Evaluating the Potential of Polygenic Risk Score to Improve Colorectal Cancer Screening

Coral Arnau-Collell[1], Anna Díez-Villanueva[2], Beatriz Bellosillo[3], Josep M. Augé[4], Jenifer Muñoz[1], Elisabet Guinó[2], Leticia Moreira[1], Anna Serradesanferm[5], Àngels Pozo[5], Isabel Torà-Rocamora[5], Laia Bonjoch[1], Gemma Ibañez-Sanz[2], Mireia Obon-Santacana[2], Ferran Moratalla-Navarro[2], Rebeca Sanz-Pamplona[2], Carmen Márquez Márquez[6], Rebeca Rueda Miret[3], Rocio Pérez Berbegal[6], Gabriel Piquer Velasco[3], Cristina Hernández Rodríguez[7], CRIPREV consortium*, Jaume Grau[5], Antoni Castells[1], Josep M. Borràs[8], Xavier Bessa[6], Victor Moreno[2], and Sergi Castellví-Bel[1]

ABSTRACT

Background: Colorectal cancer has high incidence and associated mortality worldwide. Screening programs are recommended for men and women over 50. Intermediate screens such as fecal immunochemical testing (FIT) select patients for colonoscopy with suboptimal sensitivity. Additional biomarkers could improve the current scenario.

Methods: We included 2,893 individuals with a positive FIT test. They were classified as cases when a high-risk lesion for colorectal cancer was detected after colonoscopy, whereas the control group comprised individuals with low-risk or no lesions. 65 colorectal cancer risk genetic variants were genotyped. Polygenic risk score (PRS) and additive models for risk prediction incorporating sex, age, FIT value, and PRS were generated.

Results: Risk score was higher in cases compared with controls [per allele OR = 1.04; 95% confidence interval (CI), 1.02–1.06; $P < 0.0001$]. A 2-fold increase in colorectal cancer risk was observed for subjects in the highest decile of risk alleles (≥ 65), compared with those in the first decile (≤ 54; OR = 2.22; 95% CI, 1.59–3.12; $P < 0.0001$). The model combining sex, age, FIT value, and PRS reached the highest accuracy for identifying patients with a high-risk lesion [cross-validated area under the ROC curve (AUROC): 0.64; 95% CI, 0.62–0.66].

Conclusions: This is the first investigation analyzing PRS in a two-step colorectal cancer screening program. PRS could improve current colorectal cancer screening, most likely for higher at-risk subgroups. However, its capacity is limited to predict colorectal cancer risk status and should be complemented by additional biomarkers.

Impact: PRS has capacity for risk stratification of colorectal cancer suggesting its potential for optimizing screening strategies alongside with other biomarkers.

Introduction

Colorectal cancer is recognized as one of the cancers with the highest incidence and associated mortality worldwide (1). It is generally acknowledged that a vast majority of colorectal cancer cases develop from nonmalignant precursor adenomas (2). The average duration of the development of an adenoma to colorectal cancer transition is unobserved, but it is estimated to take at least 10 years (3). This long latent phase provides an excellent window of opportunity for early detection. Therefore, colorectal cancer is particularly suitable for screening. Keeping in mind the dimension of this disease, European national health systems have started population screening programs to increase early detection and improve prevention measures. Screening for colorectal cancer offers the possibility to identify the disease at an earlier stage or at a premalignant phase. For this reason, the evidence-based European Code Against Cancer recommended that men and

[1]Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain. [2]Bellvitge Biomedical Research Institute (IDIBELL), Catalan Institute of Oncology, Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP) and University of Barcelona, Barcelona, Spain. [3]Pathology Department, Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain. [4]Biochemistry and Molecular Genetics Department, Hospital Clínic of Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain. [5]Department of Preventive Medicine and Epidemiology, Clinical Institute of Internal Medicine and Dermatology, Hospital Clinic of Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona,Barcelona, Spain. [6]Gastroenterology Department, Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain. [7]Unitat de Prevenció i Registre del Càncer, Servei d'Epidemiologia i Avaluació, Hospital del Mar, Barcelona, Spain. [8]Department of Clinical Sciences, University of Barcelona and Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain.

V. Moreno and S. Castellví-Bel contributed equally to this article.

*See Acknowledgments.

Corresponding Authors: Victor Moreno, Bellvitge Biomedical Research Institute (IDIBELL), Catalan Institute of Oncology, Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP) and University of Barcelona, Avinguda de la Granvia de l'Hospitalet, 199, L'Hospitalet de Llobregat 08908, Barcelona, Spain. Phone: 349-3260-7434; E-mail: v.moreno@iconcologia.net; and Sergi Castellví-Bel, Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Rosselló 149-153, Barcelona 08036, Spain. Phone: 349-3227-5707; E-mail: sbel@clinic.cat

women over 50 years of age should participate in colorectal cancer screening. This was given effect within the European Union (EU) by the 2003 Council Recommendation on cancer screening (4).

