently analysed for association with survival. Ten haplotypes, on 7 different genes: *MLH1, MLH3, MSH2, MSH3, MSH6, PMS2* and *PMS1* were found significant at a 5% level for OS and/or with DSS (Table III).

To determine whether multiple associated SNPs and haplotypes from the same block were all reporting the same association due to linkage disequilibrium, we performed a stepwise procedure for the 5 groups of significant SNPs and haplotypes which are in the same gene/haplotype block: MSH2, MSH3 B1, MSH3 B2, MSH3 B3 and MSH6 (Supp. Info. Table IV). Rs4638843 was sufficient

TABLE IV – SNPs AND HAPLOTYPES ASSOCIATED WITH OUTCOME IN MULTIVARIATE ANALYSIS ($p < 0.05$)

Gene	Block	SNP	Model	Multivariate analysis					
				Overall survival			Disease specific survival		
				HR[1]	95% CI	*p*-Value	HR[2]	95% CI	*p*-Value
MSH2		rs4638843	Dominant	1.35	1.03–1.77	0.03	1.36	0.96–1.94	0.08
MSH3	MSH3 B1	rs863221	Dominant	0.83	0.64–1.06	0.14	0.59	0.42–0.82	0.001
MSH3	MSH3 B2	rs33015	Recessive	1.44	1.03–2.01	0.03	1.72	1.15–2.59	0.009
MSH3	MSH3 B3	rs27385	Co-Dominant	0.67	0.45–0.98	0.04	0.93	0.6–1.45	0.80
MSH6	MSH6 B1	rs3136245	Recessive	1.62	1.02–2.56	0.04	2.03	1.18–3.48	0.01
Gene	Block	Haplotypes	Model	HR[1]	95% CI	*p*-Value	HR[2]	95% CI	*p*-Value
MSH3	MSH3 B2	h0000100	Co-Dominant	1.36	0.96–1.92	0.08	1.89	1.28–2.80	0.001
MSH3	MSH3 B3	h10000001	Co-Dominant	1.29	0.93–1.78	0.125	1.62	1.10–2.39	0.02

[1]Hazard ratio adjusted for sex, age at diagnosis, stage; with stage and diagnose age as time variable covariate–[2]Hazard ratio adjusted for stage; with stage as time variable covariate.

to explain all the association for MSH2 with OS, rs33015 and rs27385 explained the associations in MSH3 blocks 2 and 3, and rs3136245 explained the effect observed for MSH6 block 1. The two associations in MSH3 were independent (rs33015 and rs27385 are 100 kb apart with $r^2 = 0.02$). Similarly, when using DSS as end point, rs4638843, rs863221 and rs3136245 explained the effect observed on MSH2, MSH3 B1 and MSH6 B1. In contrast to the result for OS, the associations for MSH3 blocks 2 and 3 were best explained by the haplotypes h0000100 and h10000001 (Table IV).

We then assessed whether the genetic associations were affected when allowing for the effects of other covariates. Of these, age at diagnosis, sex and stage were significantly associated with OS, and only stage was associated with DSS. This difference might be expected as age and sex are important determinants of competing causes of mortality. Tumour grade and localisation were not associated with either OS or DSS. Analysis of Schoenfeld residuals showed that stage and age at diagnosis violated the proportional hazard assumption (data not shown). We therefore corrected for this by treating them as time varying covariates in which the $\log_e$ hazard ratio varies as a linear function of $\log_e$ time.

Table IV reports multivariate analyses for the 5 SNPs and 2 haplotypes resulting from the backward stepwise procedure. The multivariate model was adjusted for sex, stage and age at diagnosis when OS was the outcome and with stage only when DSS was the end point. Compare with univariate analyses, hazard estimates were not attenuated after adjustment using OS or DSS as end point and all but rs4638843 (*MSH2*) remain significant (Table IV) (Supp. Info. Table VI). We then compared the effect on prognosis of 5 SNPs and 2 haplotypes in each category of stage, grade, location, chemotherapy and radiotherapy. There were no large differences in the sub-group specific hazard ratios, but differences of borderline statistical significance were seen for stage and rs3136245, chemotherapy and rs33015 and rs3136245, location and rs3136245 and radiotherapy and rs27385 (Supp. Info. Table VII).

