RESEARCH ARTICLE

Single nucleotide polymorphisms associated with susceptibility for development of colorectal cancer: Case-control study in a Basque population

Iker Alegria-Lertxundi[1,2], Carmelo Aguirre[2,3], Luis Bujanda[4], Francisco Javier Fernández[5], Francisco Polo[6], José M. Ordovás[7,8,9], M. Carmen Etxezarraga[10,11], Iñaki Zabalza[12], Mikel Larzabal[13], Isabel Portillo[14], Marian M. de Pancorbo[2], Leire Palencia-Madrid[2], Ana M. Rocandio[1,2], Marta Arroyo-Izaga[1,2]*

1 Department of Pharmacy and Food Sciences, Faculty of Pharmacy, University of the Basque Country, UPV/EHU, Vitoria-Gasteiz, Spain, 2 BIOMICs Research Group, University of the Basque Country, UPV/EHU, Vitoria-Gasteiz, Spain, 3 Pharmacovigilance Unit, Galdakao-Usansolo Hospital, Osakidetza, Spain, 4 Department of Gastroenterology, Donostia University Hospital / BioDonostia Institute, Biomedicine Research Networking Center - CIBER of Hepatic and Digestive Diseases (CIBERehd), University of the Basque Country, UPV/EHU, San Sebastian, Spain, 5 Department of Gastroenterology, Galdakao-Usansolo Hospital, Osakidetza, Galdakao, Spain, 6 Department of Gastroenterology, Basurto University Hospital, Osakidetza, Bilbao, Spain, 7 Nutrition and Genomics Laboratory, Jean Mayer Human Nutrition Research Center on Aging, Tufts University, Boston, Massachusetts, United States of America, 8 IMDEA Food, Madrid, Spain, 9 Nutritional Genomics and Epigenomics Group, Madrid Institute for Advanced Studies (IMDEA) Food Institute, Madrid, Spain, 10 Department of Pathology, Basurto Hospital, Osakidetza, Bilbao, Spain, 11 Department of Physician and Surgeon Specialities, University of the Basque Country, Leioa, UPV/EHU, Spain, 12 Department of Pathology, Galdakao-Usansolo Hospital, Osakidetza, Galdakao, Spain, 13 Department of Pathology, Donostia University Hospital / BioDonostia Institute, Biomedicine Research Networking Center - CIBER of Hepatic and Digestive Diseases (CIBERehd), San Sebastian, Spain, 14 Colorectal Cancer Screening Programme, The Basque Health Service, Bilbao, Spain

* marta.arroyo@ehu.eus

Abstract

Given the significant population diversity in genetic variation, we aimed to investigate whether single nucleotide polymorphisms (SNPs) previously identified in studies of colorectal cancer (CRC) susceptibility were also relevant to the population of the Basque Country (North of Spain). We genotyped 230 CRC cases and 230 healthy controls for 48 previously reported CRC-susceptibility SNPs. Only the rs6687758 in *DUPS10* exhibited a statistically significant association with CRC risk based on the crude analysis. The rs6687758 *AG* genotype conferred about 2.13-fold increased risk for CRC compared to the *AA* genotype. Moreover, we found significant associations in cases between smoking status, physical activity, and the rs6687758 SNP. The results of a Genetic Risk Score (GRS) showed that the risk alleles were more frequent in cases than controls and the score was associated with CRC in crude analysis. In conclusion, we have confirmed a CRC susceptibility locus and the existence of associations between modifiable factors and the rs6687758 SNP; moreover, the GRS was associated with CRC. However, further experimental validations are needed to establish the role of this SNP, the function of the gene identified, as well as the contribution of the interaction between environmental factors and this locusto the risk of CRC.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Colorectal cancer (CRC)is the fourth most common type of tumour, being 6.1% of the total new cases of cancer diagnosed in 2018and one of the major causes of cancer-related morbidity and mortality globally(9.2% of cancer deaths)[1]. There is wide geographical variation in incidence with rates varying 8-fold (colon cancer) and 6-fold (rectal cancer) in both sexes worldwide[1]. In this sense, Spain is one of the countries with the highest incidence of CRC, and taking into account both sexes, it was the most frequent cancer diagnosed in 2018 with 13.7% of newcancer cases[2] and is the main cause of cancer related deaths [3]. Considering the magnitude of the problem, the use of screening tests for early detection and effective treatment of CRC during the initial stages would have a significant impact on public health. In this sense, US PreventiveServicesTaskForce and the American CancerSocietyrecommendthescreening for CRC byannualfaecaloccultbloodtesting (FOBT), flexible sigmoidoscopyor(every 5 years) orcolonoscopy(every 10 years), in subjects aged 50 years or older [4].

The mechanisms underlying CRC occurrence and progression are complicated and mainly involve genetic and environmental factors, such as sex[5,6], diet and physical activity [5,7]. Various oncogenes and tumour suppressors, such as *KRAS*, *APC*, *BRAF*, *TP53*, and *SMAD4*, have been identified by CRC-related studies and may be useful for diagnosing and treating CRC in the future[5,8,9].

There is a direct association between sporadic tumour occurrence and susceptibility variants carried by an individual[10]. Many candidate gene[11] and genome-wide association studies (GWAS) [12]have evaluated common genetic risk factors for CRC; however, only a few of these have been replicated in subsequent studies[10]. Thus, in this study, we aimed to test the hypothesis that some of the previously reported CRC-related SNPs are associated with CRC susceptibility in the Basque population, in which there are no previous studies of this kind. Therefore, we investigated possible associations between 48 susceptibility SNPs and development of sporadic CRC in the adult population of the Basque Country.

Methods

Design

This is an observational, matched case-control study in a population group residing in the Basque Country (Spain).

Study population

Participants in this study were recruited among patients attending, between January 2012 and December 2014, any of the three hospitals of the Osakidetza/Basque Health Service (Basurto, Galdakao and Donostia)belong of the Basque Country Colorectal Cancer Screening Programme (CRCSP)[13]. To be eligible for this CRCSP, average risk people from 50 to 69 years, asymptomatic for colorectal symptoms and registered with the Osakidetza/Basque Health Service [13]. Subjects with symptoms suggesting CRC or with high CRC risk, such as individuals with familial adenomatous polyposis or hereditary nonpolyposis are managed outside this programme and are not included in this analysis. Subjects were invited to participate in this study by the gastroenterologists who performed the colonoscopies as a confirmatory test.

The recruitment and data collection for the present study were conducted between 2014 and 2016. All the patients who were newly diagnosed with CRC (n = 601) were invited to participate in this study, that is, the individuals with a positive result, (abnormal) to an immunochemical faecal occult blood test (iFOBT), being the faecal-Haemoglobin cut-off point of 20 μg Hb/g faeces for both sexes[13] and a colonoscopy[13]. Of those, 283 refused to participate in
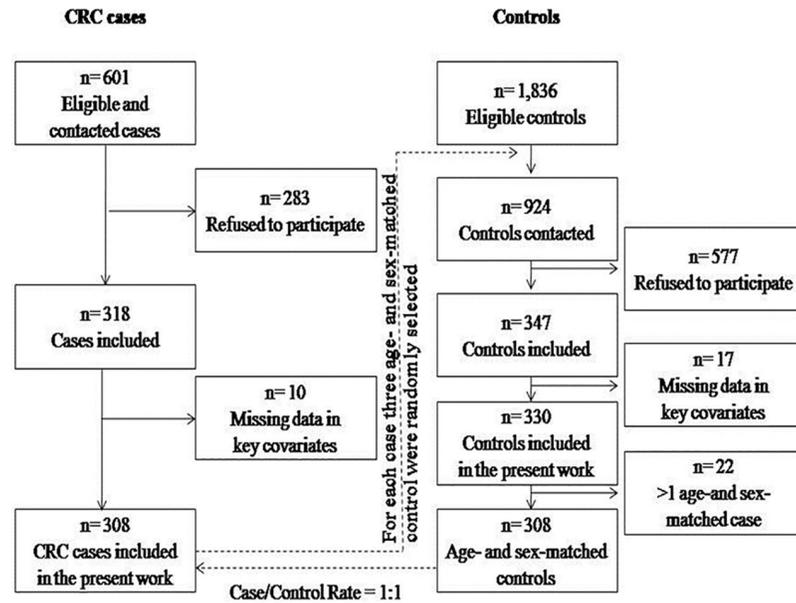
Fig 1. Flow chart of the process of obtaining the sample. CCR, Colorectal cancer.

the study, and 10 were excluded due to missing information. Ultimately, 308 subjects (66.2% men) consented to participate in the survey and completed all the questionnaires.

