available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Common variants in human CRC genes as low-risk alleles

Simone Picelli [a], *Pawel Zajac* [b], *Xiao-Lei Zhou* [a], *David Edler* [c], *Claes Lenander* [d], *Johan Dalén* [e], *Fredrik Hjern* [f], *Nils Lundqvist* [g], *Ulrik Lindforss* [h], *Lars Påhlman* [i], *Kennet Smedh* [j], *Anders Törnqvist* [k], *Jörn Holm* [l], *Martin Janson* [m], *Magnus Andersson* [n], *Susanne Ekelund* [o], *Louise Olsson* [p], *Joakim Lundeberg* [b], *Annika Lindblom* [a,*]

[a] *Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden*
[b] *Department of Gene Technology, School of Biotechnology, Royal Institute of Technology (KTH), Stockholm, Sweden*
[c] *Department of Surgery, Karolinska University Hospital, Solna, Sweden*
[d] *Centre of Gastrointestinal Disease, Ersta Hospital, Stockholm, Sweden*
[e] *Department of Surgery, St. Görans Hospital, Stockholm, Sweden*
[f] *Department of Clinical Science, Division of Surgery, Karolinska Institutet at Danderyd Hospital, Danderyd, Sweden*
[g] *Norrtälje Hospital, Norrtälje, Sweden*
[h] *Södertälje Hospital, Södertälje, Sweden*
[i] *Department of Surgery, Uppsala University Hospital, Uppsala, Sweden*
[j] *Department of Surgery, Colorectal Unit, Central Hospital, Vasterås, Sweden*
[k] *Central Hospital, Karlstad, Sweden*
[l] *Länssjukhuset Gävle-Sandviken, Gävle, Sweden*
[m] *Department of Medical & Surgical Gastroenterology, Karolinska University Hospital, Huddinge, Sweden*
[n] *Department of Surgery, Örebro University Hospital, Örebro, Sweden*
[o] *Department of Surgery, South Stockholm General Hospital, Stockholm, Sweden*
[p] *Mälarsjukhuset, Eskilstuna, Sweden*

ARTICLE INFO

Article history:
Received 23 September 2009
Accepted 14 January 2010
Available online 9 February 2010

Keywords:
Colorectal cancer
Mismatch repair genes
SNPs
Low-penetrance alleles
Polygenic model

ABSTRACT

The genetic susceptibility to colorectal cancer (CRC) has been estimated to be around 35% and yet high-penetrance germline mutations found so far explain less than 5% of all cases. Much of the remaining variations could be due to the co-inheritance of multiple low penetrant variants. The identification of all the susceptibility alleles could have public health relevance in the near future. To test the hypothesis that what are considered polymorphisms in human CRC genes could constitute low-risk alleles, we selected eight common SNPs for a pilot association study in 1785 cases and 1722 controls. One SNP, rs3219489:G>C (MUTYH Q324H) seemed to confer an increased risk of rectal cancer in homozygous status (OR = 1.52; CI = 1.06–2.17). When the analysis was restricted to our 'super-controls', healthy individuals with no family history for cancer, also rs1799977:A>G (MLH1 I219V) was associated with an increased risk in both colon and rectum patients with an odds ratio of 1.28 (CI = 1.02–1.60) and 1.34 (CI = 1.05–1.72), respectively (under the dominant model); while 2 SNPs, rs1800932:A>G (MSH6 P92P) and rs459552:T>A (APC D1822V) seemed to confer a protective effect. The latter, in particular showed an odds ratio of 0.76 (CI = 0.60–0.97) among colon patients and 0.73 (CI = 0.56–0.95) among rectal patients. In conclusion, our study suggests that common variants in human CRC genes could constitute low-risk alleles.

* *Corresponding author:* Address: Department of Molecular Medicine and Surgery L8:02, Karolinska Institutet, Stockholm, Sweden. Tel.: +46 08 51775248; fax: +46 08 51773620.
E-mail address: Annika.Lindblom@ki.se (A. Lindblom).

0959-8049/$ - see front matter
doi:10.1016/j.ejca.2010.01.013

1. Introduction

Colorectal cancer (CRC) is a complex disease caused by genetic and environmental factors, with the former accounting for a third of the cases and the latter for the remaining two-thirds.[1] Being able to divide the population into risk categories would allow tailored prevention programmes according to the risk of each individual. In order to do so, it is important to understand how the genetic background is affected by environmental factors and how this complex interplay of genetic and non-genetic factors contributes to the CRC development and progression.[2,3]

Several familial syndromes are known: Familial Adenomatous Polyposis (FAP), caused by a germline mutation in APC[4,5] and Hereditary Non-Polyposis Colorectal Carcinoma (HNPCC), caused by a germline mutation in one of the mismatch repair genes (MMR) MLH1, MSH2, MSH6 and PMS2.[6–9] However, they together account for only approximately 5% of the cases.[10] MUTYH Polyposis (MAP) is usually phenotypically identical to a classic or mild form of FAP, but the difference stands in its recessive mode of inheritance.[11]

