

RESEARCH ARTICLE

Open Access

# Case-control study for colorectal cancer genetic susceptibility in EPICOLON: previously identified variants and mucins

Anna Abulí<sup>1,2</sup>, Ceres Fernández-Rozadilla<sup>3</sup>, Virginia Alonso-Espinaco<sup>1</sup>, Jenifer Muñoz<sup>1</sup>, Victoria Gonzalo<sup>1</sup>, Xavier Bessa<sup>2</sup>, Dolors González<sup>4</sup>, Joan Clofent<sup>5</sup>, Joaquin Cubiella<sup>6</sup>, Juan D Morillas<sup>7</sup>, Joaquim Rigau<sup>8</sup>, Mercedes Latorre<sup>9</sup>, Fernando Fernández-Bañares<sup>10</sup>, Elena Peña<sup>11</sup>, Sabino Riestra<sup>12</sup>, Artemio Payá<sup>13</sup>, Rodrigo Jover<sup>13</sup>, Rosa M Xicola<sup>14</sup>, Xavier Llor<sup>14</sup>, Luis Carvajal-Carmona<sup>15</sup>, Cristina M Villanueva<sup>16</sup>, Victor Moreno<sup>17</sup>, Josep M Piqué<sup>1</sup>, Angel Carracedo<sup>3</sup>, Antoni Castells<sup>1</sup>, Montserrat Andreu<sup>2</sup>, Clara Ruiz-Ponte<sup>3</sup> and Sergi Castellví-Bel<sup>1\*</sup>, for the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association<sup>18</sup>

## Abstract

**Background:** Colorectal cancer (CRC) is the second leading cause of cancer death in developed countries. Familial aggregation in CRC is also important outside syndromic forms and, in this case, a polygenic model with several common low-penetrance alleles contributing to CRC genetic predisposition could be hypothesized. Mucins and GALNTs (N-acetylgalactosaminyltransferase) are interesting candidates for CRC genetic susceptibility and have not been previously evaluated. We present results for ten genetic variants linked to CRC risk in previous studies (previously identified category) and 18 selected variants from the mucin gene family in a case-control association study from the Spanish EPICOLON consortium.

**Methods:** CRC cases and matched controls were from EPICOLON, a prospective, multicenter, nationwide Spanish initiative, comprised of two independent stages. Stage 1 corresponded to 515 CRC cases and 515 controls, whereas stage 2 consisted of 901 CRC cases and 909 controls. Also, an independent cohort of 549 CRC cases and 599 controls outside EPICOLON was available for additional replication. Genotyping was performed for ten previously identified SNPs in *ADH1C*, *APC*, *CCDN1*, *IL6*, *IL8*, *IRS1*, *MTHFR*, *PPARG*, *VDR* and *ARL11*, and 18 selected variants in the mucin gene family.

**Results:** None of the 28 SNPs analyzed in our study was found to be associated with CRC risk. Although four SNPs were significant with a  $P$ -value  $< 0.05$  in EPICOLON stage 1 [rs698 in *ADH1C* (OR = 1.63, 95% CI = 1.06-2.50,  $P$ -value = 0.02, recessive), rs1800795 in *IL6* (OR = 1.62, 95% CI = 1.10-2.37,  $P$ -value = 0.01, recessive), rs3803185 in *ARL11* (OR = 1.58, 95% CI = 1.17-2.15,  $P$ -value = 0.007, codominant), and rs2102302 in *GALNTL2* (OR = 1.20, 95% CI = 1.00-1.44,  $P$ -value = 0.04, log-additive 0, 1, 2 alleles)], only rs3803185 achieved statistical significance in EPICOLON stage 2 (OR = 1.34, 95% CI = 1.06-1.69,  $P$ -value = 0.01, recessive). In the joint analysis for both stages, results were only significant for rs3803185 (OR = 1.12, 95% CI = 1.00-1.25,  $P$ -value = 0.04, log-additive 0, 1, 2 alleles) and borderline significant for rs698 and rs2102302. The rs3803185 variant was not significantly associated with CRC risk in an external cohort (MCC-Spain), but it still showed some borderline significance in the pooled analysis of both cohorts (OR = 1.08, 95% CI = 0.98-1.18,  $P$ -value = 0.09, log-additive 0, 1, 2 alleles).

\* Correspondence: sbel@clinic.ub.es

<sup>1</sup>Department of Gastroenterology, IDIBAPS, Hospital Clínic, CIBERehd, University of Barcelona, Barcelona, Catalonia, Spain

Full list of author information is available at the end of the article

**Conclusions:** *ARL11*, *ADH1C*, *GALNTL2* and *IL6* genetic variants may have an effect on CRC risk. Further validation and meta-analyses should be undertaken in larger CRC studies.

**Keywords:** Colorectal Neoplasms, Genetic Predisposition to Disease, Single Nucleotide Polymorphism, Mucins, Genetic Association Studies

## Background

Colorectal cancer (CRC) is the second leading cause of cancer death in developed countries [1]. Familial adenomatous polyposis and Lynch syndrome are the most frequent hereditary CRC syndromes with a more aggressive presentation, earlier onset and strong familial aggregation. However, they only correspond to a minority of the total CRC burden (~5%). Most genetic components involved in these less frequent hereditary forms were successfully identified in the past two decades and they correspond to rare, highly penetrant alleles that predispose to CRC. Genetic association analyses have been the strategy to identify predisposing CRC alleles in the last decade, firstly by studying a small number of such variants or single nucleotide polymorphisms (SNP) using candidate-gene approaches [2], and lately with an unbiased strategy by genome-wide association studies (GWAS) [3,4].

