Carcinogenesis vol.34 no.2 pp.314–318, 2013
doi:10.1093/carcin/bgs357
Advance Access publication November 16, 2012

BMP2/BMP4 colorectal cancer susceptibility loci in northern and southern European populations

Ceres Fernandez-Rozadilla[1], Claire Palles[2], Luis Carvajal-Carmona[2], Paolo Peterlongo[3], Carmela Nici[3], Silvia Veneroni[4], Manuela Pinheiro[5], Manuel R.Teixeira[5,6], Victor Moreno[7], Maria-Jesus Lamas[8], Montserrat Baiget[9], Lopez-Fernandez LA[10], Dolors Gonzalez[11], Alejandro Brea-Fernandez[1], Juan Clofent[12,13], Luis Bujanda[14], Xavier Bessa[15], Montserrat Andreu[15], Rosa Xicola[16], Xavier Llor[16], Rodrigo Jover[17], The EPICOLON Consortium[†] Antoni Castells[18], Sergi Castellvi-Bel[18], Angel Carracedo[1], Ian Tomlinson[2,19] and Clara Ruiz-Ponte[1,*]

[1]Fundación Pública Galega de Medicina Xenómica (FPGMX)-Grupo de Medicina Xenómica-Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERer)-IDIS, Santiago de Compostela 15706, Spain, [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK, [3]Department of Preventive and Predictive Medicine, IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Unit of Molecular Bases of Genetic Risk and Genetic Testing, Milan 20133, Italy, [4]Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan 20133, Italy, [5]Department of Genetics, Portuguese Oncology Institute, Porto 4099-003, Portugal, [6]Biomedical Sciences Institute (ICBAS), University of Porto, Porto 4099-003, Portugal, [7]Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO), Bellvitge Biomedical Research Centre (IDIBELL), CIBERESP, Barcelona 08907, Spain, [8]Oncology Pharmacy Unit, Complexo Hospitalario Universitario de Santiago (CHUS), Santiago de Compostela 15706, Spain, [9]Department of Genetics, Hospital de la Santa Creu I Sant Pau, Universitat Autonoma de Barcelona, CIBERER U-705, Barcelona 08025, Spain, [10]Pharmacogenetics & Pharmacogenomics Laboratory, Servicio de Farmacia, Hospital General Universitario Gregorio Marañon, Madrid 28007, Spain, [11]Department of Gastroenterology, Hospital de la Santa Creu I Sant Pau, Universitat Autonoma de Barcelona, Barcelona 08025, Spain, [12]Department of Gastroenterology, Hospital do Meixoeiro, Vigo 36214, Spain, [13]Department of Internal Medicine, Section of Digestive Diseases, Hospital Sagunto, Valencia 46520, Spain, [14]Department of Gastroenterology, Donostia Hospital-Instituto Biodonostia, CIBERehd, University of the Basque Country UPV/EHU, San Sebastian 20014, Spain, [15]Department of Gastroenterology, Hospital del Mar, Barcelona 08003, Spain, [16]Section of Digestive Diseases and Nutrition, University of Illinois, Chicago, IL 60607, USA, [17]Department of Gastroenterology, Hospital General de Alicante, Alicante 03010, Spain, [18]Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd), IDIBAPS, University of Barcelona, Barcelona 08036, Spain and [19]NIHR Comprehensive Biomedical Research Centre, University of Oxford, Oxford OX3 7BN, UK

*To whom correspondence should be addressed. Tel: +34 981955195; Fax: +34 981951473;
Email: clara.ruiz.ponte@usc.es