In the past 2 years Genome-Wide Association Studies (GWAS) have led to the discovery of several susceptibility alleles.[12–18] However, the risk associated with each variant is very low and these SNPs are suggested to explain only a small part of the remaining unknown genetic contribution.[19]

Similarly, further high-penetrance alleles are probably rare and restricted to families with multiple cases. Several genetic models have been proposed to explain the aetiology of CRC, ranging from few common alleles conferring a modest risk (the so-called 'common disease-common variant hypothesis') to a very large number of alleles conferring a higher risk (the 'common disease-rare variant hypothesis').[20,21]

Our hypothesis is that variants in genes already known to be involved in CRC development could explain at least part of the sporadic cases without family history and clearly pathogenic mutations and, perhaps, also partly some familial cases. These variants could act as low-penetrance alleles and it seems reasonable to assume that several of them together could drive the cell towards the tumourigenic process.

In the past years many SNPs of unknown pathogenicity were found in the MMR genes, APC and MUTYH by our group and in many other laboratories around the world.[22–24] In order to test our hypothesis we conducted a pilot study on eight of the most common SNPs found in our studies in the Swedish population: one in MLH1, one in APC, two in MUTYH and four in MSH6.

These variants are considered polymorphisms but could nonetheless act as low-risk alleles in CRC.

2. Materials and methods

The case cohort used in this study was composed of 1785 consecutive colorectal cancer patients of Swedish origin divided as follows: 1103 individuals with a diagnosis of colon cancer, 637 with a diagnosis of rectal cancer and 45 individuals whose exact tumour location was not specified. All cases were collected through the Family Cancer Clinic at the Karolinska Hospital (Stockholm, Sweden) and were recruited by 14 different hospitals from central Sweden. The control cohort was composed of 1306 blood donors from the general population between the age of 18 and 65 and 416 so-called 'super-controls'. The super-controls are unaffected spouses of CRC patients, which are cancer-free at the moment of the diagnosis and do not have a family history of any type of cancer. In principle, the super-controls constitute a better cohort to be used for these types of case–control studies, since they should carry even less susceptibility factors than blood donors, who we consider to represent the general population. We investigated eight SNPs in four different genes (Table 1): rs459552:T>A (APC D1822V), rs1799977:A>G (MLH1 I219V), rs1042821:G>A (MSH6 G39E), rs1800932:A>G (MSH6 P92P), rs1800935:T>C (MSH6 D180D), rs1800937:C>T (MSH6 Y214Y), rs3219484:G>A (MUTYH V22M) and rs3219489:G>C (MUTYH Q324H).

Informed consent was obtained from all participants. The study was undertaken in accordance with the Swedish legislation of ethical permission (02/489). Genomic DNA was extracted from peripheral blood by standard procedures. Genotyping and a first-quality check of rs459552:T>A (RefSeq NM_000038.4), rs1799977:A>G (RefSeq NM_000249.2), rs1800932:A>G and rs1800935:T>C (RefSeq NM_000179.2), rs3219484:G>A and rs3219489:G>C (RefSeq NM_001048171.1) were done at deCode Genetics (Reykjavik, Iceland). The remaining two SNPs in the MSH6 gene, rs1042821 and rs1800937 (RefSeq NM_000179.2) were genotyped using TaqMan SNP Genotyping Assay (Applied Biosystems, Foster City, CA). Deviations of the genotype frequency in cases and controls from those expected under Hardy–Weinberg equilibrium were calculated by χ^2 tests (1 degree of freedom). Allelic frequencies of the SNPs in the case and control groups were compared using a χ^2 test (allele 1 [common] versus allele 2 [minor]). Analyses were also performed under various types of genetic contrasts including the contrast of homozygotes (genotype 11 versus genotype 22), the dominant (genotype 11 versus genotype [12 + 22]) and recessive [genotype (11 + 12) versus genotype 22] models. In addition, Armitage's trend test, which takes into account the individuals' genotypes rather than just alleles,[25] was performed using the DeFinetti program provided as an online source (http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl). The significance level for statistical tests was set at 0.05. Odds ratios (ORs), their 95% confidence intervals (CIs) and their corresponding p-values were calculated using the same program. These analyses were performed on the entire cohort of CRC cases as well as for colon and rectum only.

3. Results

The results of the case–control association study in the whole cohort as well as the analyses stratified by tumour location are reported in Table 2.