From the EPICOLON consortium [5], several genetic association candidate-gene efforts have been pursued since 2005 aiming to identify genetic susceptibility variants for CRC. In this manner, SNPs/genes were selected to be studied from the previously identified category (variants linked to CRC risk in previous studies), from human syntenic CRC susceptibility regions identified in mouse, from the CRC carcinogenesis-related pathways Wnt and BMP, from regions 9q22 and 3q22 with positive linkage in CRC families, and from the mucin gene family [6-8].

SNPs in genes selected in the previously identified category (*ADH1C*, *APC*, *CCDN1*, *IL6*, *IL8*, *IRS1*, *MTHFR*, *PPARG*, *VDR* and *ARL11*) have been analyzed in previous independent genetic association studies and they are *a priori* attractive candidates for genetic susceptibility to CRC [6,9-24]. The expression of most genes in this category are altered in CRC and they appear to be involved in important processes for CRC risk such as hereditary CRC (*APC*), alcohol metabolism (*ADH1C*), inflammation (*IL6*, *IL8*), cell cycle regulation (*CCDN1*), energy balance (*IRS1*, *PPARG*), methylation (*MTHFR*), vitamin D (*VDR*), and the RAS superfamily (*ARL11*). On the other hand, mucins are protein constituents of the mucous barrier that protects human epithelia from adverse conditions, and they are highly glycosylated by GALNT proteins. Mucins and GALNTs, members of the mucin gene family, can be found deregulated in CRC and other neoplasms and, although

also interesting candidates, they have not been previously evaluated for CRC genetic susceptibility [25].

Here, we report results from a case-control association study for CRC risk in the EPICOLON cohort for previously identified SNPs in *ADH1C*, *APC*, *CCDN1*, *IL6*, *IL8*, *IRS1*, *MTHFR*, *PPARG*, *VDR* and *ARL11*, and selected variants within the mucin gene family.

## Methods

### Study populations

Studied subjects were mainly from EPICOLON, a prospective, multicenter, nationwide Spanish initiative [http://www.aegastro.es/aeg/ctl\\_servlet?\\_f=16&grupo=4](http://www.aegastro.es/aeg/ctl_servlet?_f=16&grupo=4) [5], comprised of two stages (2000-2001 and 2006-2008), where CRC cases and matched healthy controls were collected. DNA samples were extracted as previously described [6,7]. EPICOLON stage 1 corresponded to 515 CRC cases and 515 controls, whereas EPICOLON stage 2 consisted of 901 CRC cases and 909 controls. Additionally, an independent cohort of 549 CRC cases and 599 controls (MCC-Spain; <http://www.creal.cat>) was available for further replication of putative positive hits in EPICOLON. Cases and controls were matched for sex and age ( $\pm 5$  years) and controls were negative for personal and family cancer history. All samples were obtained with informed consent reviewed by the ethical board of the corresponding hospital.

### Gene and SNP selection

Selected SNPs were included in the previously identified category if they corresponded to genetic variants linked to CRC risk by previously published independent studies, or in the mucin gene family if located on genes encoding for mucins and GALNT proteins. Mucins are protein constituents of the mucous barrier highly glycosylated by GALNT proteins. One relevant previously identified SNP was studied in each of the following genes: *ADH1C*, *APC*, *CCDN1*, *IL6*, *IL8*, *IRS1*, *MTHFR*, *PPARG*, *VDR* and *ARL11*. For the mucin gene family, 1-2 SNPs were selected from each of the following genes: *MUC7*, *MUC12*, *MUC13*, *MUC15*, *MUC16*, *MUC17*, *MUC19*, *MUC21*, *GALNT1*, *GALNTL2*, *GALNT10*, *GALNT14*, *GALNT4* and *OVGP1*. Mucin gene selection did not intentionally include *GALTN12*. This gene is located on the 9q22 region and it was studied in our candidate-gene approach for regions with positive linkage in CRC families (Abulí et al., manuscript in

preparation). SNP selection in the mucin gene family was performed using Pupasuite, a web tool for the selection of genetic variants with potential phenotypic effect ([pupasuite.bioinfo.cipf.es](http://pupasuite.bioinfo.cipf.es)) [26]. SNPs were always prioritized if they were coding, evolutionary conserved in mouse and SNP minor allele frequency (MAF) was above 5%. Other selected SNPs with a putative regulatory effect were in promoter, intronic or 3'-UTR regions. A complete list of SNP and genes analyzed in the present study is detailed in Table 1.

High-throughput genotyping in EPICOLON cohorts was performed according to manufacturer's instructions with the TaqMan allelic discrimination and SNPlex™ systems (Applied Biosystems, Foster City, USA), and single-base primer extension chemistry matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) genotyping platform (Sequenom Inc., San Diego, USA). Genotyping in the MCC-Spain cohort was performed using

the VeraCode technology (Illumina, San Diego, USA). Genotyping was performed at the Santiago de Compostela and Barcelona nodes of the Spanish National Genotyping Centre <http://www.cegen.org>, and at the Genome Analysis Platform of the CIC-BioGUNE <http://www.cicbiogune.es>.

