

ARTICLE

Clinical relevance of 8q23, 15q13 and 18q21 SNP genotyping to evaluate colorectal cancer risk

Stéphanie Baert-Desurmont^{1,2,3}, Françoise Charbonnier^{1,2,3}, Estelle Houivet^{4,5}, Lorena Ippolito^{1,3}, Jacques Mauillon², Marion Bougeard¹, Caroline Abadie^{1,6}, David Malka⁷, Jacqueline Duffour⁸, Françoise Desseigne⁹, Chrystelle Colas¹⁰, Pascal Pujol¹¹, Sophie Lejeune¹², Catherine Dugast⁶, Bruno Buecher¹³, Laurence Faivre¹⁴, Dominique Leroux¹⁵, Paul Gesta¹⁶, Isabelle Coupier¹¹, Rosine Guimbaud¹⁷, Pascaline Berthet¹⁸, Sylvie Manouvrier¹², Estelle Cauchin¹⁹, Fabienne Prieur²⁰, Pierre Laurent-Puig²¹, Marine Lebrun²⁰, Philippe Jonveaux²², Jean Chiesa²³, Olivier Caron⁷, Marie-Emmanuelle Morin-Meschin²⁴, Florence Polycarpe-Osaer¹⁸, Sophie Giraud²⁵, Aziz Zaanani²¹, Delphine Bonnet²⁶, Ludovic Mansuy²², Valérie Bonadona⁹, Salima El Chehadeh¹⁴, François Duhoux²⁷, Marion Gauthier-Villars¹³, Jean-Christophe Saurin²⁸, Marie- Agnès Collonge-Rame²⁹, Laurence Brugières⁷, Qing Wang⁹, Brigitte Bressac-de Paillerets⁷, Jean-Marc Rey¹¹, Christine Toulas¹⁷, Marie-Pierre Buisine³⁰, Myriam Bronner²², Joanna Sokolowska²², Agnès Hardouin¹⁸, Anne-Françoise Cailleux⁵, Hakim Sebaoui³¹, Julien Blot³¹, Julie Tinat^{1,2,3}, Jacques Benichou⁴ and Thierry Frebourg^{*,1,2,3}

To determine if the at-risk single-nucleotide polymorphism (SNP) alleles for colorectal cancer (CRC) could contribute to clinical situations suggestive of an increased genetic risk for CRC, we performed a prospective national case–control study based on highly selected patients (CRC in two first-degree relatives, one before 61 years of age; or CRC diagnosed before 51 years of age; or multiple primary CRCs, the first before 61 years of age; exclusion of Lynch syndrome and polyposes) and controls without personal or familial history of CRC. SNPs were genotyped using SNaPshot, and statistical analyses were performed using Pearson's χ^2 test, Cochran–Armitage test of trend and logistic regression. We included 1029 patients and 350 controls. We confirmed the association of CRC risk with four SNPs, with odds ratio (OR) higher than previously reported: *rs16892766* on 8q23.3 (OR: 1.88, 95% confidence interval (CI): 1.30–2.72; $P=0.0007$); *rs4779584* on 15q13.3 (OR: 1.42, CI: 1.11–1.83; $P=0.0061$) and *rs4939827* and *rs58920878/Novel 1* on 18q21.1 (OR: 1.49, CI: 1.13–1.98; $P=0.007$ and OR: 1.49, CI: 1.14–1.95; $P=0.0035$). We found a significant ($P<0.0001$) cumulative effect of the at-risk alleles or genotypes with OR at 1.62 (CI: 1.10–2.37), 2.09 (CI: 1.43–3.07), 2.87 (CI: 1.76–4.70) and 3.88 (CI: 1.72–8.76) for 1, 2, 3 and at least 4 at-risk alleles, respectively, and OR at 1.71 (CI: 1.18–2.46), 2.29 (CI: 1.55–3.38) and 6.21 (CI: 2.67–14.42) for 1, 2 and 3 at-risk genotypes, respectively. Combination of SNPs may therefore explain a fraction of clinical situations suggestive of an increased risk for CRC.

European Journal of Human Genetics (2016) 24, 99–105; doi:10.1038/ejhg.2015.72; published online 15 April 2015

INTRODUCTION

Clinical situations suggestive of an inherited predisposition to colorectal cancer (CRC) are, like in other cancers, familial aggregation, early age of tumour onset and the development of multiple primary

tumours in the same individual. Molecular diagnosis of inherited CRC is stratified according to the clinical, pathological and genetic presentation of the families (see Patel and Ahnen¹). Patients presenting with adenomatous polyposis are subjected to *APC* and/or *MUTYH*