APC D1822V did not show any difference between cases and controls, but when we considered only the super-controls (n = 344), the heterozygous carrier status was associated with an odds ratio of 0.77 (CI = 0.60–0.97, p = 0.03) and the homozygous status with an odds ratio of 0.67 (CI = 0.43–1.05; p = 0.08). If a dominant model was assumed, the combined

Table 1 – Variants analysed in the present study.

rs no. dbSNP	Gene	Nucleotide position	Exon	Structural alteration	MAF CEU[a] dbSNP
rs459552	APC[b]	c.5465T>T/A	16	D1822V	T = 0.775; A = 0.225
rs1799977	MLH1[c]	c.655A>A/G	8	I219V	A = 0.667; G = 0.333
rs1042821	MSH6[d]	c.116G>G/A	1	G39E	G = 0.793; A = 0.207
rs1800932	MSH6[d]	c.276A>A/G	2	P92P	A = 0.775; G = 0.225
rs1800935	MSH6[d]	c.540T>T/C	3	D180D	T = 0.658; C = 0.342
rs1800937	MSH6[d]	c.642C>C/T	4	Y214Y	C = 0.875; T = 0.125
rs3219484	MUTYH[e]	c.64G>G/A	2	V22M	G = 0.925; A = 0.075
rs3219489	MUTYH[e]	c.972G>G/C	12	Q324H	G = 0.716; C = 0.284

a MAF, Minor Allele Frequency; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection.
b RefSeq NM_000038.4; NP_000029.2.
c RefSeq NM_000249.2; NP_000240.1.
d RefSeq NM_000179.2; NP_000170.1.
e RefSeq NM_001048171.1; NP_001041636.1.

heterozygous and homozygous variant genotypes were associated with an odds ratio of 0.75 (CI = 0.59–0.94; $p = 0.01$). The same trend was observable when we stratified the cases according to the location of the tumour, with colon and rectal cases showing odds ratios for the dominant model of 0.76 (CI = 0.60–0.97; $p = 0.03$) and 0.73 (CI = 0.56–0.95; $p = 0.02$), respectively.

Similarly, for MLH1 I219V there was no statistically significant association with CRC risk overall. However, when the analysis was restricted to the super-controls ($n = 411$) we found that being a carrier was associated with an odds ratio of 1.36 (CI = 1.08–1.71; $p < 0.01$) whereas being homozygote for the variant was associated with an odds ratio of 1.14 (CI = 0.79–1.65; $p = 0.48$). As observed for APC D1822V, we could detect an association when we assumed a dominant model of inheritance, with an odds ratio for the combined heterozygous and homozygous variant genotypes of 1.31 (CI = 1.06–1.63; $p = 0.01$). The same effect was detectable when we divided the cases according to the tumour location, with odds ratio of 1.34 (CI = 1.05–1.72; $p = 0.02$) and 1.28 (CI = 1.02–1.60; $p = 0.03$) for rectum and colon, respectively (under the dominant model as well).

Four SNPs in MSH6 were included in the analysis, rs1042821:G>A (G39E), rs1800932:A>G (P92P), rs1800935:T>C (D180D) and rs1800937:C>T (Y214Y). None of them seems to be associated with an increased risk when the whole cohort of cases versus controls was analysed but, again, a protective effect was observed for rs1800932:A>G (P92P) when only the super-controls ($n = 411$) were considered. In particular, the combination of heterozygotes and homozygotes had an odds ratio of 0.79 (CI = 0.62–0.99; $p = 0.04$) when the CRC cases were included, while it was 0.77 (CI = 0.60–0.99; $p = 0.04$) when the analysis was restricted to colon. A trend was observable also in rectum, but it did not reach significance probably due to the small sample size (OR = 0.83, CI = 0.63–1.09, $p = 0.22$).

In a similar way, rs1800935:T>C (D180D) showed a borderline association only with colon and the homozygous status, with an odds ratio of 0.74 (CI = 0.54–1.00; $p = 0.05$). The same effect only for the homozygous was observable when we considered the blood donors as controls (OR = 0.72; CI = 0.53–0.99; $p = 0.05$) (data not shown) but not when we restricted the analysis to the super-controls (OR = 0.79; CI = 0.50–1.23; $p = 0.29$).

While rs1800937:C>T (Y214Y) did not show any specific association, we could see a trend, although non-significant, for rs1042821:G>A (G39E) with an odds ratio of 2.11 (CI = 0.90–2.11; $p = 0.08$).

The last two SNPs, rs3219484:G>A (V22M) and rs3129489:G>C (Q324H) are located in the MUTYH gene. Q324H showed a significant association with rectal cancer when all the controls ($n = 1689$) were included in the analysis. We measured an odds ratio of 1.52 (CI = 1.06–2.17; $p = 0.02$) for the homozygous variant genotype, while the combination of the homozygous and the heterozygous variants, even if indicating the same phenomenon, was not significant (OR = 1.13; CI = 0.94–1.36; $p = 0.16$). Thus, as expected for the MUTYH gene, Q324H acts in a recessive manner (as observed for rs1800935:T>C [MSH6 D180D]). The same effect was found when we compared only the blood donors; in this case the magnitude of the effect is even stronger for the homozygous variant genotype, with an odds ratio of 1.75 (CI = 1.19–2.58; $p = 0.004$) (data not shown).