The AML global test of association was not significant using heterogeneity test or trend test for overall survival or cause specific survival (OS P-Het = 0.48 and *P*-Trend = 0.87, DSS P-Het = 0.07 and *P*-Trend = 0.08).

Discussion

Recently, we have found little evidence that common variants in the mismatch repair genes are associated with colorectal cancer susceptibility.[18] However, no published studies have systematically evaluated the prognostic value of common variants in the mismatch repair pathway in patients diagnosed with colorectal cancer. We have therefore assessed association between 68 tag SNPs in 7 MMR genes with survival after colorectal cancer diagnosis. The strength of our study is its large size, the duration of follow up and a systematic approach to tag all known common variants in the mismatch repair genes.

We have found no evidence that common variants in *MLH1*, *MLH3*, *PMS1*and *PMS2* are associated with survival after diagnosis of colorectal cancer. Most of the significant SNPs and haplotypes are in genes encode proteins in the MutSα or MutSβ complexes. These are involved in the recognition of the defect after an unsatisfactory DNA replication.[24] Power to detect association in these genes is limited by sample size—despite a total sample size of over 2,000 there were only 265 deaths from all causes and 147 deaths from colorectal cancer. This provides reasonable power to detect alleles with moderate risks, but the power to detect small genetic effects is limited. For example, power to detect an allele of frequency 30% that confers an all cause mortality relative hazard of 1.4 is greater than 80% at a type I error rate of 0.01, whereas power to detect a hazard ratio of 1.2 for the same allele would be just 25%. Power to detect effects on cause specific mortality is less. Furthermore, the panel of tag SNPs used in this study does not capture all the common variation in the genes of interest and it is possible that some of the known SNPs that were poorly tagged or unknown common variants in these genes are associated with prognosis.

Eight SNPs on three different genes (*MSH2*, *MSH3* and *MSH6*) were associated with OS or DSS in the univariate analysis at a nominal level of significance ($p < 0.05$). The effect was not attenuated after adjustment suggesting that the mechanisms of action of these variants (if true positive associations) are independent of tumour stage. However, none of the associations were highly significant, and these associations are still most likely to represent false positives. Where the prior probability of association is low, as here, highly stringent levels of significance are required.

The SEARCH case cohort is likely to be biased towards those with a better survival as patients with significant co-morbidity are more likely to be excluded by their general practitioner or to be non-responders. This is reflected in the fact that the estimated 5-year survival for the SEARCH cohort, after allowing for left truncation, is ∼75% compared with 57% for all cases of colorectal cancer aged under 70 diagnosed in East Anglia since 2000. Nevertheless, this will not affect the internal consistency of the study and the method of analysis provides unbiased estimates of the hazard ratios. The inclusion of prevalent cases should not affect the hazard estimates as we allowed for this left truncation in the analysis.

In summary, we have found some evidence for association of variants in mismatch repair genes and overall survival as well as disease specific survival. None were highly significant and it is possible that they represent false positives. However, as these associations are biologically plausible, they warrant replication in independent studies.

Acknowledgements

The authors thank all patients who participated in this study, the Eastern Cancer Registration and Information Centre, and the SEARCH team. TK is funded by the Foundation Dr Henri Dubois-Ferriere Dinu Lipatti. EMA is supported by the National Cancer Institute and the NIH-Cambridge Graduate Partnership Program. R.J.M. is a Sidney Sax Post Doctoral Research Fellow of the NHMRC (Australia). DFE is a Principal Fellow, PDPP is a Senior Clinical Research Fellow of Cancer Research United Kingdom.