¹Inserm U1079, University of Rouen, Institute for Research and Innovation in Biomedicine, Rouen, France; ²Department of Genetics, University Hospital, Rouen, France; ³Cancéropôle Nord-Ouest, France; ⁴Department of Biostatistics, University Hospital and Inserm U657, University of Rouen, Institute for Research and Innovation in Biomedicine, Rouen, France; ⁵Centre d'Investigation Clinique, University Hospital, Rouen, France; ⁶Department of Genetics, Comprehensive Cancer Center Eugène Marquis, Rennes, France; ⁷Department of Genetics, Institut Gustave Roussy, Villejuif, France; ⁸Department of Genetics, Comprehensive Cancer Center Val d'Aurelle, Montpellier, France; ⁹Department of Genetics, Comprehensive Cancer Center Léon Bérard, Lyon, France; ¹⁰Department of Genetics, AP-HP, Hôpital Saint Antoine, Paris, France; ¹¹Department of Genetics, University Hospital, Montpellier, France; ¹²Department of Genetics, University Hospital, Lille, France; ¹³Department of Genetics, Institut Curie, Paris, France; ¹⁴Department of Genetics, University Hospital and Comprehensive Cancer Center Georges-François Leclerc, Dijon, France; ¹⁵Department of Genetics, University Hospital, Grenoble, France; ¹⁶Department of Genetics, Hospital, Niort, France; ¹⁷Department of Genetics, Comprehensive Cancer Center Claudius Regaud, Toulouse, France; ¹⁸Department of Genetics, Comprehensive Cancer Center François Baclesse, Caen, France; ¹⁹Department of Genetics, University Hospital, Nantes, France; ²⁰Department of Genetics, University Hospital, Saint Etienne, France; ²¹Department of Genetics, Hôpital Européen Georges Pompidou, Paris, France; ²²Department of Genetics, University Hospital, Nancy, France; ²³Department of Genetics, University Hospital, Nîmes, France; ²⁴Department of Genetics, University Hospital, Angers, France; ²⁵Department of Genetics, University Hospital Edouard Herriot, Lyon, France; ²⁶Department of Genetics, University Hospital, Toulouse, France; ²⁷Department of Genetics, Cancer Centre, Cliniques Universitaires St Luc, Bruxelles, Belgium; ²⁸Department of Gastroenterology, University Hospital Edouard Herriot, Lyon, France; ²⁹Department of Genetics, University Hospital, Besançon, France; ³⁰Molecular of Genetics, University Hospital, Lille, France; ³¹Department of Clinical Research, University Hospital, Rouen, France

*Correspondence: Professor T Frebourg, Inserm U1079, Department of Genetics, Rouen University Hospital, Institute of Research and Innovation in Medicine, Faculty of Medicine, 22 Boulevard Gambetta, Rouen 76183, France. Tel: +33 2 32 88 81 82; Fax: +33 2 32 88 80 80; E-mail: frebourg@chu-rouen.fr

Received 22 September 2014; revised 19 January 2015; accepted 27 January 2015; published online 15 April 2015

analyses, according to the number of polyps and the familial history suggestive of either a dominant or a recessive mode. The rare patients presenting hamartomatous polyps are subjected to *STK11*, *SMAD4*, *BMPRIA* or *PTEN* analysis, and, for the vast majority of the patients who do not present with polyposis, the presence of a Replication Error (RER)/Microsatellite Instability (MSI) phenotype will be searched within the tumour to establish a somatic signature of the Lynch syndrome. Patients with RER+/MSI tumours will be subjected to the analysis of the MMR genes, *MSH2*, *MSH6*, *MLH1* and *PMS2*.

An important fraction of clinical situations suggestive of an increased genetic risk for CRC cannot be explained by a simple Mendelian model involving one of these key CRC genes. Analysis of the activities of the French laboratories by the French National Cancer Institute (INCa) indicates that, among 2000 index cases suspected to present an inherited form of CRC and subjected each year to the key CRC genes testing, the mutation detection rate is lower than 15%. Besides the Mendelian forms of CRC, another significant advance in the genetic determinism of CRC was the identification, owing to numerous genome-wide association studies (GWAS), of genetic risk factors corresponding to single-nucleotide polymorphisms (SNPs). So far, 20 SNPs, respectively located on 1q41, 3q26.2, 6p21.2, 8q23.3, 8q24.21, 10p14, 11q13.4, 11q23.1, 12q13, 14q22.2, 15q13.3, 16q22.1, 18q21.1, 19q13.11, 20p12.3, 20q13.33 and Xp22.2, have been reported to be associated with an increased CRC risk.²⁻¹² Each of these SNPs confers a small increase in CRC risk with an odds ratio (OR) usually in the magnitude of 1.2. Although for most of these SNPs, the association with CRC has been confirmed by replication studies, the low CRC risk increase conferred by these variants hampers their clinical use. Considering that these GWAS studies have been performed on very large numbers of patients to reach sizeable statistical powers, it is possible that the heterogeneity of CRC patient cohorts, in terms of CRC onset age and/or familial history, the heterogeneity of the patient genetic background and the lack, in most of these studies, of a strict definition of the controls may have diluted the risk associated with these SNPs.

Therefore, to further evaluate the contribution of these SNPs to the genetic determinism of CRC and to determine if these SNPs, alone or in combination, could explain a fraction of CRC occurrence in patients whose clinical presentation is suggestive of an increased genetic risk not explained by a deleterious mutation within one CRC key genes, we performed a prospective study on very carefully selected patients and controls. We selected for this study SNPs whose association with CRC in European populations had been validated by replication studies, showing the highest association with CRC, and which were in the vicinity of genes potentially involved in CRC: (1) *rs16892766* on 8q23.3, which is 140 kb centromeric to *EIF3H* encoding