4. Discussion

The sequencing of the human genome has led to the identification of a large number of variants and the availability of the technology for carrying out large-scale genome-wide association studies allows to assess the individual cancer susceptibility in a thorough manner. More specifically for CRC, in the past 2 years several studies have identified a handful of loci associated with an increased risk,[12–18] but it seems unlikely that they (even together with others which will be published in the future) will explain a significant proportion of CRC cases.

An interesting hypothesis is that it could exist a whole range of variability associated with genes already known to be involved in CRC development. This could be due to SNPs with a small effect or it is also possible that a combination of several variants in many genes works synergistically to increase the risk of CRC. In order to test this hypothesis we set up a pilot study to investigate eight common variants previously found by our group and which are common enough to increase the power of detecting even a small increase in risk.

The APC D1822V is the most common missense variant described to date.[23,26] Codon 1822 lies between the fourth and the fifth of seven 20-aminoacid repeats which bind and

Table 2 – The eight SNPs analysed in the whole cohort of cases and controls ('colorectal') and stratified by tumour location ('colon' and 'rectum'). Analyses were performed both for all the available controls ('ALL ctrls') and for the 'super-controls' only ('superctrls'). In bold are the results with $p < 0.05$.

Polymorphism	Controls used	No. of controls	Colorectal		Colon		Rectum	
			No. of cases	OR (95% CI)	No. of cases	OR (95% CI)	No. of cases	OR (95% CI)
MLH1	*ALL ctrls*							
	AA	832	819	1.00	514	1.00	287	1.00
rs1799977	AG	708	781	1.12 (0.97–1.29)	483	1.10 (0.94–1.30)	275	1.13 (0.93–1.37)
c.655A>G	GG	161	181	1.14 (0.90–1.44)	103	1.04 (0.79–1.36)	69	1.24 (0.91–1.70)
I219V	AG or GG			1.13 (0.98–1.29)		1.09 (0.94–1.27)		1.15 (0.96–1.38)
	p_{trend}			0.10		0.41		0.11
	superctrls							
	AA	217	819	1.00	514	1.00	287	1.00
	AG	152	781	**1.36 (1.08–1.71)**	483	**1.34 (1.05–1.71)**	275	**1.21 (1.05–1.78)**
	GG	42	181	1.14 (0.79–1.65)	103	1.04 (0.70–1.53)	69	1.24 (.81–1.89)
	AG or GG			**1.31 (1.06–1.63)**		**1.28 (1.02–1.60)**		**1.34 (1.05–1.72)**
	p_{trend}			0.06		0.17		0.06
MSH6	*ALL ctrls*							
rs1042821	GG	1071	898	1.00	554	1.00	316	1.00
c.116G>G/A	GA	407	336	0.99 (0.83–1.17)	216	1.03 (0.85–1.25)	116	0.97 (0.76–1.23)
G39E	AA	41	42	1.22 (0.79–1.90)	23	1.08 (0.64–1.83)	18	1.49 (0.84–2.63)
	GA or AA			1.01 (0.86–1.19)		1.03 (0.86–1.24)		1.01 (0.81–1.28)
	p_{trend}			0.72		0.71		0.57
	superctrls							
	GG	296	898	1.00	554	1.00	316	1.00
	GA	118	336	0.939 (0.73–1.20)	216	0.978 (0.75–1.28)	116	0.921 (0.68–1.25)
	AA	8	42	1.731 (0.80–3.73)	23	1.536 (0.68–3.48)	18	2.108 (0.91–4.92)
	GA or AA			0.989 (0.78–1.26)		1.013 (0.78–1.31)		0.996 (0.76–1.33)
	p_{trend}			0.69		0.68		0.57
MSH6	*ALL ctrls*							
rs1800932	AA	1200	1294	1.00	802	1.00	455	1.00
c.276A>G	AG	468	442	0.88 (0.75–1.02)	272	0.87 (0.73–1.03)	159	0.90 (0.73–1.11)
P92P	GG	40	40	0.93 (0.59–1.45)	20	0.75 (0.43–1.29)	19	1.25 (0.72–2.19)
	AG or GG			0.88 (0.76–1.02)		0.86 (0.73–1.02)		0.92 (0.75–1.13)
	p_{trend}			0.12		0.07		0.69
	superctrls							
	AA	279	1294	1.00	802	1.00	455	1.00
	AG	119	442	0.80 (0.63–1.02)	272	0.80 (0.61–1.03)	159	0.82 (0.62–1.09)
	GG	13	40	0.66 (0.35–1.26)	20	0.53 (0.26–1.09)	19	0.90 (0.44–1.84)
	AG or GG			**0.79 (0.62–0.99)**		**0.77 (0.60–0.99)**		0.83 (0.63–1.09)
	p_{trend}			**0.04**		**0.02**		0.22
MSH6	*ALL ctrls*							
rs1800935	TT	870	961	1.00	590	1.00	342	1.00
c.540T>C	TC	693	693	0.91 (0.79–1.04)	439	0.93 (0.80–1.10)	236	0.87 (0.71–1.05)
D180D	CC	140	129	0.83 (0.65–1.08)	70	**0.74 (0.54–1.00)**	56	1.02 (0.73–1.42)
	TC or CC			0.89 (0.78–1.02)		0.90 (0.77–1.05)		0.89 (0.74–1.07)
	p_{trend}			0.08		0.07		0.45
	superctrls							
	TT	212	961	1.00	590	1.00	342	1.00
	TC	165	693	0.93 (0.74–1.16)	439	0.96 (0.75–1.21)	236	0.89 (0.68–1.15)
	CC	32	129	0.89 (0.59–1.35)	70	0.79 (0.50–1.23)	56	1.09 (0.68–1.73)
	TC or CC			0.92 (0.74–1.14)		0.93 (0.74–1.17)		0.92 (0.72–1.18)
	p_{trend}			0.44		0.36		0.79
MSH6	*ALL ctrls*							
rs1800937	CC	1180	1229	1.00	743	1.00	450	1.00
c.642C>C/T	CT	334	324	0.93 (0.78–1.11)	207	0.98 (0.81–1.20)	109	0.86 (0.67–1.09)
Y214Y	TT	17	14	0.79 (0.39–1.61)	6	0.56 (0.22–1.43)	8	1.23 (0.53–2.88)
	CT or TT			0.93 (0.78–1.10)		0.96 (0.79–1.17)		0.87 (0.69–1.11)
	p_{trend}			0.33		0.54		0.37