In addition, for each case, three age- (±9.0 years) and sex-matched control patients were randomly sought from the list of CRC-free subjects (n = 1,836) who participated in the CRCSP during the same period as the cases. The matched controls were patients with positive results (abnormal) for iFOBT and negative colonoscopy results (normal). The participation rate of the controls was 37.6%, and 17 subjects were excluded due to missing information. Finally, the matched case-to-control ratio was 1:1, and the final dataset included 308 cases who were diagnosed with CRC and 308 age- and sex-matched controls. The flowchart displaying the selection process for the CRC cases and controls is shown in Fig 1. Thirty-three cases, 39 controls and 6 cases-controls initially included in this study were excluded from the genetic analysis because incomplete genotyping by insufficient DNA available for the assay, and the respective partners of cases and controls were also excluded of the study. Finally, genotyping data were obtained from 230 cases and 230 controls.

The time spent between the participation in the CRCSP and in the present study was 1.8 (1.0) years (range: 0.4–4.6) in cases and 1.6(1.5) years (range: 0.2–3.7) in controls, without significant differences (P = 0.119). Consenting participants self-completed and returned a detailed Food Frequency Questionnaire (FFQ) and one general questionnaire (GQ). The questions referred to the behaviours before participating in the CRCSP. Assistance from the study staff was available to help the patients to understand the items on the questionnaires.

This study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving patients were approved by the Clinical Research Ethics Committee of the Basque Country (reference numbers PI2011006 and PI2014042). Written informed consent was obtained from all the study participants.

Biological samples and genotyping

In this study, healthy tissues or saliva samples of 230 CRC patients and 230 controls were collected and genotyped. Samples were provided by the Basque Biobank for Research-OEHUN

www.biobancovasco.org and were processed following standard operating procedures with appropriate ethical approval. DNA was extracted using AllPrep DNA / RNA kit (Qiagen) for paraffin-embedded tissue samples and AutoGenFlex Tissue DNA Extraction kit (Autogen) for mouthwash saliva samples and then was quantified with NanoDrop™ Spectrophotometer (ThermoFisher).

Double-stranded DNA was quantified by fluorometry using theQuant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, CA) on a DTX 880 Multimode Detector (Beckman Coulter) to normalize DNA concentration. After an updated summary of the published SNPs associated with susceptibility for development of CRC [14,15], those shown in Table 1 were selected. These SNPs were organized in the context of the gene(s) at or near locus and chromosome locus. The allelic discrimination was assessed using the MassARRAY® System (Agena Bioscience) on CeGen-PRB2-ISCII (Nodo USC) following the procedure provided by the manufacturer. Quality control samples were included in the genotyping assays.

Associated data

The questionnaire mentioned above, the GQ was used to gather information on weight status (self-reported weight and height) and environmental factors (demographic factors: age and sex; and lifestyle information: physical activity (PA) and smoking consumption). These questions were taken from the Spanish Health Questionnaire [16]. Body mass index (BMI), estimated from self-reported height and weight was classified according to the WHO criteria for those under 65 years of age [17] and according to the criteria proposed by Silva Rodríguez *et al.* for those 65 years and older [18].

Diet was assessed using a short FFQ that is a modified version of the Rodríguez *et al.* questionnaire[19]. This adaptation was validated with multiple 24- recalls in a subsample of the participants[20]. It consists of 67 items and requires the subjects to recall the number of times each food item was consumed either per week or per month. The respondents might also record the consumption of other foods that were not included on the food list.

Average portion sizes were employed to convert FFQ consumptions[21]. For items that included several foods, each food's contribution was estimated with weighting coefficients that were obtained from the usual consumption data[22]. All the food items that were consumed were entered into DIAL 2.12 (2011ALCE INGENIERIA), a type of dietary assessment software, to estimate energy intake (kcal/d). Moreover, the FFQ included specific questions about their frequency of intake of five major types of alcohol beverages: beer, wine, cider, aperitif with alcohol and liquor. In terms of the amount consumed, 10 g of alcohol was considered a standard drink[23]. Participants were categorized into non-drinker/moderate consumption and risk consumption, according to the SENC criteria that consider moderate drinking is up to 1 standard drink per day for women and up to 2 standard drinks per day for men[23]. Alcohol consumption was also expressed in tertiles of ml per day according to sex (men: T1, $\leq$ 70.6; T2, 70.7–138.8; T3, $\geq$ 138.9; and women: T1 $\leq$ 5.8; T2, 5.9–69.8; T3, $\geq$ 69.9).

Additionally, socioeconomic data was assessed with an index that was obtained from the clinical databases developed by the Health Department of the Basque Government, namely the socioeconomic deprivation index (DI). This index was estimated using the MEDEA project criteria[24] from simple indicators in the 2001 Census, namely unemployment, manual workers, casual workers, low education level and low education level among young people. The DI was divided into quintiles (Q), with the first being the least disadvantaged and the fifth being the most disadvantaged. The DI was successfully assigned to 82.4% of participants, while the address information quality did not permit the linking of the remaining 17.6%.

Table 1. 48 SNPs associated with susceptibility for the development of CRC and analyzed in this study.

SNP	Gene(s) at or near locus, variant type	Chr. locus	OR[a]	Risk allele[b]	SNP	Gene(s) at or near locus, variant type	Chr. locus	OR[a]	Risk allele[b]
rs12080929	*TRABD2B*, intron variant	1p33	0.87	C [61]	rs1535	*FADS2*, intron variant, FADS1, upstream gene variant	11q12.2	1.15	A [69]
rs6687758	*DUSP10*, regulatory region variant	1q41	1.04	G [62]	rs3802842	*COLCA1*, upstream gene variant, *COLCA2*, intron variant	11q23.1	1.14	C [69]
rs6691170	*LOC105372950, DUSP10*, intergenic variant	1q41	1.01	T [62]	rs10849432	*LOC105369625*, intron variant, non coding transcript variant	12p13.31	1.14	T [70]
rs10911251	*LAMC1*, intron variant	1q25.3	1.11	A [62]	rs3217810[e]	*CCND2*, intron variant, *CCND2-AS1*, upstream gene variant	12p13.32	1.19	T [62]
rs11903757	*NABP1/SDPR*, Intergenic variant	2q.32.3	1.14	C [62]	rs3217901	*CCND2*, intron variant	12p13.32	1.10	G[71]
rs10936599	*MYNN*, upstream gene variant	3q26.2	1.02	C [62]	rs10774214	*CCND2*, intron variant, non coding transcript variant	12p13.32	1.17	T [72]
rs647161	*C5orf66*, Intron variant, non coding transcript variant	5q31.1	1.07	A [62]	rs7136702	*LARP4/DIP2B, ATF1*, intergenic variant	12q13.12	1.10	T [62]
rs2736100	*TERT*, 3 prime UTR variant	5p15.33	1.07	A [63]	rs11169552	*LARP4/DIP2B, ATF1*, upstream gene variant	12q13.12	1.05	C [62]
rs1321311	*SRSF3/CDKN1A*, regulatory region variant	6p21.2	1.07	A [62]	rs59336	*TBX3*, intron variant	12q24.21	1.15	T [62]
rs11987193	*DUSP4*, intergenic variant	8p12	0.79	T [61]	rs4444235	*BMP4/ATP5C1P1/CDKN3/ MIR5580*, downstream gene variant	14q22.2	1.11	C [73]
rs16892766	*TRPS1/EIF3H/UTP23*, downstream gene variant	8q23.3	1.25	C [63]	rs1957636	*LOC105370507*, regulatory region variant	14q22.2	1.03	T [74]
rs6983267	*CCAT2*, intron variant, non coding transcript variant, *CCAT2*, non coding transcript exon variant	8q24.21	1.15	G [63]	rs4779584	*SCG5, GREM1, FMN1*, intergenic variant	15q13.3	1.18	T [70]
rs10505477	*CASC8*, intron variant, non coding transcript variant	8q24.21	1.11	A [64]	rs16969681	*GREM 1*, downstream gene variant	15q13.3	1.18	T [75]
rs7014346	*CASC8*, intron variant, non coding transcript variant, *POU5F1B*, intron variant	8q24.21	1.20	A [65]	rs11632715	*SCG5, GREM1, FMN1*, intergenic variant	15q13.3	1.12	A [76]
rs719725	*TPD52L3/UHRF2/GLDC*, intergenic variant	9p24.1	1.08	A [61]	rs9929218	*CDH1*, intron variant	16q22.1	1.10	A [75]
rs10795668	*LOC105376400*, upstream gene variant	10p14	1.32	A [66]	rs12603526	*NXN*, intron variant	17p13.3	1.10	C [69]
rs704017	*ZMIZ1-AS1*, intron variant, non coding transcript variant	10q22.3	1.13	G [67]	rs4939827	*SMAD7*, intron variant	18q21.1	1.16	T [77]
rs1035209	*ABCC2/MRP2*, intergenic variant	10q24.2	1.13	T [68]	rs10411210	*RHPN2*, intron variant	19q13.11	1.15	C [73]
rs12241008	*VTI1A*, intron variant	10q25.2	1.19	C [38]	rs1800469	*TGFB1*, upstream gene variant *B9D2*, downstream gene variant, *TMEM91*, intron variant	19q13.2	1.09	G[69]
rs11196172	*TCF7L2*, intron varian	10q25.2	1.14	A [69]	rs2241714	*TGFB1, TMEM91*, upstream gene variant, *B9D2*, missense variant	19q13.2	1.09	C [70]
rs1665650	*HSPA12A*, intron variant	10.q25.3	0.95	T [64]	rs961253	*BMP2/HAO1/FERMT1*, upstream gene variant	20p12.3	1.12	A [73]
rs174537	*TNEM258*, downstream gene variant, *MYRF*, intron variant	11q12.2	1.16	G [64]	rs4813802	*BMP2/HAO1/FERMT1*, regulatory region variant	20p12.3	1.10	C [70]
rs4246215	*TNEM258*, upstream gene variant *FEN1*, 3 prime UTR variant, FADS1, downstream gene variant, MIR611, upstream gene variant, FADS2, intron variant	11q12.2	1.15	G [69]	rs2423279	*HAO1/PLCB1*, downstream gene variant	20p12.3	1.10	C [72]
rs174550	*FADS1*, intron variant	11q12.2	1.15	T [69]	rs5934683	*SHROOM*, upstream gene variant, GPR143, intron variant	Xp22.2	1.04	C [31]