#### Statistical methods

As quality control, genotyping success was set above 90% for SNPs. Allelic frequency description and Hardy-Weinberg equilibrium test were performed using SNPator, a web-based tool offered by the Bioinformatics division of the Spanish National Genotyping Centre ([bioinformatica.cegen.upf.es](http://bioinformatica.cegen.upf.es)) [27]. All SNPs analyzed had a genotype success rate > 90%. The genotype frequencies of all variants in the control population fitted the Hardy-Weinberg equilibrium ( $P > 0.01$ ), supporting absence of genotyping artifacts. There was no sign of underlying population stratification in EPICOLON as

**Table 1 SNPs genotyped in EPICOLON from the previously identified category and the mucin gene family to evaluate their implication in CRC genetic susceptibility.**

Gene	SNP ID	Chr	Position*	Category	SNP type	Alleles
<i>MTHFR</i>	rs1801133	1	11856378	Previously identified	Missense (A222V)	C/T
<i>IRS1</i>	rs1801278	2	227660544	Previously identified	Missense (G971R)	A/G
<i>PPARG</i>	rs1801282	3	12393125	Previously identified	Missense (P12A)	C/G
<i>IL8</i>	rs4073	4	74606024	Previously identified	Promoter	A/T
<i>ADH1C</i>	rs698	4	100260789	Previously identified	Missense (I350V)	A/G
<i>APC</i>	rs459552	5	112176756	Previously identified	Missense (V1822D)	A/T
<i>IL6</i>	rs1800795	7	22766645	Previously identified	Promoter	C/G
<i>CCDN1</i>	rs9344	11	69462910	Previously identified	Synonymous (P241P)	A/G
<i>VDR</i>	rs2228570	12	48272895	Previously identified	Missense (M1T)	C/T
<i>ARL11</i>	rs3803185	13	50205025	Previously identified	Missense (C158R)	A/G
<i>OVGP1</i>	rs10067	1	111957311	Mucin family	Missense (H604Q)	C/G
<i>GALNT14</i>	rs2288101	2	31135184	Mucin family	Missense (Q469K)	A/C
<i>GALNT14</i>	rs11676188	2	31352788	Mucin family	Intronic	C/G
<i>GALNTL2</i>	rs2102302	3	16215650	Mucin family	Promoter	A/G
<i>MUC13</i>	rs12732	3	124624568	Mucin family	3'-UTR	C/T
<i>MUC13</i>	rs4679392	3	124646594	Mucin family	Missense (I99T)	A/G
<i>MUC7</i>	rs6826961	4	71346701	Mucin family	Missense (N80K)	C/G
<i>GALNT10</i>	rs6580076	5	153783753	Mucin family	Synonymous (A382A)	C/T
<i>GALNT10</i>	rs2277937	5	153799165	Mucin family	3'-UTR	C/T
<i>MUC21</i>	rs1634730	6	30954245	Mucin family	Missense (V98A)	C/T
<i>MUC12</i>	rs11766125	7	100648117	Mucin family	Missense (T4758R)	C/G
<i>MUC17</i>	rs4729656	7	100701610	Mucin family	3'-UTR	A/T
<i>MUC15</i>	rs15783	11	26586801	Mucin family	Missense (T229I)	A/G
<i>MUC15</i>	rs11029621	11	26596998	Mucin family	3' near gene	A/G
<i>MUC19</i>	rs2933353	12	40857943	Mucin family	Missense (G1803A)	A/C
<i>GALNT4</i>	rs2230283	12	89916811	Mucin family	Missense (V506I)	A/G
<i>GALNT1</i>	rs17647532	18	33234072	Mucin family	Promoter	C/T
<i>MUC16</i>	rs1862458	19	9069792	Mucin family	Missense (S5885F)	A/G

SNP ID, single nucleotide polymorphism identification.

Chr: Chromosome; UTR: untranslated region. \*According to NCBI build 37.1 available at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp>.

tested by an independent study [7]. Genotype analysis was carried out using the SNPAssoc R library [28]. Inter-group comparisons of genotype frequency differences were performed by regression analysis for codominant, dominant, recessive and log-additive models of inheritance. We estimated the crude odds ratio (OR) and their 95% confidence intervals (95% CIs). As expected, results did not change after sex and age adjustment. The best genetic or inheritance model was selected using the Akaike information criteria. To address the issue of multiple testing, we used Bonferroni correction ( $P = 0.0125$  for four SNPs). Study power was estimated with CaTS software [29].

### Results

Twenty-eight SNPs, ten from the previously identified category and 18 from the mucin gene family, were successfully genotyped in EPICOLON stage 1. Results in EPICOLON stage 1 are shown in Additional File 1. Four SNPs were significant with a  $P$ -value  $< 0.05$  in any of the tested inheritance models: rs698 in *ADH1C* (OR = 1.63, 95% CI = 1.06-2.50,  $P$ -value = 0.02, recessive, AA-AG vs GG), rs1800795 in *IL6* (OR = 1.62, 95% CI = 1.10-2.37,  $P$ -value = 0.01, recessive, GG-GC vs CC), rs3803185 in *ARL11* (OR = 1.58, 95% CI = 1.17-2.15,  $P$ -value = 0.007, codominant, GG vs AG vs GG), and rs2102302 in *GALNTL2* (OR = 1.20, 95% CI = 1.00-1.44,  $P$ -value = 0.04, log-additive 0, 1, 2 alleles).

In the EPICOLON stage 1 cohort, a liberal  $P$ -value threshold ( $P$ -value  $< 0.05$ ) was used to avoid false-negative results. We then validated statistically-significant stage 1 results by replicating them in another independent CRC cohort (EPICOLON stage 2). Further replication in stage 2 was performed only for significant SNPs in stage 1. Results for rs698, rs1800795, rs3803185, and rs2102302 in EPICOLON stage 2 are shown in Additional File 2. Only rs3803185 maintained statistical significance in stage 2 (OR = 1.34, 95% CI = 1.06-1.69,  $P$ -value = 0.01, recessive model, AA-AG vs GG). In order to improve statistical power, results for EPICOLON stages 1 and 2 were also analyzed jointly for these four SNPs. Results were only significant for rs3803185 (OR = 1.12, 95% CI = 1.00-1.25,  $P$ -value = 0.04, log-additive 0, 1, 2 alleles) and borderline significant for rs698 and rs2102302 (Table 2).