Genome-wide association studies have successfully identified 20 colorectal cancer susceptibility loci. Amongst these, four of the signals are defined by tagging single nucleotide polymorphisms (SNPs) on regions 14q22.2 (rs4444235 and rs1957636) and 20p12.3 (rs961253 and rs4813802). These markers are located close to two of the genes involved in bone morphogenetic protein (BMP) signaling (*BMP4* and *BMP2*, respectively). By investigating these four SNPs in an initial cohort of Spanish origin, we found substantial evidence that minor allele frequencies (MAFs) may be different in northern and southern European populations. Therefore, we genotyped three additional southern European cohorts comprising a total of 2028 cases and 4273 controls. The meta-analysis results show that only one of the association signals (rs961253) is effectively replicated in the southern European populations, despite adequate power to detect all four. The other three SNPs (rs4444235, rs1957636 and rs4813802) presented discordant results in MAFs and linkage disequilibrium patterns between northern and southern European cohorts. We hypothesize that this lack of replication could be the result of differential tagging of the functional variant in both sets of populations. Were this true, it would have complex consequences in both our ability to understand the nature of the real causative variants, as well as for further study designs.

Abbreviations: BMP, bone morphogenetic protein; CRC, colorectal cancer; LD, linkage disequilibrium; MAF, minor allele frequency; OR, odds ratio; SD, standard deviation; SNP, single nucleotide polymorphisms.

[†]All members are listed in a Supplementary Note.
For the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association.

Introduction

Colorectal cancer (CRC) is one of the major forms of cancer, being the second most frequent neoplasm in both sexes and one of the most important morbidity causes in the first world (1). The genetic contribution to CRC has been estimated to be around 35% by twin studies (2). However, high-risk germline variants that cause hereditary syndromes can only explain up to 5% of the disease cases, and it seems that much of the heritable risk could be due to multiple low-penetrance alleles appearing frequently in the general population, and each conferring a modest effect on disease risk (3–6).

In this context, the implementation of association studies to genome-wide levels has successfully allowed for the identification of 20 of these low-penetrance loci associated with CRC risk in the form of tagging single nucleotide polymorphisms (SNPs) (7–9). Notably, at least four of these signals (14q22.2, 15q13.3, 20p12.3 and 20q13.3) are close to genes that are members of the bone morphogenetic protein (BMP) signaling pathway. Members of this pathway have been proven to interact with transforming growth factor-β effectors, thereby playing an essential role in colonic cell signal transduction and CRC development (10,11). It is currently believed that BMP signaling is related to CRC development through an increase in stem cell numbers near colorectal crypt bases, thus elevating the number of cells susceptible to tumor-causing mutations (12). The importance of this pathway is surely reflected on the fact that up to seven SNPs have been related to CRC susceptibility through genome-wide association studies: rs4444235 and rs1957636 with *BMP4*, rs961253 and rs4813802 with *BMP2*, rs16969681 and rs11632715 with *GREM1* and rs4939827 with *SMAD7* (7,8,13).

Thus, we genotyped the four SNPs at 14q22.2 and 20p12.3 in order to ascertain their relationship with CRC in a cohort of 1449 cases and 1450 controls of Spanish origin. These four markers have so far shown the most strongly associated signals with the CRC phenotype and, importantly, there had been no studies to the date on their relevance in southern European populations.

To our knowledge, this was the first attempt to evaluate these signals in a southern European population. Because we found substantial evidence of discrepancies at allele frequencies with the already described sample sets, we later decided to additionally genotype the four markers in three supplementary cohorts from Spain, Italy and Portugal, so as to further evaluate this potential north–south divergence.

Materials and methods

Samples and populations

EPICOLON cohort. Subjects were 1449 cases [870 males, mean age at diagnosis 73 years; standard deviation (SD) ± 0.70] and 1000 controls ascertained through the EPICOLON project and 450 additional controls

from the Spanish National DNA bank (http://www.bancoadn.org/). The EPICOLON Consortium comprises a prospective, multicentre and population-based epidemiology survey of the incidence and features of CRC in the Spanish population (14,15). Cases were selected as patients with *de novo* histologically confirmed diagnosis of colorectal adenocarcinoma. Controls were confirmed to have no cancer or history of neoplasm and no family history of CRC. They were matched with cases for hospital, sex and age (±5 years).

SPAIN2 cohort. Samples were 801 CRC patients (544 males, mean age at diagnosis 69.6 years; SD ± 0.59) recruited through a toxicology study from four different hospitals spread across the Spanish territory, and an additional 105 cases and 1287 controls (816 males, mean age 50.1 years; SD ± 7.2) from the Spanish National DNA bank.