Chr, Chromosome; OR, odds ratio; SNP, single nucleotide polymorphism

[a]Odds ratios of previous studies are reported to calculate weighted Genetic Score

[b] Superscript numbers correspond with the studies in References

Quality management

In the present research, we apply a similar quality management that those used in the IDEFICS study [25]. A unique subject identification number was attached to each recording sheet, questionnaire, and sample, as in other researches. The identification number had to be entered twice before the document could be entered into its respective database. All data were entered twice independently, and deviating entries were corrected. Inconsistencies that were identified by additional plausibility checks were rectified.

Statistical analysis

Statistical analyses were performed using SPSS 22.0 (SPSS Inc, Chicago, USA), STATA 13.0 (StataCorp LP, Texas, USA). Categorical variables are shown as a percentage, and continuous variables are shown as the means and standard deviations (s.d.). Normality was checked using Kolmogorov-Smirnov-Lilliefors test. Paired *t*-test or Wilcoxon rank-sum test was used to two related means comparison, and a $\chi 2$ test was used to evaluate differences. Tests for association and deviation from Hardy-Weinberg equilibrium were performed separately in CRC patients and healthy controls. When expected frequencies were lesser than 5, Fisher's exact test was used.

In the case-control study, we estimated the odds ratio (OR) and 95% confidence interval (95% CI) for the polymorphism selected using conditional logistic regression adjusted for age (50–59 years old *vs*. 60–69 years old), sex (women *vs*. men), BMI (underweight/normal weight *vs*. overweight/obesity), physical activity ($\geq$15 min/d *vs*. <15 min/d), smoking status (never smoker *vs*. current and former smoker and quit smoking: $\geq$ 11 years ago *vs*. < 11 years ago), alcohol consumption (T1, T2 and T3) and Deprivation Index (DI) (quintile 1–3 *vs*. quintile 4–5) as categorical variables and energy intake as quantitative (kcal/d). ORs were calculated for the codominant model, dominant model, recessive model, and allelic comparison. The most frequent genotype (homozygous) was considered the reference group to calculate ORs in a codominant and dominant model, and the most frequent genotype (homozygous) and the heterozygous genotype containing the risk allele were considered the reference group in the recessive model. The significance level was corrected using a Bonferroni correction by dividing the standard *P* value (two-tailed) (0.05) by the total number of SNPs analyzed (n = 48), assuming alpha was equal to 0.001 (α = 0.05/48).

Additionally, correspondence analysis (CA) was performed using PAST 3.21 to identify potential associations between SNPs associated with CRC and associated data. CA is a multivariate statistical technique which provides Cartesian diagrams based on the association of the variables examined. All variables were represented in graphs and the more closed are the points the more higher is the level of association between variables[26].

To assess genetic susceptibility, two methods were used as a simple, unweighted count method (count Genetic Risk Scores, c-GRS) and a weighted method (w-GRS)[27,28]. Both methods assumed each SNP to be independently associated with risk[29]. An additive genetic model was assumed: weightings of 0, 1, and 2 were given according to the number of risk alleles present[29,30].

The count method assumed that each SNP contributed equally to CRC risk and was calculated by summing the number of risk alleles across the panel of SNPs tested. This produced a score between 0 and twice the number of SNPs, i.e., representing the total number of risk alleles. The weighted GRS was calculated by multiplying each β-coefficient for the CRC phenotype from the discovery set by the number of corresponding risk alleles (0, 1, or 2 copies of the risk allele except for the SNP rs5934683 in chromosome X that was coded 0, 0.5, and 1) and then summing the products[31].

Finally, we defined the GRS as the count of risk alleles across all 48 SNPs, ranging from 0 to 95 for c-GRS and 0 to 105 for w-GRS. Since the published effects of each SNP were similar, an unweighted GRS was preferred. However, we also explored the models using weights derived from the GWAS publications and models fitted to our data[32].

Gene expression association analyses

Gene expression changes in tumour and normal colon tissue associated to SNPs with significant association with CRC risk were analyzed using publicly available data and bioinformatic tools. In the first place Genomic Data Commons Data Portal (GDC) (https://portal.gdc. cancer.gov) was used to examine data generated by the TCGA (The Cancer Genome Atlas) research network (https://www.cancer.gov/tcga), but for SNPs with unavailable data in GDC portal alternative bioinformatic tools were applied. On the one hand, gene expression data from between case and control samples of colon and rectum adenocarcinomas were compared using GEPIA (Gene Expression Profiling Interactive Analysis) (http://gepia.cancer-pku.cn/ index.html) [33]. On the other hand, GTEx (The Genotype-Tissue expression project) (https:// gtexportal.org/home/) was used to check the relationship between SNPs and the expression level of genes related to these SNPs in colon tissue of healthy donors.

Results

Table 2 shows the comparisons of associated data between cases and controls. Cases had a higher consumption of cigarettes/day and were more engaged in regular physical activity at a medium-high level as compared with controls. In addition, in the total sample, there were more smokers in men than in women (70.6% vs. 54.5%; $P<0.001$); and had a higher consumption of cigarettes/day (11.6(11.1) vs. 9.0(11.4); $P = 0.030$). Among controls 51.9% of women and 65.4% of men were smokers ($P = 0.049$); and among cases, 57.1% of women and 75.8% of men were smokers ($P = 0.004$).

The distribution of genotypes and alleles at SNPs selected in the CRC group and in the control group that deviated from the Hardy-Weinberg equilibrium are shown in Supplementary Material(S1 Table). The SNPs that were not following the Hardy-Weinberg equilibrium in cases were rs12080929 and rs5934683. None of the genotype or allele frequencies for the SNPs analysed reached statistically significant differences between cases and controls, after Bonferroni correction application.

Table 3 presents some results of the association of susceptibility genotypes and alleles with the risk of CRC in the codominant model. Other SNPs analyzed in this study are shown in Supplementary Material (S2 Table). Adjusting for potential confounders did not appreciably alter the observed ORs. Only the rs6687758 exhibited a statistically significant association with CRC risk based on the crude analysis. The AG genotype of rs6687758 conferred about 2.13-fold increased risk for CRC compared to the AA genotype.