Table 2 (*continued*)

Polymorphism	Controls used	No. of controls	Colorectal		Colon		Rectum	
			No. of cases	OR (95% CI)	No. of cases	OR (95% CI)	No. of cases	OR (95% CI)
	superctrls							
	CC	336	1229	1.00	743	1.00	450	1.00
	CT	87	334	1.02 (0.78–1.33)	207	1.08 (0.81–1.43)	109	0.93 (0.68–1.28)
	TT	2	14	1.91 (0.43–8.46)	6	1.36 (0.27–6.76)	8	2.99 (0.63–14.16)
	CT or TT			1.04 (0.80–1.35)		1.08 (0.82–1.43)		0.98 (0.72–1.34)
	p_{trend}			0.66		0.55		0.82
APC	*ALL ctrls*							
rs459552	TT	937	1033	1.00	633	1.00	372	1.00
c.5465T>A	TA	607	616	0.92 (0.80–1.06)	383	0.93 (0.79–1.10)	214	0.89 (0.73–1.08)
D1822V	AA	97	105	0.98 (0.74–1.31)	66	1.01 (0.73–1.40)	38	0.99 (0.67–1.46)
	TA or AA			0.93 (0.81–1.07)		0.94 (0.81–1.10)		0.90 (0.75–1.09)
	p_{trend}			0.41		0.61		0.41
	superctrls							
	TT	178	1033	1.00	633	1.00	372	1.00
	TA	139	616	**0.77 (0.60–0.97)**	383	0.78 (0.60–1.00)	214	**0.74 (0.56–0.97)**
	AA	27	105	0.67 (0.43–1.05)	66	0.69 (0.43–1.11)	38	0.67 (0.40–1.14)
	TA or AA			**0.75 (0.59–0.94)**		**0.76 (0.60–0.97)**		**0.73 (0.56–0.95)**
	p_{trend}			**0.01**		**0.03**		**0.02**
MYH	*ALL ctrls*							
rs3219484	GG	1420	1503	1.00	933	1.00	528	1.00
c.64G>A	GA	264	254	0.91 (0.75–1.10)	151	0.87 (0.70–1.08)	96	0.98 (0.76–1.26)
V22M	AA	7	6	0.81 (0.27–2.42)	3	0.65 (0.17–2.53)	3	1.15 (0.30–4.75)
	GA or AA			0.91 (0.75–1.10)		0.87 (0.69–1.07)		0.98 (0.76–1.26)
	p_{trend}			0.29		0.17		0.92
	superctrls							
	GG	341	1503	1.00	933	1.00	528	1.00
	GA	62	254	0.93 (0.69–1.26)	151	0.89 (0.65–1.23)	96	1.00 (0.71–1.42)
	AA	3	6	0.45 (0.11–1.82)	3	0.37 (0.07–1.82)	3	0.65 (0.13–3.22)
	GA or AA			0.91 (0.68–1.22)		0.87 (0.63–1.19)		0.98 (0.70–1.38)
	p_{trend}			0.42		0.28		0.85
MYH	*ALL ctrls*							
rs3219489	GG	992	1019	1.00	640	1.00	351	1.00
c.972G>C	GC	600	627	1.02 (0.88–1.17)	382	0.99 (0.84–1.16)	226	1.07 (0.88–1.29)
Q324H	CC	97	124	1.24 (0.94–1.65)	70	1.12 (0.81–1.54)	52	**1.52 (1.06–2.17)**
	GC or CC			1.05 (0.92–1.20)		1.01 (0.86–1.17)		1.13 (0.94–1.36)
	p_{trend}			0.25		0.74		0.06
	superctrls							
	GG	236	1019	1.00	640	1.00	351	1.00
	GC	141	627	1.03 (0.82–1.30)	382	1.00 (0.78–1.28)	226	1.08 (0.83–1.41)
	CC	33	124	0.87 (0.58–1.31)	70	0.78 (0.50–1.21)	52	1.06 (0.67–1.69)
	GC or CC			1.00 (0.80–1.24)		0.96 (0.76–1.21)		1.07 (0.84–1.38)
	p_{trend}			0.76		0.46		0.63