In order to further investigate rs3803185, we were able to genotype it additionally outside EPICOLON in a cohort from a Spanish multicase-control population study for common neoplasms (MCC-Spain). This variant was not significantly associated with CRC risk in this independent cohort. Pooled analysis for rs3803185 in the EPICOLON and MCC-Spain cohorts still showed borderline significance (OR = 1.08, 95% CI = 0.98-1.18,  $P$ -value = 0.09, log-additive 0, 1, 2 alleles).

**Table 2 SNPAssoc results for previously identified and mucin SNPs in EPICOLON cohorts.**

EPICOLON stages 1+2 (1,416 CRC cases and 1,424 controls)								
<b>ADH1C</b>								
<b>rs698</b>	Controls	%	Cases	%	OR	lower	upper	$P$ -value
Codominant								
A/A	629	46.1	616	44.2	1.00			0.13565
A/G	610	44.7	620	44.4	1.04	0.89	1.22	
G/G	125	9.2	159	11.4	1.30	1.00	1.69	
Dominant								
A/A	629	46.1	616	44.2	1.00			0.29181
A/G-G/G	735	53.9	779	55.8	1.08	0.93	1.26	
Recessive								
A/A-A/G	1239	90.8	1236	88.6	1.00			0.05243
G/G	125	9.2	159	11.4	1.28	1.00	1.63	
Log-Additive								
0, 1, 2	1364	49.4	1395	50.6	1.10	0.98	1.24	0.09043
<b>IL6</b>								
<b>rs1800795</b>	Controls	%	Cases	%	OR	lower	upper	$P$ -value
Codominant								
G/G	593	42.7	586	41.7	1.00			0.7950
G/C	623	44.9	635	45.2	1.03	0.88	1.21	
C/C	172	12.4	184	13.1	1.08	0.85	1.37	
Dominant								
G/G	593	42.7	586	41.7	1.00			0.5862
G/C-C/C	795	57.3	819	58.3	1.04	0.90	1.21	
Recessive								
G/G-G/C	1216	87.6	1221	86.9	1.00			0.5763
C/C	172	12.4	184	13.1	1.07	0.85	1.33	
log-Additive								
0, 1, 2	1388	49.7	1405	50.3	1.04	0.93	1.16	0.5034
<b>ARL11</b>								
<b>rs3803185</b>	Controls	%	Cases	%	OR	lower	upper	$P$ -value
Codominant								
A/A	359	26.9	301	23.7	1.00			0.12263
A/G	686	51.4	666	52.4	1.15	0.95	1.39	
G/G	289	21.7	304	23.9	1.26	1.01	1.57	
Dominant								
A/A	359	26.9	301	23.7	1.00			0.06417
A/G-G/G	975	73.1	970	76.3	1.18	0.99	1.41	
Recessive								
A/A-A/G	1045	78.3	967	76.1	1.00			0.15480
G/G	289	21.7	304	23.9	1.14	0.95	1.37	
log-Additive								
0, 1, 2	1334	51.2	1271	48.8	1.12	1.00	1.25	<b>0.04338</b>
<b>GALNTL2</b>								
<b>rs2102302</b>	Controls	%	Cases	%	OR	lower	upper	$P$ -value
Codominant								
A/A	579	42.1	538	39.0	1.00			0.16056
A/G	607	44.1	624	45.3	1.11	0.94	1.30	



**Table 2 SNPAssoc results for previously identified and mucin SNPs in EPICOLON cohorts. (Continued)**

G/G	189	13.7	217	15.7	1.23	0.98	1.55	
Dominant								
A/A	579	42.1	538	39.0	1.00			0.0971
A/G-G/G	796	57.9	841	61.0	1.14	0.98	1.32	
Recessive								
A/A-A/G	1186	86.3	1162	84.3	1.00			0.14317
G/G	189	13.7	217	15.7	1.17	0.95	1.45	
log-Additive								
0, 1, 2	1375	49.9	1379	50.1	1.11	1.00	1.24	0.0558

Significant *P*-values (< 0.05) are highlighted in bold.  
 OR, odds ratio.

Associations of rs698, rs1800795, rs3803185 and rs2102302 were also evaluated in 2 cohorts described in a previous GWAS [30], either by checking the original variant or a proxy SNP highly correlated with it ( $r^2 > 0.7$ ) (Table 3). Interestingly, rs3803185 showed again significance in one of the GWAS ( $P = 0.03$ ). However, it should be commented that weak associations observed in our study would have not been present if Bonferroni correction for multiple testing was applied and, therefore, they should be considered as not statistically significant.

## Discussion

Ten previously identified SNPs in *ADH1C*, *APC*, *CCDN1*, *IL6*, *IL8*, *IRS1*, *MTHFR*, *PPARG*, *VDR* and *ARL11*, and 18 selected variants in the mucin gene family were evaluated with a genetic association strategy. CRC cases and matched controls collected in two independent stages within the EPICOLON consortium were genotyped in order to evaluate its potential association with CRC risk. Mucins and GALNTs, members of the mucin gene family, have not been previously evaluated for CRC genetic susceptibility.

Significant results for previously identified SNPs in independent cohorts regarding CRC risk were previously reported for *APC* [11,12], *CCDN1* [13,14], *IL6* and *IL8* [15,16], *IRS1* [17], *MTHFR* [18-20], *PPARG* [21], and *VDR* [22,23]. Remarkably, variants in *MTHFR* and *CCDN1* were additionally supported by meta-analyses of

several individual studies [31]. However, none of them was confirmed in our cohort.

In our study, four SNPs were significant in EPICOLON stage 1 (rs698 in *ADH1C*, rs1800795 in *IL6*, rs3803185 in *ARL11*, and rs2102302 in *GALNTL2*), but only rs3803185 achieved statistical significance in EPICOLON stage 2. In the joint analysis for both stages, results were only significant for rs3803185 and borderline significant for rs698 and rs2102302. The rs3803185 variant was not significantly associated with CRC risk in an external cohort (MCC-Spain), but it still showed some borderline significance in the pooled analysis of both cohorts.