Moreover, there was an association between smoking status, physical activity and the rs6687758 SNP for CRC risk in cases (Fig 2). We did not find an association between the risk genotype for rs6687758 and other associated variables (BMI, sex, alcohol consumption, DI and age). The results of CA for all cases are shown in a Cartesian diagram. The first three axes accounted for more than 50.0% of the total variance in all cases (axis 1: 23.0%; axis 2: 19.6% and axis 3: 13.4%). An inverse association can be observed between the variable DI (which plotted at the negative end of axis 1) and age, positioned in the positive segment of axis 1. Overall, axis 1 represents a gradient that runs from low values for DI (0: Q1-Q3; 1: Q4-Q5) to high values for age (0:50–59 y; 1:60–69 y). From the genetic viewpoint, the SNP that showed

Table 2. Comparison of associated data between cases and controls with genotyping data.

	Cases (n = 230)	Controls (n = 230)	P^a-value
Age, years, mean(s.d.)	61.5(5.4)	60.9(5.5)	0.333
BMI classification, %			
NonOv/Ob	42.2	33.0	
Ov/Ob	57.8	67.0	**0.043**
Physical activity level, %			
Low	65.7	77.4	
Medium and high	34.3	22.6	**0.005**
Smoking status, %[b]			
Non-smoker	30.4	39.1	
Smoker	69.6	60.9	0.050
Cigarettes, cigarettes/day, mean(s.d.)	10.7(11.2)	8.3(10.9)	**0.007**
Number of cigarettes, %[b]			
< 15	49.3	66.9	
≥ 15	50.7	33.1	**0.003**
Alcoholic beverage intake, ml/day, mean(s.d.)	98.0(91.5)	97.2(107.5)	0.637
Tertiles of alcohol intake, ml/day[c], %			
T1	32.6	33.9	
T2	31.3	35.7	
T3	36.1	30.4	0.404
Standard drink units, classification, %			
Abstemious /low risk	72.1	79.1	
High risk	27.9	20.9	0.078
DI, %[b]			
Q1-Q3	73.5	69.6	
Q4-Q5	26.5	30.4	0.409

BMI, body mass index; DI, deprivation index; Ob, obesity; Ov, overweight, Q, quintile; s.d. standard deviation

[a]P<0.05 wassignificant

[b]Valid percentages

[c]Men: T1, ≤ 70.6; T2, 70.7–138.8; T3, ≥ 138.9; and women: T1 ≤ 5.8; T2, 5.9–69.8; T3, ≥ 69.9

https://doi.org/10.1371/journal.pone.0225779.t002

the closest association with associated variables wasrs6687758, which also plotted in the quadrant delimited by the positive segments of axis 1 and 2.

Analyses performed to study possible changes in gene expression associated with rs6687758 in tumour versus normal colon tissue showed that gene *DUSP10* is more expressed in colon sigmoid tissue when rs6687758 has GG genotype (in healthy individuals) (S1 Fig), but also, that it has higher expression in cases of colon and rectum adenocarcinomas than in healthy persons (S2 Fig).

For SNP rs719725, an increased CRC risk was found to be associated with the CC genotype in dominant and recessive models for crude analysis, compared with the AA and AC genotype (OR$_{CC}$: 1.77; 95% CI = 1.09–2.86; P = 0.020 in recessive model and OR$_{AC+CC}$: 1.64; 95% CI = 1.12–2.38; P = 0.010 in dominant model). Moreover, significantly elevated CRC risk was found to be associated with rs2736100, rs11987193and rs961253by using dominant model (for rs2736100 OR$_{AA+AC}$:1.72; 95%CI = 1.00–2.94; P = 0.048 in adjusted model; for rs11987193 OR$_{CC+CT}$: 1.45; 95%CI = 1.01–2.49; P = 0.046 in crude analysis; and for rs961253 OR$_{AA+AC}$: 1.47; 95%CI = 1.02–2.11; P = 0.038 in crude analysis).

Table 3. Association between genetic variants associated with susceptibility and the risk of CRC in the codominant model.

Gene, SNP ID[a]	N (cases/controls)	Model I[b]		Model II[c]	
		OR (95% CI)	P[d]-value	OR (95% CI)	P[d]-value
rs6687758					
AA	136/169	1.00	-	1.00	-
AG	87/51	2.13(1.39–3.25)	<0.001	1.95(1.05–3.60)	0.034
GG	7/9	1.02(0.37–2.82)	0.967	1.06(0.21–5.28)	0.945
A	359/389	1.00	-	1.00	-
G	101/69	1.60(1.13–2.28)	0.009	1.54(0.97–2.46)	0.067
rs6691170					
GG	72/87	1.00	-	1.00	-
GT	112/108	1.22(0.82–1.82)	0.331	1.20(0.64–2.26)	0.570
TT	45/31	1.79(1.01–3.16)	0.045	1.70(0.74–3.89)	0.207
G	256/282	1.00	-	1.00	-
T	202/170	1.23(0.94–1.62)	0.124	1.27(0.89–1.79)	0.185
rs719725					
AA	63/91	1.00	-	1.00	-
AC	116/106	1.46(0.97–2.18)	0.068	1.99(1.07–3.71)	0.030
CC	51/31	2.14(1.27–3.64)	0.005	1.80(0.78–4.17)	0.168
A	242/288	1.00	-	1.00	-
C	218/168	1.60(1.22–2.11)	<0.001	1.49(1.05–2.10)	0.025
rs12241008					
TT	196/204	1.00	-	1.00	-
CT	33/24	1.47(0.82–2.64)	0.192	1.49(0.75–2.95)	0.253
CC	1/2	0.50(0.05–5.51)	0.571	0.78(0.05–12.84)	0.862
T	425/435	1.00	-	1.00	-
C	35/28	1.22(0.72–2.09)	0.455	1.34(0.66–2.72)	0.412
rs7136702					
CC	80/91	1.00	-	1.00	-
CT	108/114	1.11(0.75–1.65)	0.593	1.03(0.56–1.89)	0.826
TT	42/25	1.98(1.09–3.64)	0.026	2.83(1.12–7.17)	0.028
C	268/296	1.00	-	1.00	-
T	192/164	1.34(1.03–1.74)	0.030	1.28(0.91–1.80)	0.154
rs2241714					
CC	116/101	1.00	-	1.00	-
CT	94/105	0.79(0.55–1.15)	0.217	0.54(0.31–0.95)	0.034
TT	20/23	0.72(0.37–1.38)	0.321	0.28(0.09–0.89)	0.031
C	326/307	1.00	-	1.00	-
T	134/151	0.80(0.61–1.06)	0.125	0.74(0.51–1.07)	0.114
rs961253					
CC	101/124	1.00	-	1.00	-
AC	103/76	1.65(1.11–2.46)	0.013	1.79(0.67–4.78)	0.247
AA	26/30	1.03(0.57–1.85)	0.925	1.04(0.41–2.63)	0.941
C	305/324	1.00	-	1.00	-
A	155/136	1.20(0.90–1.59)	0.208	1.11(0.76–1.62)	0.584

A, adenine; C, cytosine; CI, confidence interval; G, guanine; OR, odds ratio; rs, reference single nucleotide polymorphism; SNP, single nucleotide polymorphism; T, thymine

[a]The most frequent genotype (homozygous) was considered the reference group

[b]Model I, crude conditional logistic regression model

[c]Model II, conditional logistic regression adjusted forage, sex, BMI, physical activity, smoking status, alcohol consumption, Deprivation Index and energy intake.