down-regulate β-catenin.[27] Although, the functional significance of this substitution remains unknown, this is a change from a hydrophilic aspartate to a hydrophobic valine residue. The clinical relevance of APC D1822V is still under debate, since some studies showed that it could act as a low-penetrance allele that increases the risk of developing CRC[28,29] while others consider it as a common polymorphism without clinical consequences.[30,31] Furthermore, five previous studies investigated the possible relationship between APC D1822V with dietary and lifestyle factors. In two cases, a diet low in total saturated and unsaturated fats decreased the risk of CRC among the homozygous carriers of the variant.[27,32] In another study a significant interaction was observed with the dietary intakes of cholesterol, calcium and fibre.[33] On the other hand, two other studies failed to detect this association.[34,35] In our study we were unable to see an effect when we compared our CRC cases and the whole cohort of controls, even if a trend towards an increased risk for the dominant model was observed. On the other hand, an effect was detectable when we restricted the analysis to our smaller group of super-controls, both for the CRC cohort as a whole and for colon and rectum separately. A dominant model fits well our data in all the analyses.

The MLH1 I219V polymorphism lies in exon 8 and is conserved throughout evolution, which can be seen as a sign of its importance for the well functioning of the protein. A functional study based on a transient transfection of hMLH1 complementary DNA carrying the variant into a human

embryonic kidney fibroblast cell line lacking hMLH1 expression was able to prove *in vivo* that this SNP does not affect the mismatch repair capacity.[36] Berndt and collaborators investigated the possible role of MLH1 I219V in CRC using a community-based cohort of Caucasian American and were not able to find any association with tumours in any location (proximal or distal colon, rectum).[37] In contrast to this, a study performed on Finnish and Swedish HNPCC or HNPCC-like families showed that the SNP affects the level of expression.[38] Using the Single Nucleotide Primer Extension (SnuPE) method to be able to assess the expression of the maternal and the paternal alleles separately, they came to the conclusion that the reduced mRNA expression segregates with the cancer phenotype in the families which were tested positive.[38] The discrepancy with the study performed by Berndt and collaborators could be that in our study we genotyped a much higher number of colon and rectal cases (1100 and 631, respectively) or due to the different population used. A recent study systematically tagged all the known common variants in the MMR genes in order to test the tag SNPs (tSNPs) for association with CRC. Three SNPs (one in MSH2 and two in MSH3 gene) showed borderline associations but none remained significant after correcting for multiple testing.[39] The I219V MLH1, which was found to be associated with an increased risk in our study, failed the assay design and was not in linkage disequilibrium (LD) with any other tag SNP.[39] Hence, its role in CRC predisposition was not assessed.

We also tested four variants in the MSH6 gene, rs1042821:G>A (G39E), rs1800932:A>G (P92P), rs1800935:T>C (D180D) and rs1800937:C>T (Y214Y). For the first two silent variants we could see a protective effect when we stratified the samples according to the tumour site and when only the super-controls were included in the analysis. In particular, rs1800932:A>G (P92P) showed a protective effect for tumours in the colorectum and colon. For rs1800935:T>C (D180D) we observed the same protective effect only in colon. These two variants had previously been found by our group in two families (106 and 154) suggested to have HNPCC but where a mutation in the MMR genes had not been found.[40] Interestingly, the two patients with endometrial cancer but not CRC were both compound heterozygotes for the variants. The two SNPs are located only 5 kb apart from each other and show a high correlation ($r^2 = 0.56$). However, there was no difference in frequency of the heterozygous neither when we compare all CRC cases and controls nor in CRC cases with endometrial cancers in the family and controls (data not shown).