Italian cohort. Subjects were 622 CRC cases (379 males, mean age at diagnosis 64 years; SD ± 11.3) and 2486 blood donors (907 males, mean age at donation 44 years; SD ± 12.2). The cases are from a series of consecutive individuals affected with CRC who underwent surgery at the Fondazione IRCCS Istituto Nazionale Tumori in Milan starting October 2005. The 2486 normal controls were blood donors recruited through the Immunohematology and Transfusion Medicine Service of Fondazione IRCCS Istituto Nazionale Tumori and the Associazione Volontari Italiani Sangue Comunale in Milan.

Portuguese cohort. Samples were 495 CRC and 5 colorectal adenomas patients (235 males, mean age at diagnosis 51 years; SD ± 11.81). About four-fifths of the patients were recruited according to the Bethesda or Amsterdam criteria for Lynch syndrome and had a negative genetic testing. The remaining patients were recruited with diagnosis of either CRC at <75 years old, or an 'advanced' colorectal adenomas (villous histology, or >1 cm diameter or severe dysplasia at <60 years old). The 500 controls (269 males, mean age at donation 47; SD ± 8.7) were blood donors recruited from the Portuguese Oncology Institute of Porto.

DNA was obtained from peripheral blood by standard extraction procedures. Sample collection was undertaken with informed consent and ethical review board approval of the corresponding institution, in accordance with the tenets of the Declaration of Helsinki.

SNP genotyping and quality control

The Spanish EPICOLON and SPAIN2 cohorts were genotyped using MassARRAY iPLEX technology (Sequenom, San Diego, CA). Italian (13) and Portuguese samples were genotyped with the KASPar SNP Genotyping system (KBioscience, Herts, UK). Samples with genotyping success rates <95% were removed from the study with the help of PLINK v1.7 (16). Hardy-Weinberg equilibrium in controls was also ensured (at an $\alpha < 0.05$) in order to avoid genotyping errors and detect any hidden population stratification. Duplicate samples were included in each plate of the initial stage (up to 5% of the total samples of the EPICOLON cohort) to ensure genotyping accuracy. All samples with concordance rates <100% were removed from further analyses. Additionally, HapMap samples were also included during MassARRAY genotyping to avoid genotyping errors.

MAF analysis, 1000 Genomes data, LD mapping and statistics

Welch's two-tail statistic (a modification of Student's *t*-test assuming different variances between groups), as implemented in R, was used to test for differences between minor allele frequencies (MAFs) amongst the described populations (northern populations versus EPICOLON; northern versus southern populations) (17). The same test was used to check for differences in data retrieved from the 100 Genomes Phase I (MAFs) between the Utah residents with northern and western european ancestry (CEU)/Finnish in Finland (FIN)/ British in England and Scotland (GBR) group of populations and Iberian population from Spain (IBS)/Toscans in Italy (TSI)/EPICOLON (controls) using the data from 1000 Genomes Phase I retrieved with the help of the online tool SPSmart (http://spsmart.cesga.es/) (18). Linkage disequilibrium (LD) blocks were inspected using the HapMap3 r2 CEU and TSI populations (no other European populations were available at the time) with the help of Haploview v4.2 (19,20). The ability for these four SNPs to capture the association signal was evaluated by comparing the list of markers closely related at different r^2 values (0.8 down to 0.5) for both populations. Allelic frequencies and odds ratio (OR) with 95% confidence intervals were calculated for each population with PLINK v1.7 (16). Cochran's Q statistic and I^2 coefficients were also calculated to measure heterogeneity between the southern cohorts in order to decide whether to use a fixed or random-effect meta-analysis model (21,22). Large heterogeneity was typically defined as $I^2 \geq 75\%$. Power was calculated with CaTS power calculator (23).

Data from the northern European cohorts were obtained from Tomlinson *et al.* (2011) (13). OR comparisons between northern and southern European data sets were evaluated by logistic regression with the help of the SUEST tool in STATAv12 (Stata Corp., TX, Texas).