Participants with missing data for the confounding variables were included as a separate category for these variables

[d]P<0.001 was significant
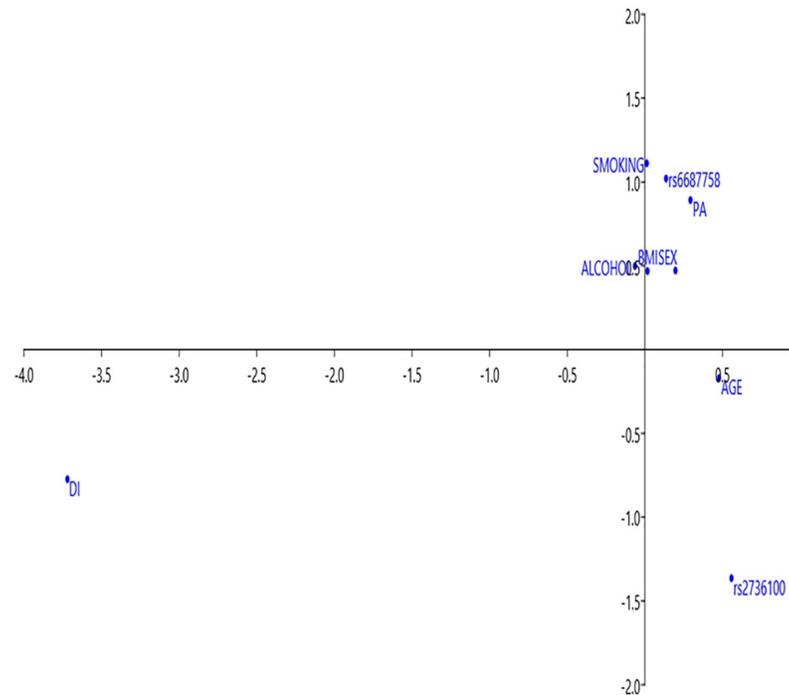
Fig 2. Cartesian diagram of correspondence analysis for studied associations between genetic and environmental factors in cases. BMI, body mass index; DI, deprivation index; PA, physical activity.

https://doi.org/10.1371/journal.pone.0225779.g002

Finally, the unweighted GRS of the sample studied was 38.6(4.6) (range: 25–52), with statistically significant differences between cases and controls (39.2(4.4) (range: 28–50.5) vs. 37.95 (4.6) (range: 25–52); $P = 0.002$). The GRS built as the unweighted count of risk alleles was significantly associated with CRC risk, with an average per-allele OR of 1.07 (95%CI = 1.02–1.11; $P = 0.002$) in crude analysis. However, this association was not statistically significant in the adjusted model (OR: 1.04; 95% CI = 1.00–1.10; $P = 0.066$). On the other hand, w-GRS was 44.7 (5.5) for the total sample, with statistically significant differences between cases and controls (45.3(5.4) (range: 32.2–58.6)vs.44.1(5.6) (range: 27.7–57.6); $P = 0.036$). The w-GRS was associated with CRC risk (OR: 1.04; 95% CI = 1.00–1.09; $P = 0.037$) in crude analysis but not in the adjusted one (OR: 1.01; 95% CI = 0.97–1.05; $P = 0.588$).

Discussion

In this study, we investigated SNPs associated with susceptibility for the development of CRC in a Basque population who took part in the population screening programme. We found that out of 48 analysed SNPs, only the rs6687758 was associated with the risk of CRC in this population. This is in agreement with previous GWAS that reported a positive association between this SNP and CRC also in European population[15,34]. Some authors have also observed relationships between this SNP and colorectal polyp risk[35]; although this SNPs is not associated significantly with adenoma risk and has their effects on the malignant stage of colorectal tumorigenesis[36]. The frequency of the risk allele of rs6687758 (G) in the European population (22.2%) [37]is similar to that registered in the cases of the present study and higher than that of the controls.

The other 47 risk SNPs did not replicate in our population. This may be due to differences in the underlying linkage patterns given the ethnic differences in populations studied. Twenty-

oneof the SNPs analyzedhave beenreplicated in Asian, American-Caucasian or African, but not in European (rs11903757, rs1321311, rs10505477, rs719725, rs704017, rs12241008, rs11196172, rs174537, rs4246215, rs174550, rs1535, rs10849432, rs3217901, rs4444235, rs11632715, rs4939827, rs10411210, rs1800469, rs2241714, rs961253 and rs4813802); and 4 were not replicated in population studies; however, they were associated with susceptibility for development of CRC in GWAS (rs1665650, rs59336, rs1957636 and rs12603526). The effect sizes of some of these associations were small (OR <1.20, P<0.05, for rs1321311, rs12241008, and rs704017)[38–40]. Additionally, it may be that the distribution of environmental factors in our population differs from that of the populations in which these genetic variants were discovered.

The SNP rs6687758 is in a regulatory region, flanking the promoter of *DUSP10*, at ~250 kb from the start of the gene. Hence, it is likely to affect the expression of this gene. Polymorphisms in *DUSP10 gene (dual* specificity protein phosphatase 10) have been previously demonstrated to be associated with CRC risk [41,42]. In this study, we confirmed this CRC susceptibility locus in the Basque population sample. Earlier analyses have found frequent dysregulation of dual specificity protein phosphatase 10 (*DUSP10/MKP-5*) in CRC [41]. *DUSP10* belongs to the dual kinase phosphatase family. These proteins are associated with cellular proliferation and differentiation, and they act as tumour suppressors [41,43].

Target kinases of DUSPs are inactivated by dephosphorylation of both phosphoserine/threonine and phosphotyrosine residues [41,42]. They act at several levels, taking part in fine-tuning signalling cascades. DUSPs negatively regulate members of the mitogen-activated protein kinase (*MAPK*) superfamily [41,44], which are implicated in some activities that are often dysregulated in cancer, such as cell proliferation, survival, and migration [41]. *MAPK* signalling also plays a key role in determining the response of tumour cells to cancer therapies, since its abnormal signalling has important consequences for the development and progression of human cancer [44].

Several studies have already shown the involvement of *DUSP*s as major modulators of critical signalling pathways dysregulated in different cancers [43], such as in the case of the overexpression of *DUSP1/MKP-1* in the early phases of cancer and its decreasing during tumour progression [42].

There is abundant evidence that *DUSP10*, in particular, may play an important role in tumorigenesis and could alter CRC risk [45,46]. It inactivates *p38* and *JNKin vitro*[41,47], and its upregulation are very common in CRC[48]. The activation of JNK protein is due to the protein kinase *G (PKG)/MEKK1/SEK1/JNK* cascade, and it is related with cell proliferation and inducing apoptosis[41,49]. Moreover, *p38* is involved in the promotion of cellular senescence as a meansof eluding oncogene-induced transformation; it participates in cell cycleregulation suppressing cell proliferation and tumorigenesis[41,49].

On the other hand, the results extracted from gene expression association analyses show a higher expression of *DUSP10* gene in CRC cases, but also that there is a higher expression of this gene in colon tissue of healthy controls when they have the GG genotype for rs6687758. Thus, it would be likely to find a relationship between higher expression of the gene and the presence of allele G in rs6687758 in tumour tissue. Nonetheless, it would be interesting to further explore this aspect through future analyses to compare gene expression between individuals carrying the risk variant and control individuals. Previous studies have pointed in the same direction that there is overall increase in patients' relapse-free survival when *DUSP10* expression is upregulated, and that *DUSP10* mRNA was increased in the tumour compared with normal tissue adjacent to the tumours [46,49,50].

We found an association between smoking status and the rs6687758 SNP for CRC risk in cases. Other authors have also observed this association[51]. Benzo[a]pyrene, one of the

carcinogenic compounds included in cigarette smoke, up-regulated *COX-2* in mouse cells[52], which in turn could either activate or be dependent on the *MAPK* pathway, suggesting a possible gene-smoking interaction [53,54]. Concerning the association between physical activity, the rs6687758 SNP and CRC risk, as far as we know, there are no precedents in the literature. However, other studies have found interactions between polymorphisms associated with growth hormone (*GH1*) and insulin-like growth factor I (*IGF-I*) (rs647161, rs2665802), physical activity and CRC [53,54]. According our results, rs6687758, medium-high physical activity level and CRC would be associated. However, this outcome, contrary o what it could be expected, could be related to changes in the lifestyles, including physical activity level, in cases after diagnosis [55].

We also analyzed unweighted and weighted GRS models. We observed that cases had more risk alleles than controls, this result was according to expectations considering the previous studies[56]. In the crude analysis, we observed that patients that had a higher number of risk alleles had a higher risk of CRC. Other authors observed similar results using an adjusted unweighted model [32]. However, some other authors did not find this association[57]. It should be noted that common allele variants generally have modest effect sizes[58], but the combination of multiple loci with modest effects into a global GRS might improve the identification of patients with genetic risk for common complex diseases, such cancer[59]. In this sense, Ortlepp *et al.*[60] concluded that more than 200 polymorphisms might be necessary for "reasonable" genetic discrimination.

Our study has several limitations and strengths. The principal limitations of this study were the small sample size that makes difficult to detect possible associations between polymorphisms and disease risk since some genotypes showed very low frequencies in our population. Another disadvantage of the small sample size is that they can produce false-positive results; in order to avoid it, the Bonferroni correction was used. The strengths of the study were that although controls tested positive in iFOBT, in CRCSP were confirmed that they were free of the disease through colonoscopy. Colonoscopy was used as diagnosis criteria to identify the cases in order to avoid false positives and negatives.