the subunit H of the eukaryotic translation initiation factor 3;^{5,8,13} *in vitro* EIF3H overexpression has been shown to increase the cellular proliferation of CRC-derived cell lines, whereas reduction of EIF3H has the opposite effect;¹³ (2) *rs6983267* on 8q24.21,^{3,5,6,8,14,15} this SNP affects a binding site for TCF4, a downstream effector of the Wnt/ β -catenin signalling pathway commonly deregulated in CRC;¹⁶ (3) *rs4779584* on 15q13.3 between the *GREM1* and *SCG5* locus.^{4,5,8,10} *GREM1* encodes one bone morphogenic antagonist secreted by intestinal subepithelial myofibroblasts and increased *GREM1* expression, resulting from an upstream region duplication, has been shown to be in some families the molecular bases of hereditary mixed polyposis syndrome, which indicates that alteration of *GREM1* expression is tumourigenic;¹⁷ (4) *rs4939827* on 18q21.1 within intron 3 of *SMAD7*, encoding an intracellular antagonist of the TGF β pathway frequently inactivated in CRC.^{2,5,6,11,15} The association of CRC with this SNP was subsequently shown to be due to another SNP, *rs58920878/Novel 1*, also located within *SMAD7* intron 3 and regulating *SMAD7* expression.¹⁸ We also selected for this study the 10p14 SNP *rs10795668*, initially detected by a large GWAS study⁵, as a recent meta-analysis confirmed its association with increased CRC risk,¹⁹ and *rs3802842* on 11q23.1,⁶ as a strong association with CRC had been reported in Dutch CRC cohort enriched in familial and/or early CRC cases.⁸

The medical objective of this study was to determine whether or not the increased risk conferred by these SNPs could be sufficient to justify their genotyping in clinical practice.

MATERIALS AND METHODS

Subjects

Patients, from Caucasian origin, were selected according to three different criteria suggestive of an increased genetic risk for CRC (Table 1). All patients were recruited by the French network of Cancer Genetics Departments to ensure the strict observation of inclusion criteria. Controls, recruited by the Clinical Investigation Center of Rouen University Hospital, were healthy volunteers from Caucasian origin, between 45 and 60 years of age, without any personal or familial history among their first-degree relatives of CRC. For each subject, informed consent both for study and performance of genetic analyses was obtained. The study was approved by the local ethical committee.

Genotyping

For controls and most of the patients, DNA was extracted from peripheral blood using the Flexigen Kit (Qiagen, Courtaboeuf, France) and quality of DNA was assessed using the NanoVue spectrophotometer (GE Healthcare Life Sciences, Velizy-Villacoublay, France). Genotyping of the variants *rs16892766*: A>C (chr8. GRCh38: g.116618444A>C), *rs6983267*: G>T (chr8. GRCh38: g.127401060G>T), *rs10795668*: G>A (chr10. GRCh38: g.8659256G>A), *rs3802842*: C>A (chr11. GRCh38: g.111300984C>A), *rs4779584*: T>C (chr15. GRCh38: g.32702555T>C), *rs4939827*: T>C (chr18. GRCh38: g.48927093T>C) and *rs58920878*: C>G (chr18. GRCh38: g.48923195C>G)

Table 1 Inclusion and exclusion criteria

Inclusion criteria

- CRC in two first-degree relatives, one being diagnosed before 61 years of age
- *or* CRC diagnosed before 51 years of age *or* advanced colorectal adenoma (diameter over 1 cm, and/or tubulovillous or villous and/or with high-grade dysplasia) before 41 years of age
- *or* Multiple primary CRCs in the same individual, the first one being diagnosed before 61 years of age

Exclusion criteria

- Lynch syndrome, as defined by the presence in the patient of a deleterious germline MMR gene mutation and/or a RER+/MSI tumour, in the context of a suggestive presentation (familial history, early age of onset)
- Adenomatous polyposis, as defined by more than 10 histologically proven adenomas
- Hamartomatous polyposis, as defined by the presence of histologically proven hamartomas

was performed using the SNaPshot methodology following the manufacturer's protocol (Applied Biosystems, Meylan, France). This genotyping method was validated by Sanger sequencing of the 7 SNPs in 10 samples. Quality of the genotyping was checked by analysing 10% of the samples in duplicate, by two different technicians. The concordance was 100%.

Statistical analyses

The target sample size was calculated for a 2:1 case-to-control ratio, no individual case-control matching, two-sided 0.05 Type I error and to obtain at least 80% power for an OR value of 2 in the range 0.06–0.90 for allele frequencies. The OR value of 2 was used as the aim of this study was to determine genetic variations conferring a clinically relevant increase in CRC risk. Based on these criteria, the target sample size was set at 700 cases and 350 controls. Sample size calculations were performed using the POWER program (from EPITOME software available on the US NCI website <http://dceg.cancer.gov/tools/design/power>).

For each SNP, genotype frequencies in cases and controls were tested for deviations from expected frequencies under the Hardy–Weinberg equilibrium, using the χ^2 goodness-of-fit test. For each SNP, frequencies of the three possible genotypes were compared between cases and controls using Pearson's χ^2 test. For each of the two genotypes including the at-risk allele reported in the literature, the OR was estimated along with its 95% confidence interval (CI) from logistic regression using the remaining genotype as the reference and with adjustment for sex. The same analyses were performed after dichotomization of genotypes according to a dominant model based on the at-risk allele. The ability of the three-genotype model to better explain risk (ie, to distinguish between CRC cases and controls) was compared with that of the two-level dominant model using the likelihood ratio test within logistic regression to assess the appropriateness of the dominant model. We applied Bonferroni's correction for multiple testing. To evaluate the potential cumulative effect of at-risk alleles, we compared the distribution of the number of at-risk alleles and at-risk genotypes between patients and controls, using the Cochran–Armitage test of trend. OR with 95% CIs were obtained from logistic regression. These analyses were performed using the SAS software (version 9.2; SAS Institute Inc., Cary, NC, USA). Data are available at: <http://www.gwascentral.org/study/HGVST1828>.