Results

After genotyping the four SNPs (rs4444235, rs1957636, rs961253 and rs4813802) in our initial Spanish cohort, we found that MAFs in EPICOLON departed from the proportions described for several populations in the literature (Table I). The differences in MAFs were significant in all four SNPs for the case and control groups. Given these results and the fact that all the previously reported populations had a northern European origin, we decided to further explore this potential heterogeneity between northern and southern European populations for complementary evidence. By these means, we searched the 1000 Genomes data for the estimated MAFs for these markers (Table II) (18). We observe that in this case, the differences for the northern versus southern European MAFs are not statistically significant, a fact that could also be due to the low sample sizes of some of the populations.

LD blocks (±100 kb regions from each marker) were also checked for the 14q22.2 and 20p12.3 regions in the CEU and TSI populations separately. Overall, there were no great visible differences in the block patterns (taken as D′ pairwise measures) between these populations (Supplementary Figures 1–4, available at *Carcinogenesis* Online). Fine resolution LD screenings, however, revealed that at least rs4444235 and rs1957636 at 14q22.2 showed different tagging abilities in both populations (Table III).

SNP genotyping in other cohorts

With the results obtained in the initial comparisons, we decided to additionally genotype other southern European cohorts, in

Table I. MAF differences between EPICOLON and the northern European populations

SNP/locus/minor allele	Cohort	MAF cases	MAF controls
rs4444235	Tomlinson *et al.* (13) average[a]	0.484	0.457
Chr14: 54,410,919	EPICOLON	0.556	0.537
C	Welch test *P*-value (Tomlinson *et al.* versus EPICOLON)	4.251E-11*	1.084E-10*
rs1957636	Tomlinson *et al.* (13) average[a]	0.421	0.398
Chr14:54,560,018	EPICOLON	0.407	0.394
A	Welch test *P*-value (Tomlinson *et al.* versus EPICOLON)	2.729E-03*	0.01356*
rs961253	Tomlinson *et al.* (13) average[a]	0.380	0.353
Chr20:6,404,281	EPICOLON	0.350	0.333
A	Welch test *P*-value (Tomlinson *et al.* versus EPICOLON)	1.810E-05*	3.506E-03*
rs4813802	Tomlinson *et al.* (13) average[a]	0.382	0.360
Chr20:6,699,595	EPICOLON	0.319	0.321
G	Welch test *P*-value (Npops versus EPICOLON)	1.188E-09*	5.193E-07*

Frequencies in case and control groups are shown for each of the cohorts, as well as Welch's statistic *P*-value. Note that for rs4444235, frequencies are shown always for the C allele, even though it does not constitute the minor allele for some of the cohorts.

[a]Corresponds to the average values for the MAFs present in the study by Tomlinson *et al.* (13); these are UK1, Scotland1, UK2, Scotland2, VQ58, CCFR, Australia, Helsinki, Cambridge, COIN/NBS UK3, Scotland3 and UK4.

*Denotes statistically significant *P*-values.

Table II. Allele frequencies for the four SNPs in the 1000 Genomes

SNP	Chr	Position	Allele	Population	*N*	MAF	Welch *t*-test (CEU/FIN/GBR versus EPICOLON$_{CT}$[a]/IBS/TSI)
rs4444235	14	54,410,919	C	CEU	87	0.443	0.062
				FIN	93	0.446	
				GBR	88	0.466	
				IBS	14	0.643	
				TSI	98	0.546	
rs1957636	14	54,560,018	A	CEU	87	0.431	0.310
				FIN	93	0.392	
				GBR	88	0.455	
				IBS	14	0.429	
				TSI	98	0.362	
rs961253	20	6,404,281	A	CEU	87	0.437	0.506
				FIN	93	0.323	
				GBR	88	0.347	
				IBS	14	0.321	
				TSI	98	0.367	
rs4813802	20	6,699,595	G	CEU	87	0.356	0.141
				FIN	93	0.328	
				GBR	88	0.386	
				IBS	14	0.321	
				TSI	98	0.311	

Frequencies for the 1000 Genomes northern (CEU, FIN and GBR) and southern (IBS and TSI) populations.
[a]EPICOLON$_{CT}$: EPICOLON controls as depicted in Table I.