Indeed, colorectal cancer is highly preventable by detecting and removing adenomas through colonoscopy screening, but this procedure is very costly to be implemented as population screening and has an associated morbidity (5). Intermediate screens to detect occult blood in feces such as fecal immunochemical testing (FIT) are therefore often used to select patients for colonoscopy with suboptimal sensitivity (6, 7). This two-step strategy for colorectal cancer screening is the most common worldwide (8), but results in a high false positive rate due to the suboptimal specificity of the occult blood detection, necessitating unnecessary colonoscopies (9). Therefore, additional biomarkers added to the first step to the current scenario could improve colorectal cancer screening.

As for other complex diseases, colorectal cancer is caused by both genetic and environmental factors (10). Twin studies showed that around 13% to 30% of the variation in colorectal cancer susceptibility involves inherited genetic differences (11, 12). Some of the known colorectal cancer predisposition factors were already discovered in the past two decades through genome-wide association studies (13, 14). Right after their identification, the hope was raised for genetic profiling using the combination of these common, low-penetrance genetic variants to be able to identify high-risk individuals in the population that could benefit from preventive and therapeutic interventions (15). Indeed, polygenic risk scores (PRS) combining the individual, weak effects on disease risk have been developed in the past for common diseases such as colorectal cancer. Their predictive potential was limited, most likely evidencing their usefulness but their shortcomings when used alone without other clinical or environmental data (16, 17). PRS models for colorectal cancer were developed by using individual genome-wide association study (GWAS) genetic variants (from 10 to more than 100) but have recently incorporated genome-wide data to improve risk prediction by using all available SNPs genotyped in GWAS (18). Certainly, genome-wide PRS have proven to identify individuals with risk equivalent to monogenic mutations (19), which could justify its application in health care systems.

Using PRS to screen the population at medium risk for colorectal cancer is an attractive alternative to improve current results in this setting. Frampton and colleagues demonstrated that personalized screening programs for colorectal cancer, in which eligibility was based on PRS in addition to age, had the potential to greatly reduce the number of individuals screened while still detecting nearly as many cases (20). Some more recent studies have also tested the potential application of PRS on colorectal cancer screening programs, showing its value to define a personalized, risk-adapted starting ages for screening (17, 21) or personalized screening intervals after negative findings from colonoscopy (22).

In this study, almost 3,000 individuals from the Barcelona colorectal cancer screening program were enrolled. We were able to develop a PRS and evaluate its effectiveness in determining which individuals should undergo a colonoscopy after a positive FIT result considering their colorectal cancer–associated genetic risk. This study is the first investigation analyzing PRS in a two-step colorectal cancer screening program.

Materials and Methods

Study population

Our study included 2,893 subjects aged between 50 to 69 years. Participants were recruited initially in 2009 (first round) and then during 2017 to 2019 (second round) from the Barcelona colorectal cancer screening program with a positive FIT in 3 different hospitals (23). In both rounds, individuals with a FIT-positive result (≥20 µg of hemoglobin/g of feces) were advised to undergo a colonoscopy.

The histologic classification of polyps and cancer was based on World Health Organization (WHO) criteria (24) and additional evidence (25), as summarized in **Table 1**. It included low-risk adenomas (LRA) and high-risk adenomas (HRA). All invasive carcinomas (stages I–IV) were classified as colorectal cancer as well as carcinoma *in situ* (CIS) although currently this latter lesion is classified as HRA. While developing this study, the intermediate risk adenoma (IRA) was incorporated as a new category in the European Guideline for Quality Assurance in Colorectal Cancer Screening and Diagnosis (25, 26). However, before its publication IRA was included in the HRA group. Our cohort was divided in two groups (cases and controls) taking into account the outcome of the colonoscopy and their link with colorectal cancer and their different clinical surveillance after its finding. Therefore, the cases group included high-risk lesions (HRL) such as colorectal cancer, intramucosal carcinoma, HRA, IRA, and polyposis cases, whereas the control group comprised LRA and normal examination after colonoscopy (27).

Table 1. Cancer and polyps classification used in the current study was based on WHO criteria (24) and additional evidence (25).