*ARL11*, also known as *ARLTS1* (ADP-ribosylation factor-like tumor suppressor gene 1), is a tumor suppressor gene that belongs to the ARF family of the Ras superfamily of small GTPases that are known to be involved in multiple regulatory pathways altered in human carcinogenesis. It has been suggested that *ARL11* SNPs, especially rs3803185, may act as low penetrance variants in several neoplasms including CRC [6,24,32]. The *ADH1C* gene encodes for class I alcohol dehydrogenase, gamma subunit, which is a member of the alcohol dehydrogenase family. Members of this enzyme family metabolize a wide variety of substrates, including ethanol. There is a noticeable association between alcohol consumption and CRC risk [33], and it is plausible that this association could also be modified by germline variants in enzymes that metabolize ethanol. *GALNTL2* (UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 2), also known as *GALNT15*, is ubiquitously expressed in human tissues [34]. This gene has never been investigated in CRC susceptibility and it has only been reported to be studied as a genetic factor involved in longevity [35]. Chronic inflammation is in the etiology of CRC and release of large amount of cytokines and growth factors may influence the carcinogenesis process. The *IL6* (interleukin 6) gene, among others, has been analyzed in previous independent genetic association studies and it is an *a priori* attractive candidate for genetic susceptibility to CRC [15,16].

Finally, as limitations of our study, it should be commented that our cohort sample size may probably be

**Table 3 Association results for four selected SNPs evaluated in two external GWAS (CORGI and VQ58), either by checking the original variant or a proxy SNP highly correlated with it ( $r^2 > 0$ ).**

	Present in I5?	Present in I3?	Proxy	$r^2$ and $D'$	<i>P</i> -value (CORGI)	<i>P</i> -value (VQ58)
rs698	no	no	rs1789924	1.000 1.000	0.1806	0.1658
rs1800795	no	no	rs1554606	0.868 0.932	0.1743	0.3346
rs3803185	no	no	rs4942859	0.716 1.000	0.5943	<b>0.03525</b>
rs2102302	no	no	rs2730351	0.853 1.000	0.1447	0.4666

Significant *P*-values (< 0.05) are highlighted in bold.

I5, Illumina HumanHap550; I3 Illumina HumanHap300.

not large enough to reach stronger conclusions for the analyzed variants. However, our study (1,416 CRC cases and 1,424 controls for the EPICOLON cohorts) had an estimated 80% power to detect an OR as low as 1.3 with a MAF of 0.30, and 1.34 for a MAF down to 0.20, assuming a log-additive model and  $\alpha = 0.05$ . Also, since we did not follow a GWAS strategy, our results apply only for the selected SNPs. Nevertheless, gene/SNP selection was biased to include those with previously published positive association results for CRC risk or those mucin SNPs with a putative functional effect.

In summary, none of the 28 SNPs analyzed in our study could be associated with CRC risk. However, variants in *ARL11*, *ADHIC*, *GALNTL2* and *IL6* may have an effect on CRC risk. Mucins and GALNTs, included in the mucin gene family, have been found deregulated in CRC and other neoplasms, and they are interesting candidates for CRC genetic susceptibility [27]. Since they have not been previously evaluated and despite our mostly negative results, we consider that genetic variation in the mucin gene family should be further explored in larger CRC cohorts in order to draw more solid conclusions. Also, additional case-control studies in larger CRC cohorts and meta-analyses could be useful to confirm or refute the role of *ARL11*, *ADHIC*, *GALNTL2* and *IL6* variants in CRC susceptibility.

## Conclusions

None of the 28 SNPs analyzed in our study could be associated with CRC risk. However, *ARL11*, *ADHIC*, *GALNTL2* and *IL6* genetic variants may have an effect on CRC risk. Further validation and meta-analyses should be undertaken in larger CRC cohorts.

## Additional material

**Additional file 1: Results for previously identified and mucin SNPs in EPICOLON stage 1.** SNPassoc results for previously identified and mucin SNPs in EPICOLON stage 1. *P*-values for some SNPs are highlighted in bold if significant ( $P < 0.05$ ).

**Additional file 2: Results for previously identified and mucin SNPs in EPICOLON stage 2.** SNPassoc results for previously identified and mucin SNPs in EPICOLON stage 2. *P*-values for some SNPs are highlighted in bold if significant ( $P < 0.05$ ).

## Abbreviations

CRC: colorectal cancer; SNP: single nucleotide polymorphism; GWAS: genome-wide association study; GALNT: N-acetylgalactosaminyltransferase; MCC: multicase-control; MAF: minor allele frequency; MALDI-TOF MS: matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; 95% CI: 95% confidence interval; OR: odds ratio; ARLTS1: ADP-ribosylation factor-like tumor suppressor gene 1; GALNTL2: UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-like 2; Chr: Chromosome; UTR: untranslated region; I5: Illumina HumanHap550; I3: Illumina HumanHap300; AIC: Akaike information content.