Table III. LD patterns for each of the four SNPs in HapMap CEU compared with TSI

SNP	CEU		TSI	
	r^2	Tagged SNPs	r^2	Tagged SNPs
rs4444235	>0.7	rs17563 rs2071047	>0.7	
	>0.5		>0.5	rs12587398
rs1957636	>0.8		>0.8	rs7160450
	>0.7		>0.7	rs12887156 rs1953743
	>0.6	rs7149949 rs7492415	>0.6	rs12892552
rs961253	>0.8 to >0.5	No differences	>0.8 to >0.5	No differences
rs4813802	>0.8 to >0.5	No differences	>0.8 to >0.5	No differences

Tagging relationships between rs4444235, rs1957636, rs961253 and rs4813802 for different r^2 thresholds (0.8 down to 0.5) and their relationship with LD patterns. Only r^2 values with differential tagging have been included for visualization purposes.

order to obtain more data and have more reliable results regarding the differences that we had observed. Therefore, we genotyped a total of 2028 cases and 4273 controls from three new cohorts: the Spanish SPAIN2 cohort, the Italian ITALY cohort and the Portuguese PORTUGAL cohort. The results from this second genotyping round are illustrated in Table IV. Briefly, SNPs rs1957636, rs961253 and rs4813802 showed higher frequencies in northern compared with southern European populations, whereas the opposite was observed for rs4444235. All MAF comparisons between northern and southern European populations remained significant, with the exception of the rs961253 cases group, which is probably influenced by the outlier result from the SPAIN2 cohort being similar to the values for the previous northern European populations.

Once the MAF comparisons were made, we also performed the association analyses for each of the four markers in the southern European data sets both individually and altogether. We found that the only association signal that could be effectively replicated was that of rs961253 (meta-analysis $P = 0.010$; OR = 1.091). None of the other three signals tagged by markers rs4444235, rs1957636 and rs4813802 could be replicated in the southern European data sets, despite our study being powered enough to detect them (>80%). OR from the meta-analysis of the four populations were consistent in directions with the ones described in the literature for each of the SNPs. However,

Table IV. MAF values for the four SNPs in the southern European cohorts

SNP/minor allele	Cohort	MAF cases	MAF controls
rs4444235/Allele: C	EPICOLON	0.556	0.537
	SPAIN2	0.542	0.550
	ITALY	0.544	0.527
	PORTUGAL	0.551	0.543
	Welch test *P*-value (Tomlinson *et al.*[a] versus southern[b])	1.036E-07*	2.417E-06*
rs1957636/Allele: A	EPICOLON	0.407	0.394
	SPAIN2	0.413	0.404
	ITALY	0.407	0.408
	PORTUGAL	0.410	0.390
	Welch test *P*-value (Tomlinson *et al.*[a] versus southern[b])	9.873E-03*	0.01444*
rs961253/Allele: A	EPICOLON	0.350	0.333
	SPAIN2	0.372	0.322
	ITALY	0.309	0.323
	PORTUGAL	0.356	0.331
	Welch test *P*-value (Tomlinson *et al.*[a] versus southern[b])	0.07824	8.284E-04*
rs4813802/Allele: G	EPICOLON	0.319	0.321
	SPAIN2	0.323	0.310
	ITALY	0.275	0.301
	PORTUGAL	0.329	0.298
	Welch test *P*-value (Tomlinson *et al.*[a] versus southern[b])	7.397E-03*	8.816E-05*

Allele frequencies in each EPICOLON, SPAIN2, ITALY and PORTUGAL are depicted for comparison purposes with Table I.
[a]Corresponds to the average values for the MAFs present in the study by Tomlinson *et al.* (13); these are UK1, Scotland1, UK2, Scotland2, VQ58, CCFR, Australia, Helsinki, Cambridge, COIN/NBS UK3, Scotland3 and UK4.
[b]Southern: southern European populations (EPICOLON, SPAIN2, ITALY and PORTUGAL).
*Denotes statistically significant *P*-values.