RESULTS

According to the criteria presented in Table 1, we included 1029 patients (412 males (40%), 617 females (60%)) whose clinical presentation was suggestive of an increased genetic risk for CRC but which did not correspond to a known Mendelian form of CRC. None of the patients presented with either adenomatous polyposis defined on the basis of at least 10 adenomas or hamartomatous polyposes. In all the patients, the diagnosis of Lynch syndrome had been excluded, either on the basis of the absence of an RER/MSI phenotype and a normal immunostaining of the MMR proteins in the tumour or on the absence of detectable germline *MSH2*, *MLH1* and *MSH6* mutations and rearrangements, when tumour analysis was not possible. Patient clinical characteristics are presented in Table 2. Taking into account that some of these patients fulfilled several inclusion criteria, 363 had a family history of CRC in two first-degree relatives, one having been diagnosed before 61 years of age, 887 had been diagnosed with CRC before age 51 years or with advanced colorectal adenoma before age 41, and 44 patients presented with multiple primary CRCs, the first one being diagnosed before 61 years. We recruited 350 controls (113 males (32.3%) and 237 females (67.7%)) of 45–60 years of age without personal or familial history of CRC, among their first-degree relatives. Mean ages (\pm SD) of cases at inclusion and controls were 48.7 ± 10.2 and 51.7 ± 4.5 years, respectively. The seven SNPs analysed in this study, *rs16892766* on 8q23.3, *rs6983267* on 8q24.21, *rs10795668* on 10p14, *rs3802842* on 11q23.1, *rs4779584* on 15q13.3, *rs4939827* and *rs58920878/Novel 1* on 18q21.1, were in Hardy–Weinberg equilibrium, both in controls and patients (data not shown). Table 3 displays

Table 2 Clinical characteristics of the 1029 patients included in the study

Characteristics	No. (%)
Patients with CRC ^a	952 (92.5%)
<i>Number of CRC sites</i>	
1	901 (94.6%)
2	44 (4.6%)
3	7 (0.7%)
Median age at CRC diagnosis (min; max)	44 (15; 62)
Mean age of CRC (SD)	43.1 (8.3)
<i>Localization of CRC</i>	
Caecum and right colon	188 (19.7%)
Transverse colon	34 (3.6%)
Left colon and sigmoid	350 (36.8%)
Recto-sigmoid junction	103 (10.8%)
Rectum	261 (27.4%)
Not indicated	16 (1.7%)
Patients with adenomas ^b	280 (27.2%)

^aOther patients presented with advanced adenomas.

^bUp to 10 adenomas.

genotype frequencies in CRC patients and controls. As shown in Table 3, there was a statistically significant difference between CRC patients and controls regarding the distributions of *rs16892766* on 8q23.3, *rs4779584* on 15q13.3, *rs4939827* and *rs58920878/Novel 1* on 18q21.1, but not those of *rs6983267* on 8q24.21, *rs10795668* on 10p14 and *rs3802842* on 11q23.1. For all these SNPs, there was no evidence that CRC risk was better explained by distinguishing the three genotypes than by grouping them according to a dominant model (likelihood ratio test, all *P*-value between 0.19 and 0.74). The association of the at-risk level of *rs16892766* on 8q23.3 according to the dominant model was the highest of the four SNPs associated with CRC risk, with an OR estimated at 1.88 (95% CI: 1.30–2.72), whereas the remaining three SNPs showed lower (but still significant, even after Bonferroni's correction) associations with very close ORs in the range 1.42–1.49 (Table 3). Almost all the controls (80/89) that were homozygote for the *rs4939827 C* allele were also homozygote for the *rs58920878/Novel 1 C* allele and the majority of the homozygote controls for the *rs4939827 T* allele (67/104) were homozygote for the *rs58920878/Novel 1 G* allele, confirming that, as previously shown,¹⁸ the two SNPs are in complete linkage disequilibrium and constitute a haplotypic block on 18q21.1.

Associations of CRC risk with SNPs were further assessed within the subgroups as defined by the inclusion criteria. Based on our results above, only dominant models were considered and only the three SNPs, *rs16892766* on 8q23.3, *rs4779584* on 15q13.3 and *rs58920878/Novel 1* on 18q21.1, were assessed. The *rs58920878/Novel 1* SNP was considered rather than *rs4939827* on 18q21.1 in view of previous results showing it to support the association of CRC with 18q21.1,¹⁸ but the results were similar with either SNP. As shown in Tables 4 and 5, even after Bonferroni's correction for multiple testing, we observed, as compared with controls, a significant enrichment of the at-risk allele for the 8q23.3 and 18q21.1 SNPs in patients selected on the basis of the occurrence of CRC in two first-degree relatives and for the 8q23.3, 15q13.3 and 18q21.1 SNPs in patients who developed CRC before 51 years of age or advanced colorectal adenoma before 41 years. The number of patients selected on the basis of the development of