Classification	Criteria
Normal examination	No presence of adenomas/serrated polyps/cancer/ inflammatory bowel disease
LRA	1 or 2 lesions smaller than 10 mm, showing a tubular histology and low-grade dysplasia, or
	1 or 2 serrated lesions smaller than 10 mm, without dysplasia
IRA	3–4 adenomatous polyps smaller than 10 mm, showing a tubular histology and low-grade dysplasia, or
	1–4 adenomatous polyps between 10–19 mm showing a tubular histology and low-grade dysplasia, or
	1–4 adenomatous polyps smaller than 20 mm, with a tubulovillous or villous histology, or high-grade dysplasia and/or intramucosal carcinoma (CIS), or
	3–4 serrated polyps smaller than 10 mm, without dysplasia, or
	1–4 serrated polyps between 10–19 mm without dysplasia, or
	1–4 serrated polyps smaller than 20 mm with dysplasia
HRA	Either adenomatous or serrated polyp ≥20 mm, or
	more than 5 adenomatous or serrated polyps
CRC	CIS, or
	All invasive carcinomas (stages I–IV)

Abbreviation: CRC, colorectal cancer.

The cases group included 1,221 subjects, comprising 755 men and 466 women. Controls were 1,672 subjects with no findings relevant for colorectal cancer and included 795 men and 877 women. Cases and controls included in this study were negative for a family history of hereditary or familial colorectal cancer or adenomatous polyposis (i.e., ≥2 first-degree relatives with colorectal cancer/adenomatous polyposis or 1 first-degree relative diagnosed before the age of 60). Cases with polyposis corresponded to de novo cases. Subjects with inflammatory bowel disease were also excluded.

Standard extraction procedures were used to obtain DNA from frozen peripheral blood for all samples. Mutations in genes for germ-line predisposition to colorectal cancer with a high-penetrance could not be excluded. However, since cases did not report a relevant family history of colorectal cancer or adenomatous polyposis, the role of germline predisposition in cases can be considered less likely and all the participants were over 50 years of age. Sex, age, and FIT values of all individuals included in the study are summarized in **Table 2**. This study was conducted in alignment with the Declaration of Helsinki and approved by the corresponding institutional ethics committee (Hospital Clínic of Barcelona; Barcelona, Spain; HCB/2017/0193) and written informed consent was obtained from all individuals.

SNP genotyping and quality control

We selected and genotyped 65 SNPs previously linked with colorectal cancer risk in all available DNA samples. We did not observe any inconsistency in the genotyping results between the two inclusion rounds. We selected SNPs from association results of previous GWAS, which were available before mid-2017 (www.ebi.ac.uk/gwas/, GWAS Catalog; Supplementary Table S1). We used the Biomark 96.96 Genotyping dynamic array (Fluidigm) and TaqMan assays (Thermo Fisher Scientific) to perform genotyping. Fluidigm SNP Genotyping Analysis and PLINK software were used to assess the quality of the data. We eliminated samples and SNPs with a genotyping success rate below 90% from following analyses (including rs7229639, rs3764482). We also eliminated a monomorphic SNP (rs10904849). We also tested genotyping quality by duplicating 26 samples. In this case, 100% genotype concordance was achieved. Hardy–Weinberg equilibrium (HWE) checking for deviation of the genotype frequencies in the controls from those expected was assessed by the χ^2 test (1df). Each SNP was in HWE in controls, with the exception of rs647161 and rs12603526 ($P = 9.28e-04$ and $6.42e-04$, respectively). The possibility of genotyping artifacts was excluded also after manually inspecting genotype assignments for them. Therefore, both SNPs were not excluded. After data quality filtering, the final number of available samples was 2,829 (1,200 cases and 1,629 controls) and 62 SNPs. In the remaining individuals, 99.07% was the overall genotyping success rate.

Statistical analysis

We first evaluated the association between each SNP and colorectal cancer through unconditional logistic regression to observe how these SNPs behaved in our population and whether they followed the trend of previously reported studies. Models were unadjusted to capture the raw association of each SNP with colorectal cancer. ORs and 95% confidence intervals (CI) were derived from the model.

We then calculated a polygenic risk score (PRS). This score was defined as the count of risk alleles across all available SNPs. We used all 62 SNPs and not only those that were significant in our cohort, since all the SNPs selected in this study were previously shown to be associated with colorectal cancer. The total number of risk alleles was calculated for all samples and coded as 0, 1, or 2 for each SNP assuming an additive genetic effect. To allow for missing values in some SNPs, the PRS values were proportionally rescaled according to the number of non-missing SNPs.