## Acknowledgements and Funding

We are sincerely grateful to all patients participating in this study who were recruited in 25 (EPICOLON 1) and 14 (EPICOLON 2) Spanish hospitals as part of the EPICOLON project (a list of members is provided below). We are also grateful to the Spanish National Genotyping Center (CEGEN-ISCIII)-USC and UPF nodes, and the Genome Analysis Platform of the CIC-BioGUNE. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona. SCB is supported by a contract from the Fondo de Investigación Sanitaria (CP 03-0070). CF has obtained a FPU Fellowship from the Ministerio de Educación. VAE and JM are supported by a contract from the CIBERehd. CIBERehd, CIBERER and CIBERESP are funded by the Instituto de Salud Carlos III. This work was supported by grants from the Fondo de Investigación Sanitaria/FEDER (06/1384, 08/0024, 08/0533, 08/1276, 10/00918), Instituto de Salud Carlos III (Acción Transversal de Cáncer), Ministerio de Ciencia e Innovación (SAF2010-19273), Asociación Española contra el Cáncer (Fundación Científica y Junta de Barcelona), Fundació Olga Torres (SCB and CRP), and EU FP7 (FOOD-CT-2006-036224, CMV and CHIBCHA Consortium, LCC, ACar and SCB). Members of the EPICOLON Consortium (Gastrointestinal Oncology Group of the Spanish Gastroenterological Association)

Hospital 12 de Octubre, Madrid: Juan Diego Morillas (local coordinator), Raquel Muñoz, Marisa Manzano, Francisco Colina, Jose Díaz, Carolina Ibarrola, Guadalupe López, Alberto Ibáñez; Hospital Clínic, Barcelona: Antoni Castells (local coordinator), Virginia Piñol, Sergi Castellví-Bel, Francesc Balaguer, Victoria Gonzalo, Teresa Ocaña, María Dolores Giráldez, María Pellisé, Anna Serradesanferm, Leticia Moreira, Miriam Cuatrecasas, Josep M. Piqué; Hospital Clínic Universitario, Zaragoza: Ángel Lanás (local coordinator), Javier Alcedo, Javier Ortego; Hospital Cristal-Piñor, Complejo Hospitalario de Ourense: Joaquín Cubiella (local coordinator), M<sup>a</sup> Soledad Díez, Mercedes Salgado, Eloy Sánchez, Mariano Vega; Hospital del Mar, Barcelona: Montserrat Andreu (local coordinator), Anna Abuli, Xavier Bessa, Mar Iglesias, Agustín Seoane, Felipe Bory, Gemma Navarro, Beatriz Bellosillo; Josep M<sup>a</sup> Dedeu, Cristina Álvarez, Begoña Gonzalez; Hospital San Eloy, Baracaldo and Hospital Donostia, CIBERehd, University of Country Basque, San Sebastián: Luis Bujanda (local coordinator) Ángel Cosme, Inés Gil, Mikel Larzabal, Carlos Placer, María del Mar Ramírez, Elisabeth Hijona, Jose M. Enríquez-Navascués y Jose L. Elosegui; Hospital General Universitario de Alicante: Artemio payá (EPICOLON I local coordinator), Rodrigo Jover (EPICOLON II local coordinator), Cristina Alenda, Laura Sempere, Nuria Acame, Estefanía Rojas, Lucía Pérez-Carbonell; Hospital General de Granollers: Joaquim Rigau (local coordinator), Ángel Serrano, Anna Giménez; Hospital General de Vic: Joan Saló (local coordinator), Eduard Batiste-Alentorn, Josefina Autonell, Ramon Barniol; Hospital General Universitario de Guadalajara and Fundación para la Formación e Investigación Sanitarias Murcia: Ana María García (local coordinator), Fernando Carballo, Antonio Bienvenido, Eduardo Sanz, Fernando González, Jaime Sánchez, Akiko Ono; Hospital General Universitario de Valencia: Mercedes Latorre (local coordinator), Enrique Medina, Jaime Cuquerella, Pilar Canelles, Miguel Martorell, José Ángel García, Francisco Quiles, Elisa Orti; CHUVI-Hospital Meixoeiro, Vigo: EPICOLON I: Juan Clófent (local coordinator), Jaime Seoane, Antoni Tardío, Eugenia Sanchez. EPICOLON II M<sup>a</sup> Luisa de Castro (local coordinator), Antoni Tardío, Juan Clófent, Vicent Hernández; Hospital Universitari Germans Trias i Pujol, Badalona and Section of Digestive Diseases and Nutrition, University of Illinois at Chicago, IL, USA: Xavier Llor (local coordinator), Rosa M. Xicola, Marta Piñol, Mercè Rosinach, Anna Roca, Elisenda Pons, José M. Hernández, Miquel A. Gassull; Hospital Universitari Mútua de Terrassa: Fernando Fernández-Bañares (local coordinator), Josep M. Viver, Antonio Salas, Jorge Espinós, Montserrat Forné, María Esteve; Hospital Universitari Arnau de Vilanova, Lleida: Josep M. Reñé (local coordinator), Carmen Piñol, Juan Buenestado, Joan Viñas; Hospital Universitario de Canarias: Enrique Quintero (local coordinator), David Nicolás, Adolfo Parra, Antonio Martín; Hospital Universitario La Fe, Valencia: Lidia Argüello (local coordinator), Vicente Pons, Virginia Pertejo, Teresa Sala; Hospital Sant Pau, Barcelona: Dolors Gonzalez (local coordinator) Eva Roman, Teresa Ramon, María Poca, M<sup>a</sup> Mar Concepción, Marta Martín, Lourdes Pétriz; Hospital Xeral Cies, Vigo: Daniel Martínez (local coordinator); Fundación Pública Galega de Medicina Xenómica (FPGMX), CIBERER, Genomic Medicine Group-University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain: Ángel Carracedo (local coordinator), Clara Ruiz-Ponte, Ceres Fernández-Rozadilla, M<sup>a</sup> Magdalena Castro; Hospital Universitario Central de Asturias: Sabino Riestra (local coordinator), Luis Rodrigo; Hospital de Galdácano, Vizcaya: Javier Fernández (local coordinator), Jose Luis Cabriada;

Fundación Hospital de Calahorra (La Rioja) La Rioja: Luis Carreño (local coordinator), Susana Oquiñena, Federico Bolado; Hospital Royo Villanova, Zaragoza: Elena Peña (local coordinator), José Manuel Blas, Gloria Ceña, Juan José Sebastián; Hospital Universitario Reina Sofía, Córdoba: Antonio Naranjo (local coordinator).