In **conclusion**, most SNPs analyzed were not associated with risk of CRC. Only one of the 48 SNPs analyzed, rs6687758, was associated with risk of CRC, in this population (on crude analysis). Moreover, there were significant associations between smoking status, physical activity, the rs6687758SNP and CRC risk. On the other hand, the results of the GRS showed that the risk alleles were more frequent in cases than controls and this score was associated with this type of cancer in crude analysis. Therefore, in this study, we have confirmed a CRC susceptibility locus and the existence of associations between modifiable factors such as smoking and physical activity and the presence of the risk genotype for rs6687758. However, further experimental validations are needed to establish the role of this SNP, the function of the gene identified, as well as the contribution of the interaction between environmental factors and this polymorphism to the risk of CRC.

Supporting information

S1 Table. Deviation from Hardy-Weinberg equilibrium and differences in allele frequencies and genotype distribution between cases and controls. A, adenine; C, cytosine; G, guanine; HWE, Hardy-Weinberg equilibrium; rs, reference single nucleotide polymorphism; SNP, single nucleotide polymorphism; T, thymine; [a]Valid percentages; [b]P<0.001 was significant; [c]Differences in allele frequencies and genotype distribution between cases and controls. (PDF)

S2 Table. Association between genetic variants of susceptibility and the risk of CRC in the codominant model. A, adenine; C, cytosine; CI, confidence interval; G, guanine; NA, no available data; OR, odds ratio; rs, reference single nucleotide polymorphism; SNP, single nucleotide polymorphism; T, thymine; [a]The most frequent genotype was considered the reference group; [b]Model I, crude conditional logistic regression model; [c]Model II, conditional logistic regression adjusted for: age, sex, BMI, physical activity, smoking status, alcohol consumption, Deprivation Index and energy intake. Participants with missing data for the confounding variables were included as a separate category for these variables; [d]$P<0.001$ was significant.
(PDF)

S1 Fig. eQTL violin plot showing gene association results for *DUSP10* gene, rs6687758 and colon sigmoid healthy tissue. A, adenine. G, guanine. Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2).
(TIFF)

S2 Fig. Box plot for comparing the difference of expression for DUSP10 gene between cases (in red) and controls (in grey) in colon adenocarcinoma and rectum adenocarcinoma. COAD, colon adenocarcinoma. N, normal. READ, rectum adenocarcinoma. T, tumour. The method for differential analysis is one-way ANOVA, using disease state (Tumor or Normal) as variable for calculating differential expression. Data Source: TCGA and GTEx data, using GEPIA.
(TIFF)

Acknowledgments

Author Contributions

Conceptualization: Iker Alegria-Lertxundi, Carmelo Aguirre, Luis Bujanda, Francisco Javier Fernández, Francisco Polo, M. Carmen Etxezarraga, Iñaki Zabalza, Mikel Larzabal, Isabel Portillo, Marian M. de Pancorbo, Ana M. Rocandio, Marta Arroyo-Izaga.

Data curation: Iker Alegria-Lertxundi, Carmelo Aguirre, Luis Bujanda, Francisco Javier Fernández, Francisco Polo, M. Carmen Etxezarraga, Iñaki Zabalza, Mikel Larzabal, Isabel Portillo, Marta Arroyo-Izaga.

Formal analysis: Iker Alegria-Lertxundi, Marian M. de Pancorbo, Leire Palencia-Madrid, Marta Arroyo-Izaga.

Funding acquisition: Carmelo Aguirre, Luis Bujanda, Francisco Javier Fernández, Francisco Polo, M. Carmen Etxezarraga, Iñaki Zabalza, Mikel Larzabal, Isabel Portillo, Marian M. de Pancorbo, Ana M. Rocandio, Marta Arroyo-Izaga.

Investigation: Iker Alegria-Lertxundi, Carmelo Aguirre, Francisco Javier Fernández, Francisco Polo, José M. Ordovás, M. Carmen Etxezarraga, Iñaki Zabalza, Mikel Larzabal, Isabel Portillo, Marian M. de Pancorbo, Leire Palencia-Madrid, Marta Arroyo-Izaga.

Methodology: Iker Alegria-Lertxundi, José M. Ordovás, Marian M. de Pancorbo, Leire Palencia-Madrid, Marta Arroyo-Izaga.

Project administration: Marta Arroyo-Izaga.

Software: Iker Alegria-Lertxundi, Marta Arroyo-Izaga.

Supervision: Luis Bujanda, José M. Ordovás, Marian M. de Pancorbo, Marta Arroyo-Izaga.

Writing – original draft: Iker Alegria-Lertxundi, Marta Arroyo-Izaga.

Writing – review & editing: Iker Alegria-Lertxundi, Carmelo Aguirre, Luis Bujanda, Francisco Javier Fernández, Francisco Polo, José M. Ordovás, M. Carmen Etxezarraga, Iñaki Zabalza, Mikel Larzabal, Isabel Portillo, Marian M. de Pancorbo, Leire Palencia-Madrid, Ana M. Rocandio, Marta Arroyo-Izaga.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality world wide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov; 68(6):394–424. https://doi.org/10.3322/caac.21492 PMID: 30207593

2. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer;2018. [cited 2019 Apr 4]. https://gco.iarc.fr/today

3. Castillejo A, Mata-Balaguer T, Montenegro P, Ochoa E, Lázaro R, Martínez-Cantó A, et al. The TGFBR1*6A allele is not associated with susceptibility to colorectal cancer in a Spanish population: a case-control study. BMC Cancer. 2009; 9:193. https://doi.org/10.1186/1471-2407-9-193 PMID: 19538729

4. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin. 2018; 68(4):250–281. https://doi.org/10.3322/caac.21457 PMID: 29846947

5. Chen Y, Du M, Chen W, Zhu L, Wu C, Zhang Z, Wang M, Chu H, Gu D, Chen J. Polymorphism rs2682818 in miR-618 is associated with colorectal cancer susceptibility in a Han Chinese population. Cancer Med. 2018; 7(4):1194–1200. https://doi.org/10.1002/cam4.1409 PMID: 29533012

6. Siegel RL, Miller KD, Jemal A. Cancerstatistics, 2019. CA Cancer J Clin. 2019; 69(1):7–34. https://doi.org/10.3322/caac.21551 PMID: 30620402

7. Marley AR, Nan H. Epidemiology of colorectal cancer. Int J Mol Epidemiol Genet. 2016; 7(3):105–114. PMID: 27766137

8. Sameer AS. Colorectalcancer: molecular mutations and polymorphisms. Front Oncol. 2013; 3:114. https://doi.org/10.3389/fonc.2013.00114 PMID: 23717813

9. Zhong R, Liu L, Zou L, Sheng W, Zhu B, Xiang H,et al. Genetic variations in the TGFβ signaling pathway, smoking and risk of colorectal cancer in a Chinese population. Carcinogenesis. 2013; 34(4):936–42. https://doi.org/10.1093/carcin/bgs395 PMID: 23275154

10. Yang CY, Lu RH, Lin CH, Jen CH, Tung CY, Yang SH, Lin JK, Jiang JK, Lin CH. Single nucleotide polymorphisms associated with colorectal cancer susceptibility and loss of heterozygosity in a Taiwanese population. PLoS One. 2014; 9(6):e100060. https://doi.org/10.1371/journal.pone.0100060 PMID: 24968322

11. Dong LM, Potter JD, White E, Ulrich CM, Cardon LR, Peters U. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. JAMA. 2008; 299(20):2423–36. https://doi.org/10.1001/jama.299.20.2423 PMID: 18505952

12. Castellví-Bel S, Abulí A, Castells A. Meta-analysis of several GWAS sets yields additional genetic susceptibility variants for colorectal cancer: first X-linked component identified. Gastroenterology. 2012; 143(6):1684–5. https://doi.org/10.1053/j.gastro.2012.10.008 PMID: 23073136

13. Idigoras I, Arrospide A, Portillo I, Arana-Arri E, Martínez-Indart L, Mar J, et al. Evaluation of the colorectal cancer screening Programme in the Basque Country (Spain) and its effectiveness based on the Miscan-colon model. BMC Public Health. 2017 Aug 1; 18(1):78. doi: 10.1186/s12889-017-4639-3. Erratum in: BMC Public Health. 2017;17 (1):736. PMID: 28764731

14. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. Gastroenterology. 2015; 149(5):1177–1190.e3. https://doi.org/10.1053/j.gastro.2015.06.047 PMID: 26216840

15. Real LM, Ruiz A, Gayán J, González-Pérez A, Sáez ME, Ramírez-Lorca R, et al. A colorectal cancer susceptibility new variant at 4q26 in the Spanish population identified by genome-wide association analysis. PLoS One. 2014; 9(6):e101178. https://doi.org/10.1371/journal.pone.0101178 PMID: 24978480

16. Ministerio de Sanidad, Servicios Sociales e Igualdad / Instituto Nacional de Estadística. Encuesta Nacional de Salud. España 2011/12.Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad;2012. http://www.msssi.gob.es/estadEstudios/estadisticas/sisInfSanSNS/nivelSalud.htm

17. WorldHealthOrganization (WHO). Reportof WHO consultation: Obesity: preventing and managingtheglobal epidemic.Geneva; 2000.