Table V. Association values for the southern European cohorts

SNP	Allele	EPICOLON (1449/1450)		SPAIN2 (906/1392)		ITALY (622/2486)		PORTUGAL (500/500)		Meta-analysis			Tomlinson *et al.* (13)	Tomlinson *et al.* (13) OR versus southern OR
		OR	*P*	OR	*P*	OR	*P*	OR	*P*	OR	I^2	*P*	OR	*P*
rs4444235	C	1.049	0.391	0.967	0.591	1.038	0.585	1.032	0.722	1.021	0.00	0.533	1.091	0.300
rs1957636	A	1.065	0.260	1.036	0.565	0.993	0.908	1.091	0.350	1.041	0.00	0.219	1.084	0.167
rs961253	A	1.078	0.197	1.252	5.656E-04*	0.942	0.388	1.109	0.255	1.091	66.83	0.010*	1.117	0.993
rs4813802	G	1.008	0.899	1.065	0.337	0.881	0.075	1.154	0.137	1.009	52.07	0.798	1.093	0.036*

P-values and ORs are shown for each population independently plus the meta-analysis of all four southern European populations. The last column corresponds to the *P*-value for OR comparisons between the northern and southern European populations for each of the markers. Numbers in brackets below the cohort correspond to number of cases and controls. The last column shows the *P*-value for the test comparing the OR between the Tomlinson *et al.* (13) report and the results from the meta-analysis of our four southern European populations.
*Denotes statistically significant *P*-values.

rs4813802 proved to have significant differences between OR in the northern and southern European cohorts ($P = 0.036$) (Table V).

Discussion

SNPs rs4444235 and rs1957636 on chromosome 14q22.2, and rs961253 and rs4813802 on 20p12.3 have been related to CRC susceptibility in previous association studies (13). Although their relationship with CRC risk has been undoubtedly established, all of the cohorts described so far happened to have a northern European origin. Thus, we thought it would be interesting to study the behavior of these SNPs and their related association signals in an independent case-control series of southern European origin (EPICOLON). The discrepancies between EPICOLON and the northern European populations further justified the latter inclusion of three new cohorts (SPAIN2, ITALY and PORTUGAL) in the analyses.

We found that allele frequencies at the four investigated SNPs were significantly different between northern and southern European populations. This divergence was observable both when comparing MAFs between the northern described populations and EPICOLON, and to a lesser extent for the 1000 Genomes European data available for these markers. LD block analysis performed separately for the CEU and TSI HapMap3 populations at both the broad and fine levels was less conclusive. The fine scale analysis shows some evidence of variation in the tagging abilities of the two SNPs close to *BMP4* (rs4444235 and rs1957636). This is an important point, because in order for these association signals to be replicated in other cohorts, the allelic architecture of the populations, namely LD structure, needs to be shared (24). Given these differences, it is easy to understand why only one of these signals (rs961253) could be positively replicated in the association analysis in either the four populations separately or their meta-analysis (despite the study being powered enough to detect all four).

It is important to note that each of the markers represents an independent association signal with its own distinctive features and different evidence of discrepancy between the northern and southern European cohorts. For instance, rs4444235 shows significant MAF discrepancies amongst the northern and southern European cohorts at each stage of this investigation. Furthermore, markers tagged by rs4444235 are different in the CEU and TSI HapMap3 populations (Table III). This is also true of rs1957636. For rs4813802, the evidence of differential behavior between northern and southern populations in further supported by the fact that ORs seem to be significantly different for both groups. It is interesting that the only replicated association signal (that of rs961253) shows much weaker evidence of differences between the northern and southern European data sets. MAFs are not significantly different in northern and southern European cases when all southern European cohorts are considered together. Also, the LD analysis shows that rs961253 tags the same markers in CEU and TSI populations (Table III), hence explaining our ability to replicate this association in the southern European cohorts.