An unweighted PRS was preferred since the published effects of each SNP were similar. We also explored the weighted model derived from the GWAS publications (used ORs in Supplementary Table S2) and the model fitted to our data, but the results were essentially the same, and the weights may be biased due to the winner's course effect. Therefore, we opted to use the unweighted score. A two-sided *t* test was applied to compare between cases and controls. Afterward, PRS effect was fitted in a logistic regression model to assess for genetic susceptibility comparing cases and controls. We also performed a comparison with extreme phenotypes (CRC and HRA vs. negative colonoscopy) excluding subjects with other diagnosis.

Next, we explored the combination of PRS combined with age, sex, and the FIT value to see if we could improve its predictive capacity and if it would be worthwhile to incorporate the PRS in the actual colorectal cancer screening program. To do so, we tested four different models. The first one was calculated taking into account sex and age, the second one taking into account sex, age, and PRS, the third one taking into account sex, age, and FIT value, and the fourth was a combined model taking into account all variables (sex, age, FIT value, and PRS). FIT value was used as a qualitative variable cutting it in three categories. All models were developed using a general linear model.

Table 2. Characteristics of individuals included in the study.

	Individuals n (%)	Sex Women, n (%)	Age Mean (max–min)	FIT value Mean (max–min)
CRC	123 (4.25)	43 (34.9[a])	61.41 (51–71)	2,565.6 (108–43,828)
Stage I	68 (55.28)	26 (60.47)	61.26 (51–70)	1,353.1 (115–8,592)
Stage II	21 (17.07)	5 (11.63)	61.76 (51–70)	4,701 (179–33,322)
Stage III	22 (17.89)	7 (16.27)	60.32 (51–70)	5,028.9 (108–43,828)
Stage IV	12 (9.76)	5 (11.63)	63.67 (55–69)	1,183.8 (108–4,093)
Intramucosal carcinoma	35 (1.21)	15 (0.41)	59.54 (51–70)	584.6 (126–3,982)
Polyposis	32 (1.12)	10 (0.35)	59.47 (50–70)	491.5 (102–5,944)
HRA	399 (13.79)	144 (4.98)	61.74 (50–71)	855.4 (100–29,262)
IRA	632 (21.85)	254 (8.78)	60.72 (49–70)	1,023.3 (100–45,653)
LRA	666 (23.02)	288 (9.96)	60.61 (50–70)	886.2 (101–55,811)
Normal examination	1,006 (34.76)	589 (20.32)	59.72 (49–71)	830 (100–50,579)

Abbreviations: max, maximum; min, minimum; CRC, colorectal cancer.
[a]Percent of females among all colorectal cancer cases.

The predictive accuracy of the different models was assessed with the area under the ROC curve (AUROC). We also compared AUROC values in the different models using the Delong test. To account for potential overfitting that could overestimate the effect of the different models, a 10-fold cross-validation was used to estimate AUROC. The cross-validated AUROC and Delong P were calculated with all samples combined. In each step of the cross-validation the model was estimated with 90% of the samples and predictions applied to the remaining 10%. This step was repeated 10 times. The AUROC and Delong P were calculated with the combined predictions of the 10 steps. We then selected four cut-offs (P10, P20, P80, and P90). Subsequently, we calculated the positive and negative predictive value of our "combined model" by comparing the observed case/control status and the predicted probability estimated by our "combined model".

All analyses were performed using PLINK v1.09 and R statistical software (version 3.4.1, R Foundation for Statistical Computing). All statistical tests were two-sided, and P values < 0.05 were considered statistically significant. Bonferroni correction was used for multiple testing adjustment ($P_{adjusted} = 0.05/62$, SNPs = 8.06e-04).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Results

Association tests for individual SNPs

The cases group included 1,221 subjects, comprising 755 men and 466 women. Controls were 1,672 subjects with no relevant findings related to colorectal cancer and included 795 men and 877 women. Afterward, a total of 1,200 cases and 1,629 controls were successfully genotyped for 62 SNPs previously linked with genetic susceptibility to colorectal cancer. Firstly, the frequency of colorectal cancer risk alleles between cases and controls was compared and those significantly enriched in the cases cohort were detected (Supplementary Table S2). 13 colorectal cancer SNPs showed statistically significant associations

with colorectal cancer (rs10936599, rs704017, rs1035209, rs174537, rs1535, rs174550, rs3217810, rs4939827, rs2241714, rs1800469, rs961253, rs6061231, and rs4925386) and these genetic associations were similar to those previously reported for colorectal cancer genetic susceptibility. Despite being not significant, most remaining SNPs (47/62), showed ORs in the same directions as those previously described in the literature for colorectal cancer susceptibility.