18. Silva Rodrigues RA, Martinez Espinosa M, Duarte Melo C, RodriguesPerracini M, Rezende Fett WC, et al. New values anthropometry for classification of nutritional status in theelderly. J NutrHealthAging. 2014; 18(7):655–61. https://doi.org/10.1007/s12603-014-0451-2 PMID: 25226103

19. Rodríguez IT, Ballart JF, Pastor GC, Jordà EB, Val VA. Validation of a short questionnaire on frequency of dietaryintake: reproducibility and validity. NutrHosp. 2008; 23(3):242–52. Spanish.

20. Alegria-Lertxundi I, Alvarez M, Rocandio AM, de Pancorbo MM, Arroyo-Izaga M. NutritionalAdequacy and DietQuality in ColorectalCancerPatientsPostsurgery: A PilotStudy. NutrCancer. 2016; 68(4):577–88. https://doi.org/10.1080/01635581.2016.1158299 PMID: 27144653

21. Carbajal A, Sánchez-Muniz FJ. Guía de prácticas. In: García-Arias MT, García-Fernández MC, editors. Nutrición y dietética. Secretariado de Publicaciones y Medios Audiovisuales, Universidad de León; 2003. p. 1a-130a.

22. Gobierno Vasco, Departamento de Agricultura Pesca y Alimentación. Estudio cuantitativo del consumo de alimentos en la CAPV, año 2007. Vitoria-Gasteiz: Servicio Central de Publicaciones del GobiernoVasco; 2008.

23. Aranceta J, Serra Majem L, Arija V, Gil A, Martínez de Vitoria E, Ortega R, et al. Objetivos nutricionales para la poblaciónEspañola Consenso de la Sociedad Española de Nutrición Comunitaria 2011. RevEspNutr Comunitaria. 2011; 17(4):178–199.

24. Domínguez-Berjón MF, Borrell C, Cano-Serral G, Esnaola S, Nolasco A, Pasarín MI, et al. Constructing a deprivation index based on census data in large Spanish cities(the MEDEA project). GacSanit. 2008; 22(3):179–87.

25. Ahrens W, Bammann K, Siani A, Buchecker K, De Henauw S, Iacoviello L, Hebestreit A, Krogh V, Lissner L, Mårild S, Molnár D, Moreno LA, Pitsiladis YP, Reisch L, Tornaritis M, Veidebaum T, Pigeot I; IDEFICS Consortium. The IDEFICS cohort: design, characteristics and participation in the baseline survey. Int J Obes (Lond). 2011; 35 Suppl1:S3–15. https://doi.org/10.1038/ijo.2011.30 PMID: 21483420

26. Celorrio D, Muñoz X, Amiano P, Dorronsoro M, Bujanda L, Sánchez MJ et al. Influence of Dopaminergic System Genetic Variation and LifestyleFactorsonExcessive Alcohol Consumption. Alcohol Alcohol. 2016; 51(3):258–67. https://doi.org/10.1093/alcalc/agv114 PMID: 26447226

27. Piccolo SR, Abo RP, Allen-Brady K, Camp NJ, Knight S, Anderson JL,et al. Evaluation of genetic risk scores for lipid level susinggenome-widemarkers in the Framingham Heart Study. BMC Proc. 2009; 3 Suppl7:S46.

28. Jung KJ, Won D, Jeon C, Kim S, Kim TI, Jee SH, et al. A colorectal cancer prediction model using traditional and genetic risk scores in Koreans. BMC Genet. 2015; 16:49. https://doi.org/10.1186/s12863-015-0207-y PMID: 25956580

29. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, et al. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. Ann Intern Med. 2009; 150 (8):541–50. https://doi.org/10.7326/0003-4819-150-8-200904210-00008 PMID: 19380854

30. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet. 2006; 7 (10):781–91. https://doi.org/10.1038/nrg1916 PMID: 16983374

31. Ibáñez-Sanz G, Díez-Villanueva A, Alonso MH, Rodríguez-Moranta F, Pérez-Gómez B, Bustamante M, et al. Risk Model for Colorectal Cancer in Spanish Population Using Environmental and Genetic Factors: Results from the MCC-Spain study. Sci Rep. 2017; 7:43263. https://doi.org/10.1038/srep43263 PMID: 28233817

32. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E,et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017; 45(D1):D896–D901. https://doi.org/10.1093/nar/gkw1133 PMID: 27899670

33. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017 Jul 3; 45(W1):W98–W102. https://doi.org/10.1093/nar/gkx247 PMID: 28407145

34. 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(11):973–7. https://doi.org/10.1038/ng.670 PMID: 20972440

35. Zhang B, Shrubsole MJ, Li G, Cai Q, Edwards T, Smalley WE, et al. Association of genetic variants for colorectal cancer differs by subtypes of polyps in the colorectum. Carcinogenesis. 2012; 33(12):2417–23. https://doi.org/10.1093/carcin/bgs308 PMID: 23027627

36. Carvajal-Carmona LG, Zauber AG, Jones AM, Howarth K, Wang J, Cheng T et al. Much of the genetic risk of colorectal cancer is likely to be mediated through susceptibility to adenomas. Gastroenterology. 2013; 144(1):53–5. https://doi.org/10.1053/j.gastro.2012.09.016 PMID: 22999960

37. Ensembl genome database project [Internet]; 2018 https://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=1:221991106-221992106;v=rs6687758;vdb=variation;vf=3766274

38. Wang H, Burnett T, Kono S, Haiman CA, Iwasaki M, Wilkens LR, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. Nat Commun. 2014; 5:4613. https://doi.org/10.1038/ncomms5613 PMID: 25105248

39. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet. 2012; 44(7):770–6. https://doi.org/10.1038/ng.2293 PMID: 22634755

40. Abe M, Ito H, Oze I, Nomura M, Ogawa Y, Matsuo K. The more from East-Asian, the better: risk prediction of colorectal cancer risk by GWAS-identified SNPs among Japanese. J Cancer Res Clin Oncol. 2017; 143(12):2481–2492. https://doi.org/10.1007/s00432-017-2505-4 PMID: 28849422

41. Duan X, Gao Y, Yang H, Feng T, Jin T, Long Y, et al. Poly morphisms in the DUSP10 gene are associated with sex-specific colorectal cancer risk in a Han population. Int J Clin Exp Pathol. 2015; 8(2):2018–25. PMID: 25973098

42. Bermudez O, Pagès G, Gimond C. The dual-specificity MAP kinas phosphatases: critical roles in development and cancer. Am J Physiol Cell Physiol. 2010; 299(2):C189–202. https://doi.org/10.1152/ajpcell.00347.2009 PMID: 20463170

43. Patterson KI, Brummer T, O'Brien PM, Daly RJ. Dual-specificity phosphatases: critical regulators with diverse cellular targets. Biochem J. 2009; 418(3):475–89. https://doi.org/10.1042/bj20082234 PMID: 19228121

44. Keyse SM. Dual-specificity MAP kinase phosphatases (MKPs) and cancer. Cancer Metastasis Rev. 2008; 27(2):253–61. https://doi.org/10.1007/s10555-008-9123-1 PMID: 18330678

45. Tanoue T, Moriguchi T, Nishida E. Molecular cloning and characterization of a novel dual specificity phosphatase, MKP-5. J Biol Chem. 1999; 274(28):19949–56. https://doi.org/10.1074/jbc.274.28.19949 PMID: 10391943