The differences observed in allele frequencies as well as the lack of replication for the association signals in the southern European cohorts may arise from a variety of scenarios, each of which may apply to either one of the three unreplicated markers: (i) tagging of the real causative variant may be different in northern and southern European populations, with this 'imperfect' tagging resulting in the underestimation of the effect of the real functional variant; (ii) although the Common Disease–Common Variant hypothesis is thought to explain a wide proportion of disease heritability, we cannot rule out the possibility that other types of genetic variation, namely rarer variants, are responsible for these association signals (25); (iii) although the least probable, the assumption that the susceptibility variants are the same in all European populations is not necessarily true, and therefore, it could be the case that the variants tagged in the northern cohorts do not constitute CRC risk factors in the southern populations. The case might be a little different for rs4444235 however. It has been discovered that this variant may have some direct implications in the *cis*-regulation of *BMP4*, and thus situations (i) and (ii) are not applicable for this marker (26). Therefore, we can only assume that in this case either the described OR has been underestimated and much larger cohorts are needed to detect this subtle effect, or that it does indeed not constitute a CRC risk variant in the studied southern European populations.

Although Europe is thought to be overall homogeneous, it has already been widely documented that there are two north–south west–east gradients that determine the genetic diversity of European populations (27,28). These differences between the northern and southern European populations are compatible with previous theories that have postulated a southeast–northwest expansion in the continent or larger effective population sizes in southern than northern Europe (with the possible exception of population isolates, such as the Finns) (29). This is also supported by the fact that genetic diversity tends to be larger and LD smaller in southern Europe (30). Therefore, allele frequencies may vary greatly between populations for some particular loci, due to founder effects and genetic drift, and this may make some SNPs informative in one population but not in another. This fact has some important implications in association studies, where mismatches in ancestry between cohorts may lead to false-positive findings or a decreased power to detect associations (31).

Our results suggest that although other CRC susceptibility loci (8q23.3, 10p14, 11q23 and 15q13.3) have been described to behave similarly in all the cohorts studied (13,32,33), European populations may not be as homogeneous as initially thought for the three unreplicated loci with regards to CRC susceptibility. Whatever the underlying reason for these differences may be, we must highlight that these could have some serious implications in further fine-mapping, signal-refining or functional works and should be taken into account in subsequent study designs when considering these loci. The eventual finding of the real causative variants may be the key to fully understand the reason differences such as those described in this study. Clarification of CRC association signals and their

behavior in both northern and southern European populations could be extremely helpful in the detection of individuals at higher risk of developing CRC.

Supplementary material

Supplementary Figures 1–4 can be found at http://carcin.oxfordjournals.org/

Funding

Fondo de Investigación Sanitario/FEDER (08/1276, 08/0024, PS09/02368, 11/00219 and 11/00681); Instituto de Salud Carlos III (Accion Transversal de Cancer); Xunta de Galicia (PGIDIT07PXIB9101209PR); Ministerio de Economía y Competitividad (SAF07-64873 and SAF2010-19273); Fundación Privada Olga Torres (C.R.P.). C.F.R. and S.C-B. are supported by grants from Fondo de Investigación Sanitaria (PS09/02368 to C.F.R. and CP03/0070 to S.C-B.); A.Car., M.T., S.C-B., L.C.C. and I.T. are supported by the FP7 CHIBCHA Consortium.

Acknowledgements

We are sincerely grateful to all patients who participated in this study, recruited as part of the EPICOLON Project. We acknowledge the Spanish National DNA bank (BNADN) for the availability of the samples. We also thank all the individuals who participated to the Italian study and the personnel of Tissue Bank of Fondazione IRCCS Istituto dei Tumori for sample collection and all pathologists for their contribution and collaboration.

Conflict of Interest Statement: None declared.