Rs1042821:G>A (G39E) is probably the most interesting variant in MSH6 tested in the present study, since an earlier work suggested a slightly higher frequency for the A allele among CRC cases (23%) than controls (15%), even though a statistical comparison was not performed.[41] More recently, Berndt and collaborators reported that AA carriers have a very high risk of developing rectal but not colon cancer (RR = 3.25; CI = 1.08–9.83; $p = 0.04$),[37] while Campbell and collaborators found an increased risk for male carriers, both heterozygous and homozygous, while no association was observed among females.[42] In our study there was no difference in frequency neither between cases and controls nor between females and males (data not shown). Similarly, the last SNP on MSH6, rs1800937:C>T (Y214Y) did not show any significant association.

We investigated two variants in the MUTYH gene, rs3219484:G>A (V22M) and rs3219489:G>C (Q324H), both previously reported by us and others,[22,43] and generally considered as common polymorphisms. V22M represents a semi-conservative change and is located 2 bp from the catalytic ASP site within the N-terminal domain of the expressed protein while Q324H is on exon 12 and it seems to be outside any functional domain.

No difference was observed for V22M, while we could see an increased risk associated with the homozygous variant of Q324H (OR = 1.52 [1.06–2.17]; $p = 0.02$), compatible with a recessive inheritance. The same effect was found when we compared only the blood donors, with an odds ratio for the homozygous variant genotype of 1.75 (CI = 1.19–2.58; $p = 0.004$).

Interestingly, we also observed that, when a recessive inheritance was suggested, like for rs1800935:T>C (MSH6 D180D) and rs3219489:G>C (MUTYH Q324H) the significant association was of the same magnitude or stronger in the blood donor cohort, rather than in the super-controls. In contrast, a dominant effect was more evident when using only the super-controls that were selected to carry as few risk alleles as possible.

In this study we have investigated the role of eight SNPs in MLH1, MSH6, APC and MUTYH as low-penetrance alleles in CRC in our cohort of 1785 cases and 1722 healthy controls. We found evidence that these CRC genes could act as low-penetrance alleles. There is also a possibility of important interactions between the loci. Recently, Koessler and collaborators performed a similar analysis and found a protective effect associated with being homozygous for one SNP in MSH6 and one SNP in MSH2.[39] We investigated all the possible two-loci interactions using the case–control approach in a chi-squared test. No evidence of interaction was found in our study (after correction for multiple testing) but this could be due to either the small number of SNPs tested or to the limited size of our cohort.

Genome-wide association studies have been designed to detect common alleles in the population of European ancestry and the studies published so far indicate that the risk associated with such variants is generally modest. The commonly used 550K chip from Illumina captures on average ~80% of common SNPs (i.e. those with $r^2 > 0.8$), but only ~12% of SNPs with a MAF of 5–10%, limiting the power to detect this relatively rare class of variants.[12] In our case, all the SNPs analysed have a high frequency in the general population. Nonetheless they could have been missed due to submaximal LD with tagging SNPs or because of the limited sample size of the cohorts that have been used so far. In conclusion, the approach we chose in this study could help reveal additional low-susceptibility loci in the genes already known to be responsible for CRC and deserves to be tested in even bigger cohorts.

Conflict of interest statement

None declared.

Acknowledgements

This work was funded by the Swedish Cancer Foundation, the Stockholm Cancer Society and the Nilsson-Ehle Foundation. We are indebted to the families for their cooperation. We also thank Berith Wejderot, Emilie Hultin and Afshin Ahmadian for their technical assistance.