46. Nomura M, Shiiba K, Katagiri C, Kasugai I, Masuda K, Sato I, et al. Novel function of MKP-5/DUSP10, a phosphatase of stress-activated kinases, on ERK-dependent gene expression, and up regulation of its gene expression in colon carcinomas. Oncol Rep. 2012; 28(3):931–6. https://doi.org/10.3892/or.2012.1862 PMID: 22711061

47. Fang JY, Richardson BC. The MAPK signaling pathways and colorectal cancer. Lancet Oncol. 2005; 6(5):322–7. https://doi.org/10.1016/S1470-2045(05)70168-6 PMID: 15863380

48. Han J, Sun P. The pathways to tumour suppression via route p38. Trends Biochem Sci. 2007; 32(8):364–71. https://doi.org/10.1016/j.tibs.2007.06.007 PMID: 17624785

49. Png CW, Weerasooriya M, Guo J, James SJ, Poh HM, Osato M, et al. DUSP10 regulates intestinal epithelial cell growth and colorectal tumorigenesis. Oncogene. 2016; 35(2):206–17. https://doi.org/10.1038/onc.2015.74 PMID: 25772234

50. Jiménez-Martínez M, Stamatakis K, Fresno M. The Dual-Specificity Phosphatase 10 (DUSP10): Its Role in Cancer, Inflammation, and Immunity. Int J Mol Sci. 2019; 20(7). pii: E1626. https://doi.org/10.3390/ijms20071626 PMID: 30939861

51. Song N, Shin A, Jung HS, Oh JH, Kim J. Effects of interactions between common genetic variants and smoking on colorectal cancer. BMC Cancer. 2017; 17(1):869. https://doi.org/10.1186/s12885-017-3886-0 PMID: 29258461

52. Ouyang W, Ma Q, Li J, Zhang D, Ding J, Huang Y, Xing MM, et al. Benzo[a]pyrene diol-epoxide (B[a]PDE) upregulates COX-2 expression through MAPKs/AP-1 and IKK beta/NF-kappa B in mouse epidermal Cl41 cells. Mol Carcinog. 2007; 46(1):32–41. https://doi.org/10.1002/mc.20260 PMID: 16921490

53. Gunathilake MN, Lee J, Cho YA, Oh JH, Chang HJ, Sohn DK, et al. Interaction between physical activity, PITX1 rs647161 genetic polymorphism and colorectal cancer risk in a Korean population: a case-control study. Oncotarget. 2018; 9(7):7590–7603. https://doi.org/10.18632/oncotarget.24136 PMID: 29484135

54. Khoury-Shakour S, Gruber SB, Lejbkowicz F, Rennert HS, Raskin L, Pinchev M, et al. Recreational physical activity modifies the association between a common GH1 polymorphism and colorectal cancer risk. Cancer Epidemiol Biomarkers Prev. 2008; 17(12):3314–8. https://doi.org/10.1158/1055-9965.EPI-08-0062 PMID: 19064544

55. Dennis DL, Waring JL, Payeur N, Cosby C, Daudt HM. Making lifestyle changes after colorectal cancer: insights for program development. Curr Oncol. 2013; 20(6):e493–511. https://doi.org/10.3747/co.20.1514 PMID: 24311950

56. Dunlop MG, Tenesa A, Farrington SM, Ballereau S, Brewster DH, Koessler T, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. Gut. 2013; 62(6):871–81. https://doi.org/10.1136/gutjnl-2011-300537 PMID: 22490517

57. Phipps AI, Newcomb PA, Garcia-Albeniz X, Hutter CM, White E, Fuchs CS, et al. Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. Gastroenterology. 2012; 143(1):51–4.e4. https://doi.org/10.1053/j.gastro.2012.04.052 PMID: 22580541

58. Hindorff LA, Gillanders EM, Manolio TA. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. Carcinogenesis. 2011; 32(7):945–54. https://doi.org/10.1093/carcin/bgr056 PMID: 21459759

59. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007; 17(10):1520–8. https://doi.org/10.1101/gr.6665407 PMID: 17785532

60. Ortlepp JR, Lauscher J, Janssens U, Minkenberg R, Hanrath P, Hoffmann R. Analysis of several hundred genetic polymorphisms may improve assessment of the individual genetic burden for coronary artery disease. Eur J Intern Med. 2002; 13(8):485–492. https://doi.org/10.1016/s0953-6205(02)00182-6 PMID: 12446192

61. Fernandez-Rozadilla C, Cazier JB, Tomlinson IP, Carvajal-Carmona LG, Palles C, Lamas MJ, et al. A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. BMC Genomics. 2013; 14:55. https://doi.org/10.1186/1471-2164-14-55 PMID: 23350875

62. Kantor ED, Hutter CM, Minnier J, Berndt SI, Brenner H, Caan BJ, et al. Gene-environment interaction involving recently identified colorectal cancer susceptibility Loci. Cancer Epidemiol Biomarkers Prev. 2014; 23(9):1824–33. https://doi.org/10.1158/1055-9965.EPI-14-0062 PMID: 24994789

63. Abulí A, Castells A, Bujanda L, Lozano JJ, Bessa X, Hernández C, et al. Genetic Variants associated with Colorectal Adenoma Susceptibility. PLoSOne. 2016; 11(4):e0153084. https://doi.org/10.1371/journal.pone.0153084 PMID: 27078840

64. Yao K, Hua L, Wei L, Meng J, Hu J. Correlation Between CASC8, SMAD7 Polymorphisms and the Susceptibility to Colorectal Cancer: An Updated Meta-Analysis Basedon GWAS Results. Medicine (Baltimore). 2015; 94(46):e1884. https://doi.org/10.1097/MD.0000000000001884 PMID: 26579801

65. Li L, Lv L, Liang Y, Shen X, Zhou S, Zhu J, et al. Association of 8q23-24 region (8q23.3 loci and 8q24.21 loci) with susceptibility to colorectal cancer: a systematic and updated meta-analysis. Int J Clin Exp Med. 2015; 8(11):21001–13. PMID: 26885031

66. Hong Y, Wu G, Li W, Liu D, He K. A comprehensive meta-analysis of genetic associations between five key SNPs and colorectal cancer risk. Oncotarget. 2016 Nov 8; 7(45):73945–73959. https://doi.org/10.18632/oncotarget.12154 PMID: 27661122

67. Song N, Shin A, Park JW, Kim J, Oh JH. Common risk variants for colorectal cancer: an evaluation of associations with age at cancer onset. Sci Rep. 2017; 7:40644. https://doi.org/10.1038/srep40644 PMID: 28084440

68. Whiffin N, Hosking FJ, Farrington SM, Palles C, Dobbins SE, Zgaga L, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. Hum Mol Genet. 2014; 23(17):4729–37. https://doi.org/10.1093/hmg/ddu177 PMID: 24737748

69. Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet. 2014; 46(6):533–42. https://doi.org/10.1038/ng.2985 PMID: 24836286

70. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancerrisk. Hum Genet. 2012; 131(2):217–34. https://doi.org/10.1007/s00439-011-1055-0 PMID: 21761138

71. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013; 144(4):799–807.e24. https://doi.org/10.1053/j.gastro.2012.12.020 PMID: 23266556

72. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet. 2013; 45(2):191–6. https://doi.org/10.1038/ng.2505 PMID: 23263487

73. Lubbe SJ, Whiffin N, Chandler I, Broderick P, Houlston RS. Relationship between 16 susceptibility loci and colorectal cancer phenotype in 3146 patients. Carcinogenesis. 2012; 33(1):108–12. https://doi.org/10.1093/carcin/bgr243 PMID: 22045029

74. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 2008; 40(12):1426–35. https://doi.org/10.1038/ng.262 PMID: 19011631

75. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, Howarth K, 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(6):e1002105. https://doi.org/10.1371/journal.pgen.1002105 PMID: 21655089

76. Kupfer SS, Skol AD, Hong E, Ludvik A, Kittles RA, Keku TO, et al. Shared and independent colorectal cancer risk alleles in TGFβ-related genes in African and European Americans. Carcinogenesis. 2014; 35(9):2025–30. https://doi.org/10.1093/carcin/bgu088 PMID: 24753543

77. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet. 2007; 39(11):1315–7. https://doi.org/10.1038/ng.2007.18 PMID: 17934461