References

1. Ferlay,J. *et al.* (2004) GLOBOCAN 2002: cancer incidence, mortality and prevalence worldwide. *IARC Cancer Base No. 5, version 2.0*, IARC, Lyon.
2. Lichtenstein,P. *et al.* (2000) Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.*, **343**, 78–85.
3. de la Chapelle,A. (2004) Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer,* **4**, 769–780.
4. Aaltonen,L.A. (2000) Hereditary intestinal cancer. *Semin. Cancer Biol.*, **10**, 289–298.
5. Castells,A. *et al.* (2009) Concepts in familial colorectal cancer: where do we stand and what is the future? *Gastroenterology*, **137**, 404–409.
6. Houlston,R.S. *et al.* (2004) The search for low-penetrance cancer susceptibility alleles. *Oncogene*, **23**, 6471–6476.
7. Houlston,R.S. *et al.* (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.*, **40**, 1426–1435.
8. Houlston,R.S. *et al.* (2010) 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.*, **42**, 973–977.
9. Dunlop,M.G. *et al.* (2012) Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat. Genet.*, **44**, 770–776.
10. Deng,H. *et al.* (2007) Bone morphogenetic protein-4 is overexpressed in colonic adenocarcinomas and promotes migration and invasion of HCT116 cells. *Exp. Cell Res.*, **313**, 1033–1044.
11. Deng,H. *et al.* (2009) Overexpression of bone morphogenetic protein 4 enhances the invasiveness of Smad4-deficient human colorectal cancer cells. *Cancer Lett.*, **281**, 220–231.
12. He,X.C. *et al.* (2004) BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat. Genet.*, **36**, 1117–1121.
13. Tomlinson,I.P. *et al.* (2011) Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.*, **7**, e1002105.
14. Piñol,V. *et al.* (2005) Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA*, **293**, 1986–1994.
15. Fernández-Rozadilla,C. *et al.* (2010) Single nucleotide polymorphisms in the Wnt and BMP pathways and colorectal cancer risk in a Spanish cohort. *PLoS ONE*, **5**(9), e12673.
16. Purcell,S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
17. Ihaka,R. *et al.* (1996) R: a language for data analysis and graphics. *J. Comput. Graph Stat.*, **5**, 299–314.
18. 1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
19. International HapMap Consortium. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
20. Barrett,J.C. (2009) Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harb. Protoc.*, **10**, pdb.ip71.
21. Petitti,D.B. (2000) *Meta-analysis, Decision Analysis, and Cost-effectiveness Analysis: Methods for Quantitative Synthesis in Medicine.* Oxford University Press, New York, NY.
22. Higgins,J.P. *et al.* (2002) Quantifying heterogeneity in a meta-analysis. *Stat. Med.*, **21**, 1539–1558.
23. Skol,A.D. *et al.* (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, **38**, 209–213.
24. Houlston,R.S. (2012) COGENT (COlorectal cancer GENeTics) revisited. *Mutagenesis*, **27**, 143–151.
25. Weiss,K.M. *et al.* (2000) How many diseases does it take to map a gene with SNPs? *Nat. Genet.*, **26**, 151–157.
26. Lubbe,S.J. *et al.* (2012) The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. *Oncogene*, **31**, 3777–3784.
27. Heath,S.C. *et al.* (2008) Investigation of the fine structure of European populations with applications to disease association studies. *Eur. J. Hum. Genet.*, **16**, 1413–1429.
28. Seldin,M.F. *et al.* (2006) European population substructure: clustering of northern and southern populations. *PLoS Genet.*, **2**, e143.
29. Cavalli-Sforza,L.L. *et al.* (1993) Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur. J. Hum. Genet.*, **1**, 3–18.
30. Lao,O. *et al.* (2008) Correlation between genetic and geographic structure in Europe. *Curr. Biol.*, **18**, 1241–1248.
31. Novembre,J. *et al.* (2008) Genes mirror geography within Europe. *Nature*, **456**, 98–101.
32. Tomlinson,I.P. *et al.* (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.*, **40**, 623–630.
33. Pittman,A.M. *et al.* (2008) Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. *Hum. Mol. Genet.*, **17**, 3720–3727.

Received July 3, 2012; revised November 2, 2012; accepted November 6, 2012