References

1. Cancer Research, UK. Available at: http://info.cancerresearchuk.org.
2. Haydon AM, MacInnis RJ, English DR, Giles GG. Effect of physical activity and body size on survival after diagnosis with colorectal cancer. Gut 2006;55:62–7.
3. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ 2007;335:1134.
4. Leitch EF, Chakrabarti M, Crozier JE, McKee RF, Anderson JH, Horgan PG, McMillan DC. Comparison of the prognostic value of selected markers of the systemic inflammatory response in patients with colorectal cancer. Br J Cancer 2007;97:1266–70.
5. Galon J, Fridman WH, Pages F. The adaptive immunologic microenvironment in colorectal cancer: a novel perspective. Cancer Res 2007;67:1883–6.
6. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoue F, Bruneval P, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 2006;313:1960–4.
7. Kim GP, Colangelo LH, Wieand HS, Paik S, Kirsch IR, Wolmark N, Allegra CJ. Prognostic and predictive roles of high-degree microsatellite instability in colon cancer: a National Cancer Institute-National Surgical Adjuvant Breast and Bowel Project Collaborative Study. J Clin Oncol 2007;25:767–72.
8. Benatti P, Gafa R, Barana D, Marino M, Scarselli A, Pedroni M, Maestri I, Guerzoni L, Roncucci L, Menigatti M, Roncari B, Maffei S, et al. Microsatellite instability and colorectal cancer prognosis. Clin Cancer Res 2005;11:8332–40.
9. Dotor E, Cuatrecases M, Martinez-Iniesta M, Navarro M, Vilardell F, Guino E, Pareja L, Figueras A, Mollevi DG, Serrano T, de Oca J, Peinado MA, et al. Tumor thymidylate synthase 1494del6 genotype as a prognostic factor in colorectal cancer patients receiving fluorouracil-based adjuvant treatment. J Clin Oncol 2006;24:1603–11.
10. 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;112:733–7.
11. Kim JG, Chae YS, Sohn SK, Cho YY, Moon JH, Park JY, Jeon SW, Lee IT, Choi GS, Jun SH. Vascular endothelial growth factor gene polymorphisms associated with prognosis for patients with colorectal cancer. Clin Cancer Res 2008;14:62–6.
12. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 2007;39:989–94.
13. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 2007;39:984–8.
14. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007;39:1315–17.
15. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, Walther A, Spain S, Pittman A, Kemp Z, Sullivan K, Heinimann K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 2008;40:26–8.
16. Schafmayer C, Buch S, Egberts JH, Franke A, Brosch M, El Sharawy A, Conring M, Koschnick M, Schwiedernoch S, Katalinic A, Kremer B, Folsch UR, et al. Genetic investigation of DNA-repair pathway genes PMS2. MLH1, MSH2, MSH6, MUTYH, OGG1 and MTH1 in sporadic colon cancer. Int J Cancer 2007;121:555–8.
17. Berndt SI, Platz EA, Fallin MD, Thuita LW, Hoffman SC, Helzlsouer KJ. Mismatch repair polymorphisms and the risk of colorectal cancer. Int J Cancer 2007;120:1548–54.
18. Koessler T, Oestergaard MZ, Song H, Tyrer J, Perkins B, Dunning A, Easton D, Pharoah PP. Common variants in mismatch repair genes and risk of colorectal cancer. Gut 2008;57:1097–101.
19. Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354:2751–63.
20. Brookmeyer R. Biased sampling of cohorts. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: Wiley, 2005.
21. Cnaan A, Ryan L. Survival analysis in natural history studies of disease. Stat Med 1989;8:1255–68.
22. Hess KR. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. Stat Med 1995;14:1707–23.
23. Tyrer J, Pharoah PD, Easton DF. The admixture maximum likelihood test: a novel experiment-wise test of association between disease and multiple SNPs. Genet Epidemiol 2006;30:636–43.
24. Jiricny J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 2006;7:335–46.

Appendix

Cox extended model formula for overall survival:

$$h_i(t|x_i) = h_0(t)\exp\,(\beta_1\text{ stage} + \beta_2\text{Age at diagnosis} + \beta_3\text{sex} + \ln\,(t)(\beta_4\text{stage} + \beta_5\text{Age at diagnosis}))$$

with $\beta_n = \ln$ (HR β_n) and t follow-up time in years (range: 0–14.9).

Example for SNP rs4638843:

$$h_i(t|x_i) = 1.35\ \exp\,(1.99 - 0.05 + 0.35 + \ln\,(t)\,(-0.84 + 0.06)).$$