Table 3 Assessment of associations between colorectal tumour risk and each of the seven at-risk alleles of SNPs^a

Locus	Genotype	Controls (N= 350), no. (%)	Patients (N= 1029), no. (%)	P-value ^b	Odds ratio ^c	95% Confidence interval ^d
8q23.3						
<i>Rs16892766: A>C</i>						
At-risk allele: C						
	AA	311 (88.8)	832 (80.8)		1	
	CA	36 (10.3)	186 (18.1)	0.0032**	1.93	1.32–2.82
	CC	3 (0.9)	11 (1.1)		1.38	0.38–5.00
	AA	311 (88.8)	832 (80.8)		1	
Estimated MAF ^d	CA and CC	39 (11.2)	197 (19.2)	0.0007**	1.88	1.30–2.72
Reported MAF ^{e,5,8}	0.06					
	0.07–0.09					
8q24.21						
<i>Rs6983267: G>T</i>						
At-risk allele: G						
	TT	83 (23.7)	204 (19.8)		1	
	TG	168 (48.0)	509 (49.5)	0.3252	1.23	0.90–1.68
	GG	99 (28.3)	316 (30.7)		1.27	0.91–1.79
	TT	83 (23.7)	204 (19.8)		1	
Estimated MAF	TG and GG	251 (76.3)	825 (80.2)	0.1390	1.25	0.93–1.67
Reported MAF ^{5,8}	0.48					
	0.48–0.52					
10p14						
<i>Rs10795668: G>A</i>						
At-risk allele: G						
	AA	38 (10.9)	89 (8.7)		1	
	GA	143 (40.8)	395 (38.4)	0.2561	1.17	0.76–1.79
	GG	169 (48.3)	545 (52.9)		1.36	0.90–2.07
	AA	38 (10.9)	89 (8.7)		1	
Estimated MAF	GA and GG	312 (89.1)	940 (91.4)	0.2403	1.27	0.85–1.90
Reported MAF ^{5,8}	0.31					
	0.31–0.33					
11q23.1						
<i>Rs3802842: C>A</i>						
At-risk allele: C						
	AA	195 (55.7)	516 (50.1)		1	
	CA	131 (37.4)	426 (41.4)	0.1841	1.22	0.95–1.58
	CC	24 (6.9)	87 (8.5)		1.39	0.86–2.25
	AA	195 (55.7)	516 (50.1)		1	
Estimated MAF	CA and CC	155 (44.3)	513 (49.9)	0.0759	1.25	0.98–1.59
Reported MAF ⁸	0.26					
	0.27					
15q13.3						
<i>Rs4779584: T>C</i>						
At-risk allele: T						
	CC	223 (63.7)	563 (54.7)		1	
	TC	115 (32.9)	402 (39.1)	0.0116	1.35	1.04–1.79
	TT	12 (3.4)	64 (6.2)		2.07	1.10–3.92
	CC	223 (63.7)	563 (54.7)		1	
Estimated MAF	TC and TT	127 (36.3)	466 (45.3)	0.0061**	1.42	1.11–1.83
Reported MAF ^{4,8}	0.20					
	0.18–0.22					
18q21.1						
<i>Rs4939827: T>C</i>						
At-risk allele: T						
	CC	89 (25.4)	191 (18.6)		1	
	CT	157 (44.9)	493 (48.0)	0.0258	1.46	1.07–1.99
	TT	104 (29.7)	343 (33.4)		1.52	1.09–2.13
	CC	89 (25.4)	191 (18.6)		1	
Estimated MAF	CT and TT	261 (74.6)	836 (81.4)	0.007**	1.49	1.13–1.98
Reported MAF ^{5,11}	0.48					
	0.47–0.49					
<i>Rs58920878: C>G</i>						
At-risk allele: G						
	CC	110 (31.4)	243 (23.6)		1	
	GC	162 (46.3)	514 (50.0)	0.0123	1.45	1.09–1.94
	GG	78 (22.3)	271 (26.4)		1.57	1.12–2.21
	CC	110 (31.4)	243 (23.6)		1	
Estimated MAF	GC and GG	240 (68.6)	785 (76.4)	0.0035**	1.49	1.14–1.95
Reported MAF ¹⁸	0.45					
	0.43					

Bold-type figures correspond to $P < 0.05$; ** $P < 0.0071$ (significance threshold according to Bonferroni's correction for multiple testing with seven SNPs tested).

^a*Rs16892766* on 8q23.3, *rs6983267* on 8q24, *rs10795668* on 10p14, *rs3802842* on 11q23.1, *rs4779584* on 15q13.3 and *rs4939827* and *rs58920878/Novel 1* on 18q21.1.

^bAccording to Pearson's χ^2 test.

^cEstimated from unconditional logistic regression with case-control status as the response and with adjustment for sex.

^dMinor allelic frequency estimated from the study controls.

^eAccording to published GWAS studies.