REFERENCES

1. Lichtenstein P, Holm NV, Verkasalo PK, 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.
2. Heavey PM, McKenna D, Rowland IR. Colorectal cancer and the relationship between genes and the environment. *Nutr Cancer* 2004;**48**:124–41.
3. Baglioni S, Genuardi M. Simple and complex genetics of colorectal cancer susceptibility. *Am J Med Genet C Semin Med Genet* 2004;**129C**:35–43.
4. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 1991;**66**:589–600.
5. Kinzler KW, Nilbert MC, Vogelstein B, et al. Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* 1991;**251**:1366–70.
6. Lindblom A, Tannergard P, Werelius B, Nordenskjold M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet* 1993;**5**:279–82.
7. Peltomaki P, Aaltonen LA, Sistonen P, et al. Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 1993;**260**:810–2.
8. Nicolaides NC, Papadopoulos N, Liu B, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 1994;**371**:75–80.
9. Papadopoulos N, Nicolaides NC, Liu B, et al. Mutations of GTBP in genetically unstable cells. *Science* 1995;**268**:1915–7.
10. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. *N Engl J Med* 2003;**348**:919–32.
11. Sieber OM, Lipton L, Crabtree M, et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N Engl J Med* 2003;**348**:791–9.
12. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008;**40**:1426–35.
13. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 2007;**39**:1315–7.
14. Jaeger E, Webb E, Howarth 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.
15. Tenesa A, Farrington SM, Prendergast JG, 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**:631–7.
16. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 2008;**40**:623–30.
17. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;**39**:989–94.
18. Tomlinson I, Webb E, Carvajal-Carmona L, 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.
19. Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet* 2009;**10**:353–8.
20. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant or not? *Hum Mol Genet* 2002;**11**:2417–23.
21. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 2008;**82**:100–12.
22. Zhou XL, Djureinovic T, Werelius B, et al. Germline mutations in the MYH gene in Swedish familial and sporadic colorectal cancer. *Genet Test* 2005;**9**:147–51.
23. Zhou XL, Eriksson U, Werelius B, et al. Definition of candidate low risk APC alleles in a Swedish population. *Int J Cancer* 2004;**110**:550–7.
24. Liu HX, Zhou XL, Liu T, et al. The role of hMLH3 in familial colorectal cancer. *Cancer Res* 2003;**63**:1894–9.
25. Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics* 1997;**53**:1253–61.
26. Powell SM, Zilz N, Beazer-Barclay Y, et al. APC mutations occur early during colorectal tumorigenesis. *Nature* 1992;**359**:235–7.
27. Slattery ML, Samowitz W, Ballard L, et al. A molecular variant of the APC gene at codon 1822: its association with diet, lifestyle, and risk of colon cancer. *Cancer Res* 2001;**61**:1000–4.
28. Wallis YL, Morton DG, McKeown CM, Macdonald F. Molecular analysis of the APC gene in 205 families: extended genotype–phenotype correlations in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition. *J Med Genet* 1999;**36**:14–20.
29. Cleary SP, Kim H, Croitoru ME, et al. Missense polymorphisms in the adenomatous polyposis coli gene and colorectal cancer risk. *Dis Colon Rectum* 2008;**51**:1467–73. discussion 1473–1464.
30. Chen J, Giovannucci E, Hankinson SE, et al. A prospective study of methylenetetrahydrofolate reductase and methionine synthase gene polymorphisms, and risk of colorectal adenoma. *Carcinogenesis* 1998;**19**:2129–32.
31. Ruiz-Ponte C, Vega A, Conde R, Barros F, Carracedo A. The Asp1822Val variant of the APC gene is a common polymorphism without clinical implications. *J Med Genet* 2001;**38**:E33.
32. Theodoratou E, Campbell H, Tenesa A, et al. Modification of the associations between lifestyle, dietary factors and colorectal cancer risk by APC variants. *Carcinogenesis* 2008;**29**:1774–80.
33. Guerreiro CS, Cravo ML, Brito M, et al. The D1822V APC polymorphism interacts with fat, calcium, and fiber intakes in modulating the risk of colorectal cancer in Portuguese persons. *Am J Clin Nutr* 2007;**85**:1592–7.
34. Menendez M, Gonzalez S, Blanco I, et al. Colorectal cancer risk and the APC D1822V variant. *Int J Cancer* 2004;**112**:161–3.
35. Tranah GJ, Giovannucci E, Ma J, Fuchs C, Hunter DJ. APC Asp1822Val and Gly2502Ser polymorphisms and risk of colorectal cancer and adenoma. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:863–70.
36. Trojan J, Zeuzem S, Randolph A, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology* 2002;**122**:211–9.
37. Berndt SI, Platz EA, Fallin MD, et al. Mismatch repair polymorphisms and the risk of colorectal cancer. *Int J Cancer* 2007;**120**:1548–54.

38. Renkonen E, Zhang Y, Lohi H, et al. Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. *J Clin Oncol* 2003;**21**:3629–37.
39. Koessler T, Oestergaard MZ, Song H, et al. Common variants in mismatch repair genes and risk of colorectal cancer. *Gut* 2008;**57**:1097–101.
40. Liu T, Chen J, Salahshor S, et al. Screening families with endometrial, colorectal cancers for germline mutations. *J Med Genet* 2001;**38**:E29.
41. Kolodner RD, Tytell JD, Schmeits JL, et al. Germ-line msh6 mutations in colorectal cancer families. *Cancer Res* 1999;**59**:5068–74.
42. Campbell PT, Curtin K, Ulrich CM, et al. Mismatch repair polymorphisms and risk of colon cancer, tumour microsatellite instability and interactions with lifestyle factors. *Gut* 2009;**58**:661–7.
43. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C → T:A mutations in colorectal tumors. *Nat Genet* 2002;**30**:227–32.