#### Author details

<sup>1</sup>Department of Gastroenterology, IDIBAPS, Hospital Clínic, CIBERehd, University of Barcelona, Barcelona, Catalonia, Spain. <sup>2</sup>Gastroenterology Department, Parc de Salut Mar, Institut Municipal d'Investigació Mèdica (IMIM), Pompeu Fabra University, Barcelona, Catalonia, Spain. <sup>3</sup>Galician Public Foundation of Genomic Medicine, CIBERER, Genomics Medicine Group, Hospital Clínic, Santiago de Compostela, University of Santiago de Compostela, Galicia, Spain. <sup>4</sup>Hospital Sant Pau, Barcelona, Catalonia, Spain. <sup>5</sup>Department of Gastroenterology, Hospital Meixoeiro, Vigo, Spain. <sup>6</sup>Department of Gastroenterology, Hospital de Ourense, Galicia, Spain. <sup>7</sup>Department of Gastroenterology, Hospital 12 de Octubre, Madrid, Spain. <sup>8</sup>Department of Medicine, Hospital General de Granollers, Barcelona, Catalonia, Spain. <sup>9</sup>Department of Gastroenterology, Hospital General Universitario de Valencia, Valencia, Spain. <sup>10</sup>Department of Gastroenterology, Hospital Mutua de Terrassa, Barcelona, Catalonia, Spain. <sup>11</sup>Department of Gastroenterology, Hospital Royo Villanova, Zaragoza, Spain. <sup>12</sup>Department of Gastroenterology, Hospital Central de Asturias, Oviedo, Asturias, Spain. <sup>13</sup>Department of Pathology and Gastroenterology, Hospital General d'Alacant, Alicante, Spain. <sup>14</sup>Section of Digestive Diseases and Nutrition and Cancer Center, University of Illinois at Chicago, IL 60612, USA. <sup>15</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. <sup>16</sup>Centre for Research in Environmental Epidemiology (CREAL), IMIM (Hospital del Mar Research Institute). CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. <sup>17</sup>IDIBELL-Institut Català d'Oncologia (ICO), CIBER Epidemiología y Salud Pública (CIBERESP), University of Barcelona, L'Hospitalet de Llobregat, Barcelona, Spain. <sup>18</sup>All authors are listed in the Acknowledgements section.

#### Authors' contributions

Samples were collected within the EPICOLON by XB, DG, JCu, JCI, JDM, JR, ML, FF-B, EP, SR, RJ, XLI, VM, AC and MA and other members of this consortium. Study was designed by SCB and CRP. Data acquisition was performed by all authors. Quality control was performed by AA, CFR, VG, XB, RJ, XLL, MA and SCB. Data was analyzed and interpreted by AA, CFR, CRP and SCB. Statistical analysis was performed by AA, CV, VM and SCB. Manuscript was prepared by AA and SCB. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 22 February 2011 Accepted: 5 August 2011