Table 4 Assessment of associations between colorectal cancer risk and each of the three at-risk alleles of SNPs^a in patients selected according to the occurrence of CRC in two first-degree relatives, one being diagnosed before 61 years of age

Locus	Genotype	Controls (N = 350), no. (%)	Patients (N = 363), no. (%)	P-value ^b	Odds ratio ^c	95% Confidence interval
8q23.3 <i>Rs16892766</i>	AA	311 (88.8)	292 (80.4)	0.0031**	1	1.24–2.92
	CA and CC	39 (11.2)	71 (19.6)			
15q13.3 <i>Rs4779584</i>	CC	223 (63.7)	204 (56.2)	0.1142	1	0.94–1.74
	TC and TT	127 (36.3)	159 (43.8)			
18q21.1 <i>Rs58920878/Novel 1</i>	CC	110 (31.4)	78 (21.5)	0.0014**	1	1.24–2.46
	GC and GG	240 (68.6)	285 (78.5)			

Bold-type figures correspond to $P < 0.05$; ** $P < 0.0167$ (significance threshold according to Bonferroni's correction for multiple testing with three SNPs tested).

^a*Rs16892766* on 8q23.3, *rs4779584* on 15q13.3 and *rs58920878/Novel 1* on 18q21.1.

^bAccording to Pearson's χ^2 test.

^cEstimated from unconditional logistic regression with case-control status as the response and with adjustment for sex.

Table 5 Assessment of associations between colorectal cancer risk and each of the three at-risk alleles of SNPs^a in patients with CRC diagnosed before 51 years of age or advanced colorectal adenoma before 41 years of age

Locus	Genotype	Controls (N = 350), no. (%)	Patients (N = 887), no. (%)	P-value ^b	Odds ratio ^c	95% Confidence interval
8q23.3 <i>Rs16892766</i>	AA	311 (88.8)	711 (80.2)	0.0004**	1	1.36–2.86
	CA and CC	39 (11.2)	176 (19.8)			
15q13.3 <i>Rs4779584</i>	CC	223 (63.7)	481 (54.2)	0.0036**	1	1.13–1.89
	TC and TT	127 (36.3)	406 (45.8)			
18q21.1 <i>Rs58920878/Novel 1</i>	CC	110 (31.4)	210 (23.7)	0.0045**	1	1.13–1.96
	GC and GG	240 (68.6)	676 (76.3)			

Bold-type figures correspond to $P < 0.05$; ** $P < 0.0167$ (significance threshold according to Bonferroni's correction for multiple testing with three SNPs tested).

^a*Rs16892766* on 8q23.3, *rs4779584* on 15q13.3 and *rs58920878/Novel 1* on 18q21.1.

^bAccording to Pearson's χ^2 test.

^cEstimated from unconditional logistic regression with case-control status as the response and with adjustment for sex.

multiple primary CRCs ($n = 44$) hampered the comparison in this subgroup. Of note, independently of the inclusion criteria, the enrichment of the at-risk alleles for the three SNPs in CRC cases was confirmed in young patients using a 51-year threshold for CRC occurrence (data not shown).

We then evaluated the potential cumulative effect of the 8q23.3 on *rs16892766*, 15q13.3 on *rs4779584* and *rs58920878* on 18q21.1 at-risk alleles on CRC risk. The overall number of at-risk alleles for the three loci was obtained by summing over the three loci the number of at-risk alleles for each locus, that is, 1 for heterozygote and 2 for homozygote subjects. None of the controls or CRC patients had six at-risk alleles. As only five cases and none of the controls had five at-risk alleles, subjects with four and five at-risk alleles were grouped in one category. CRC risk increased with the number of at-risk alleles ($P < 0.0001$, trend test), with OR values > 2 for individuals harbouring at least two at-risk alleles relative to those with none up to a maximal OR of 3.88 (95% CI: 1.72–8.76) for at least four at-risk alleles relative to none (Table 6 and Figure 1). In view of our above results on the appropriateness of a dominant model, the same analysis was repeated by counting the number of at-risk genotypes, whether homozygous or

heterozygous. Similarly, CRC increased with the number of at-risk genotypes ($P < 0.0001$, trend test), with OR values > 2 for individuals harbouring at least two at-risk genotypes relative to those with none, and a maximal OR of 6.21 (95% CI: 2.67–14.42) for three at-risk genotypes (Table 6 and Figure 1). We also calculated the cumulative OR, by considering, as the reference, the most frequent number of at-risk alleles or genotypes observed in the control population, and we found again a statistical significance ($P < 0.0001$, data not shown).

DISCUSSION

This study on highly selected patients and controls confirms the association of *rs16892766* on 8q23.3, *rs4779584* on 15q13.3 and *rs58920878* on 18q21.1 at-risk alleles with CRC, with ORs in the range 1.4–1.9, thus higher than reported from GWAS studies (Table 3). In contrast, we did not replicate other previously reported associations. It could be argued that our study was underpowered. It is important to highlight here that our study was adequately powered for ORs of 2 or more and had indeed little power for ORs < 1.5 . This was a deliberate design option in view of the little clinical value of genotyping SNPs with such low levels of CRC risk.

The higher ORs we found compared with GWAS studies may be because our decision to include selected CRC cases with a high likelihood of significant genetic determinism in contrast with GWAS studies with non-selective recruitment of CRC cases. It could also be

Table 6 Assessment of associations between colorectal tumour risk and the number of at-risk alleles or genotypes at the 8q23.3, 15q13.3 and 18q21.1 loci

	Controls (N= 350), no. (%)	Patients (N= 1028), no. (%)	P-value ^a (trend test)	Odds ratio ^b	95% Confidence ^b interval
Number of at-risk alleles					
0	60 (17.1)	95 (9.2)		1	
1	131 (37.4)	335 (32.6)		1.62	1.10–2.37
2	117 (33.4)	391 (38.0)	<0.0001	2.09	1.43–3.07
3	34 (9.7)	157 (15.3)		2.87	1.76–4.70
≥4 ^c	8 (2.3)	50 (4.9)		3.88	1.72–8.76
Number of at-risk genotypes					
0	60 (17.1)	95 (9.2)		1	
1	181 (51.7)	489 (47.6)		1.71	1.18–2.46
2	102 (29.1)	374 (36.4)	<0.0001	2.29	1.55–3.38
3	7 (2.1)	70 (6.8)		6.21	2.67–14.42

Bold-type figures correspond to $P < 0.0001$ using the Cochran-Armitage test of trend.

^aAccording to Pearson's χ^2 test.

^bEstimated from unconditional logistic regression with case/control status as the response and with adjustment for sex.

^cCorresponding to 4 or 5 at-risk alleles.

argued that these higher OR values might be because 'super-controls', without any personal or familial history among their first-degree relatives of CRC, were used in this study. Nevertheless, as indicated in Table 3, for all seven SNPs considered, we observed in our control sample a MAF (minor allele frequency) in excellent agreement with previous estimates obtained on large Caucasian populations, which indicates that there was probably no bias in our control sample. A recent study, based on a meta-analysis of previous GWAS studies and high coverage sequencing of the at-risk loci, showed that the association of these GWAS tag SNPs with CRC is not due to the combined effects of rare causal variants but reflects linkage disequilibrium with a common variant located at a vicinity of a gene involved in CRC.¹² That study also led to the identification of SNPs more strongly associated with CRC than the original tag SNPs, with ORs between 1.08 and 1.61.¹² It is therefore possible that genotyping of these recently identified top SNPs would have allowed us to obtain even larger OR than those we observed.

An enrichment of the 15q13.3 and 18q21.1 at-risk alleles had already been reported in a study performed on 995 Dutch selected CRC cases and 1340 controls⁸ with OR estimated to be 1.45 and 1.23, respectively. The *rs4779584* at-risk allele on 15q13 was also recently found enriched in a series of 252 genetically unexplained index patients with >10 colorectal adenomas *versus* 745 controls, and interestingly, the observed OR (1.5) was also higher compared with that reported in GWAS studies.²⁰ Similar to the Dutch study,⁸ we could not replicate the previously reported association with the 10p14 SNP (Table 3), *rs10795668*, possibly because of the lack of statistical power to detect low ORs, as previously indicated. Our study and that

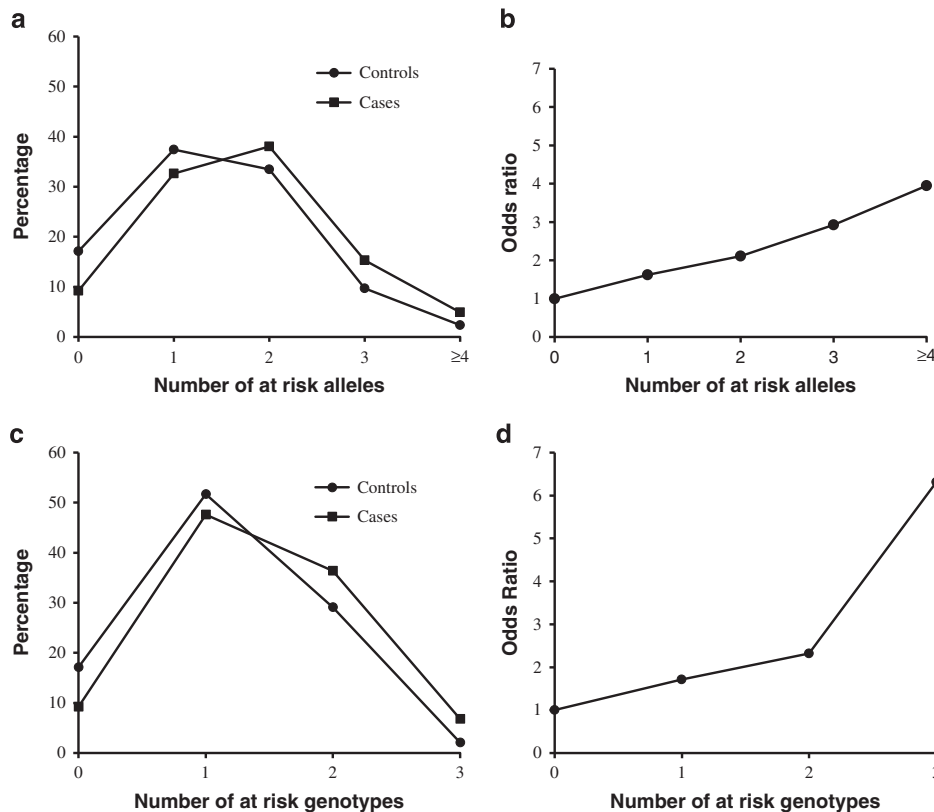


Figure 1 Number of at-risk alleles or genotypes in cases and controls. (a) Distribution of the number of at-risk alleles in cases and controls. (b) OR according to the number of at-risk alleles. (c) Distribution of the number of at-risk genotypes in cases and controls. (d) OR according to the number of at-risk genotypes.