Published: 5 August 2011

#### References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM: **GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide**. IARC CancerBase No. 10 Lyon, France: International Agency for Research on Cancer; 2010 [http://globocan.iarc.fr].
2. Kemp Z, Thirlwell C, Sieber O, Silver A, Tomlinson I: **An update on the genetics of colorectal cancer**. *Hum Mol Genet* 2004, **13**:R117-R185.
3. Tenesa A, Dunlop MG: **New insights into the aetiology of colorectal cancer from genome-wide association studies**. *Nat Rev Genet* 2009, **10**(6):353-8.
4. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, *et al*: **Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33**. *Nat Genet* 2010, **42**:973-7.
5. Piñol V, Castells A, Andreu M, Castellví-Bel S, Alenda C, Llor X, *et al*: **Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer**. *JAMA* 2005, **293**:1986-94.
6. Castellví-Bel S, Castells A, de Cid R, Muñoz J, Balaguer F, Gonzalo V, *et al*: **Association of the ARLTS1 Cys148Arg variant with sporadic and familial colorectal cancer**. *Carcinogenesis* 2007, **28**:1687-91.
7. Fernández-Rozadilla C, Tarrío R, Clófent J, de Castro L, Brea-Fernández A, Andreu M, *et al*: **Colorectal cancer susceptibility quantitative trait loci in mice as a novel approach to detect low-penetrance variants in humans: a two-stage case-control study**. *Cancer Epidemiol Biomarkers Prev* 2010, **19**(2):619-23.
8. Fernández-Rozadilla C, de Castro L, Clófent J, Brea-Fernández A, Bessa X, Abulí A, *et al*: **Single nucleotide polymorphisms in the Wnt and BMP pathways and colorectal cancer risk in a Spanish cohort**. *PLoS ONE* 2010, **5**(9):pii: e12673.
9. Tiemersma EW, Wark PA, Ocké MC, Bunschoten A, Otten MH, Kok FJ, *et al*: **Alcohol consumption, alcohol dehydrogenase 3 polymorphism, and colorectal adenomas**. *Cancer Epidemiol Biomarkers Prev* 2003, **12**(5):419-25.
10. van der Logt EM, Bergevoet SM, Roelofs HM, Te Morsche RH, Dijk Y, Wobbes T, *et al*: **Role of epoxide hydrolase, NAD(P)H:quinone oxidoreductase, cytochrome P450 2E1 or alcohol dehydrogenase genotypes in susceptibility to colorectal cancer**. *Mutat Res* 2006, **593**(1-2):39-49.
11. Slattery ML, Samowitz W, Ballard L, Schaffer D, Leppert M, Potter JD: **A molecular variant of the APC gene at codon 1822: its association with diet, lifestyle, and risk of colon cancer**. *Cancer Res* 2001, **61**(3):1000-4.
12. Menéndez M, González S, Blanco I, Guinó E, Peris M, Peinado MA, *et al*: **Colorectal cancer risk and the APC D1822V variant**. *Int J Cancer* 2004, **112**(1):161-3.
13. Le Marchand L, Seifried A, Lum-Jones A, Donlon T, Wilkens LR: **Association of the cyclin D1 A870G polymorphism with advanced colorectal cancer**. *JAMA* 2003, **290**(21):2843-8.
14. Tan XL, Nieters A, Kropp S, Hoffmeister M, Brenner H, Chang-Claude J: **The association of cyclin D1 G870A and E-cadherin C-160A polymorphisms with the risk of colorectal cancer in a case control study and meta-analysis**. *Int J Cancer* 2008, **122**(11):2573-80.
15. Landi S, Moreno V, Gioia-Patricola L, Guino E, Navarro M, de Oca J, *et al*: **Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer**. *Cancer Res* 2003, **63**(13):3560-6.
16. Gunter MJ, Canzian F, Landi S, Chanock SJ, Sinha R, Rothman N: **Inflammation-related gene polymorphisms and colorectal adenoma**. *Cancer Epidemiol Biomarkers Prev* 2006, **15**(6):1126-31.
17. Slattery ML, Murtaugh M, Caan B, Ma KN, Neuhausen S, Samowitz W: **Energy balance, insulin-related genes and risk of colon and rectal cancer**. *Int J Cancer* 2005, **115**(1):148-54.
18. Koushik A, Kraft P, Fuchs CS, Hankinson SE, Willett WC, Giovannucci EL, *et al*: **Nonsynonymous polymorphisms in genes in the one-carbon metabolism pathway and associations with colorectal cancer**. *Cancer Epidemiol Biomarkers Prev* 2006, **15**(12):2408-17.
19. Hubner RA, Lubbe S, Chandler I, Houlston RS: **MTHFR C677T has differential influence on risk of MSI and MSS colorectal cancer**. *Hum Mol Genet* 2007, **16**(9):1072-7.
20. Levine AJ, Figueiredo JC, Lee W, Conti DV, Kennedy K, Duggan DJ, *et al*: **Genetic variability in the MTHFR gene and colorectal cancer risk using the colorectal cancer family registry**. *Cancer Epidemiol Biomarkers Prev* 2010, **19**(1):89-100.
21. Meirhaeghe A, Amouyel P: **Impact of genetic variation of PPARgamma in humans**. *Mol Genet Metab* 2004, **83**(1-2):93-102.
22. Sweeney C, Curtin K, Murtaugh MA, Caan BJ, Potter JD, Slattery ML: **Haplotype analysis of common vitamin D receptor variants and colon and rectal cancers**. *Cancer Epidemiol Biomarkers Prev* 2006, **15**(4):744-9.
23. Slattery ML, Sweeney C, Murtaugh M, Ma KN, Caan BJ, Potter JD, *et al*: **Associations between vitamin D, vitamin D receptor gene and the androgen receptor gene with colon and rectal cancer**. *Int J Cancer* 2006, **118**(12):3140-6.
24. Frank B, Hemminki K, Brenner H, Hoffmeister M, Chang-Claude J, Burwinkel B: **ARLTS1 variants and risk of colorectal cancer**. *Cancer Lett* 2006, **244**(2):172-5.
25. Kufe DW: **Mucins in cancer: function, prognosis and therapy**. *Nat Rev Cancer* 2009, **9**(12):874-85.
26. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, *et al*: **PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes**. *Nucleic Acids Res* 2006, **34** Web Server: W621-5.
27. Morcillo-Suarez C, Alegre J, Sangros R, Gazave E, de Cid R, Milne R, *et al*: **SNP analysis to results (SNPator): a web-based environment oriented to**

- statistical genomics analyses upon SNP data. *Bioinformatics* 2008, **24**(14):1643-4.
28. González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, *et al*: SNPassoc: an R package to perform whole genome association studies. *Bioinformatics* 2007, **23**(5):644-5.
  29. Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006, **38**:209-13.
  30. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain 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.
  31. Tomlinson IP, Dunlop M, Campbell H, Zanke B, Gallinger S, Hudson T, *et al*: COGENT (COlorectal cancer GENEtics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer. *Br J Cancer* 2010, **102**(2):447-54.
  32. Yendamuri S, Trapasso F, Calin GA: ARLTS1 - a novel tumor suppressor gene. *Cancer Lett* 2008, **264**(1):11-20.
  33. Bongaerts BW, van den Brandt PA, Goldbohm RA, de Goeij AF, Weijnenberg MP: Alcohol consumption, type of alcoholic beverage and risk of colorectal cancer at specific subsites. *Int J Cancer* 2008, **123**(10):2411-7.
  34. Cheng L, Tachibana K, Iwasaki H, Kameyama A, Zhang Y, Kubota T, *et al*: Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T15. *FEBS Lett* 2004, **566**(1-3):17-24.
  35. Flachsbart F, Franke A, Kleindorp R, Caliebe A, Blanché H, Schreiber S, *et al*: Investigation of genetic susceptibility factors for human longevity - A targeted nonsynonymous SNP study. *Mutat Res* 2010, **694**(1-2):13-9.

#### Pre-publication history

The pre-publication history for this paper can be accessed here:  
<http://www.biomedcentral.com/1471-2407/11/339/prepub>

doi:10.1186/1471-2407-11-339

Cite this article as: Abulí *et al.*: Case-control study for colorectal cancer genetic susceptibility in EPICOLON: previously identified variants and mucins. *BMC Cancer* 2011 **11**:339.

Submit your next manuscript to BioMed Central  
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