of Middeldorp *et al*⁸ found different results for 8q23.3, 8q24.21 and 11q23.1 SNPs as, in the Dutch study, the association with 8q23.3 *rs16892766* did not remain significant after correction for multiple testing, whereas the associations with the 8q24.21 and 11q23.1 at-risk alleles were highly significant. This apparent discrepancy between both studies might be explained by the fact that our study was underpowered to detect these associations or by a difference in the selection of the patients or in the genetic background between the French and Dutch samples. Both studies found that CRC risk increased with the number of at-risk alleles, which supports an oligogenic determinism of CRC. This oligogenic determinism of CRC might explain a fraction of the clinical situations suggestive of an increased genetic risk but not explained by a deleterious mutation within a key CRC gene (ie, of the Mendelian type). A large international study,²¹ performed on over 40 000 individuals mainly from North Europe origin, had previously shown that it should be possible to identify population subgroups with increased risk for CRC, by combining SNP genotyping and familial history. Our study, on a limited number of highly selected cases and controls, confirms the clinical relevance of 8q23.3, 15q13.3 and 18q21.1 SNP genotyping. Indeed, the OR values that we found in individuals harboring at least two at-risk alleles or genotypes are, to our knowledge, the highest values reported, and the level of increased risk (OR > 2) is of the same magnitude than that estimated for first relatives of patients with CRC;²² these patients benefit from CRC detection based either on faecal immunochemical testing or colonoscopy every 5 years, starting 5 years before the age of tumour onset in the index case. As our overall finding that the presence of at least two at-risk alleles or genotypes was associated with a substantially increased risk of CRC also applied specifically to CRC cases diagnosed before 51 years of age, our study suggests that this genotyping might be considered from 40 years of age.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the French National Cancer Institute (INCa-DGOS_1332).

AUTHOR CONTRIBUTIONS

SB: coordination of the study, data analysis, writing the manuscript; FC: genotyping and data control; EH: statistical analyses; LI: genotyping; JM: patient recruitment; MB: genotyping; CA, DM, JD, FD, CC, PJ, SL, CD, BB, LF, DL, PG, IC, RG, PB, SM, EC, FP, PL-P, ML, PJ, JC, OC, M-EM-M, FP-O, SG, AZ, DB, LM, VB, SEC, FD, MG-V, J-CS, M-AC-R, LB, QW, BB-dP, J-MR, CT, M-PB, MB, JS, AH: patient recruitment; A-FC, HS: control recruitment; JB: coordination of the study and patient data control; JT: patient recruitment and coordination of the study; JB: statistical analyses and writing the manuscript; TF: planning, overseeing of the study, data analysis and writing the manuscript.

- 2 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**: 1315–1317.
- 3 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**: 984–988.
- 4 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**: 26–28.
- 5 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**: 623–630.
- 6 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**: 631–637.
- 7 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**: 1426–1435.
- 8 Middeldorp A, Jagmohan-Changur S, van Eijk R *et al*: Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 3062–3067.
- 9 Houlston RS, Cheadle J, Dobbins SE *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.
- 10 Tomlinson IP, Carvajal-Carmona LG, Dobbins SE *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.
- 11 Song Q, Zhu B, Hu W *et al*: A common SMAD7 variant is associated with risk of colorectal cancer: evidence from a case-control study and a meta-analysis. *PLoS One* 2012; **7**: e33318.
- 12 Whiffin N, Dobbins SE, Hosking FJ *et al*: Deciphering the genetic architecture of low-penetrance susceptibility to colorectal cancer. *Hum Mol Genet* 2013; **22**: 5075–5082.
- 13 Pittman AM, Naranjo S, Jalava SE *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.
- 14 Berndt SI, Potter JD, Hazra A *et al*: Pooled analysis of genetic variation at chromosome 8q24 and colorectal neoplasia risk. *Hum Mol Genet* 2008; **17**: 2665–2672.
- 15 Curtin K, Lin WY, George R *et al*: Colorectal Cancer Study Group: Meta association of colorectal cancer confirms risk alleles at 8q24 and 18q21. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 616–621.
- 16 Tuupanen S, Turunen M, Lehtonen R *et al*: The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 2009; **41**: 885–890.
- 17 Jaeger E, Leedham S, Lewis A *et al*: Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nat Genet* 2012; **44**: 699–703.
- 18 Pittman AM, Naranjo S, Webb E *et al*: The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res* 2009; **19**: 987–993.
- 19 Qin Q, Liu L, Zhong R *et al*: The genetic variant on chromosome 10p14 is associated with risk of colorectal cancer: results from a case-control study and a meta-analysis. *PLoS One* 2013; **22**: e64310.
- 20 Hes FJ, Ruano D, Nieuwenhuis M *et al*: Colorectal cancer risk variants on 11q23 and 15q13 are associated with unexplained adenomatous polyposis. *J Med Genet* 2014; **51**: 55–60.
- 21 Dunlop MG, Tenesa A, Farrington SM *et al*: Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut* 2013; **62**: 871–881.
- 22 Butterworth AS, Higgins JP, Pharoah P: Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. *Eur J Cancer* 2006; **42**: 216–227.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

1 Patel SG, Ahnen DJ: Familial colon cancer syndromes: an update of a rapidly evolving field. *Curr Gastroenterol Rep* 2012; **14**: 428–438.