Polygenic risk score

We then calculated a PRS and compared the number of colorectal cancer risk alleles between cases and controls. **Figure 1** shows the distribution of risk by allele number for the 62 genotyped SNPs, both for cases and controls. Risk alleles followed a normal distribution in both cases and controls. It was apparent a shift towards a higher number of risk alleles in cases coherent with a cumulative impact of colorectal cancer risk alleles. The mean number of risk alleles in control individuals was 56.47 compared with 57.54 in cases and there was a statistically significant difference in the mean number of risk alleles between cases and controls (difference: -1.07; two-sided t test $P = 1.6e\text{-}07$).

Then, we calculated PRS for cases and controls taking into account the cumulative number of colorectal cancer risk alleles. We took into account 56, the median number of risk alleles in controls, as reference. We grouped cases and controls when subjects carried ≤46 risk alleles and ≥70 alleles, since they corresponded to a small number of subjects. We observed that the risk score was higher in the cases group compared with the control cohort (per allele OR = 1.04; 95% CI, 1.02–1.06; $P < 0.0001$). A 2-fold increase in colorectal cancer risk for subjects in the highest decile of risk alleles (≥65), compared with those in the first decile (≤54; OR = 2.22, 95% CI 1.59–3.12; $P < 0.0001$) was also detected. As presented in **Fig. 1**, a linear increase in risk per allele was apparent, indicating the independent additive contribution of each allele to predispose for a HRL. We also checked that there was no association between PRS and age (OR = 0.99; 95% CI, 0.92–1.07; $P = 0.86$) and sex (OR = 0.99; 95% CI, 0.91–1.07; $P = 0.75$). Results shown here correspond to the unweighted model. We also explored the
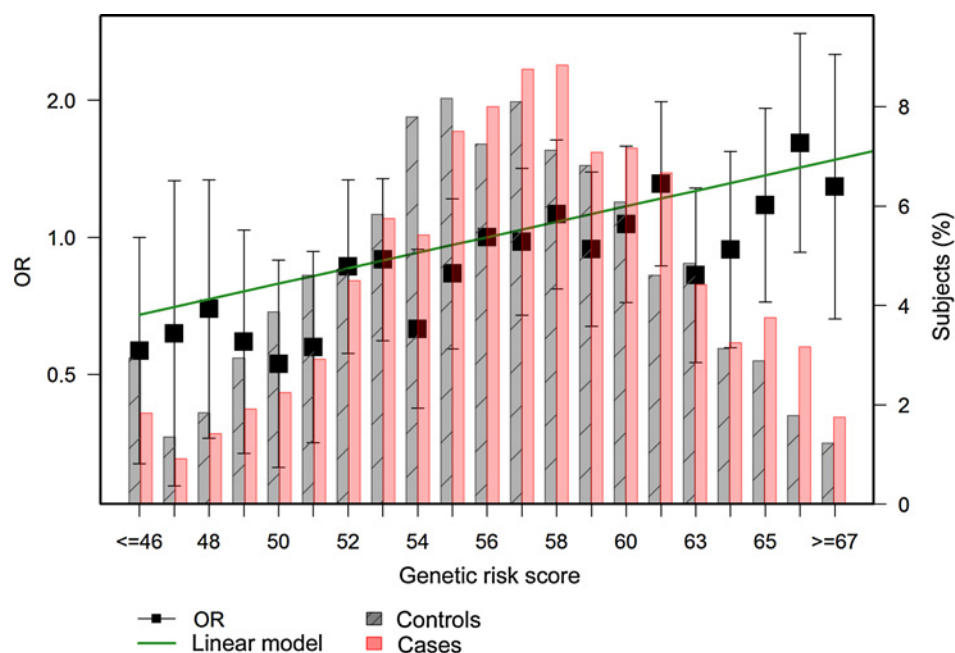

Figure 1.
Unweighted PRS. Distribution of risk by allele number for the 62 SNPs genotyped. The presence of multiple colorectal cancer risk alleles is displayed for SPS cases (bold bars) and controls (stripped bars).

weighted model derived from the GWAS publications and the model fitted to our data, but the results were essentially the same, and the weights may be biased due to the winner's course effect (Supplementary Fig. S1). Finally, a comparison of extreme phenotypes (colorectal cancer and HRA as cases vs. negative colonoscopy as controls) was also explored (Supplementary Fig. S2). When doing so, results did not change relevantly but showed some improvement. For instance, we detected that the risk score was higher in the cases group compared with the control cohort (OR = 1.05; 95% CI, 1.01–1.08; P = 0.015). We also observed that there was a 3-fold increase in colorectal cancer risk was detected for subjects in the highest decile of risk alleles (≥64), compared with those in the first decile (≤51; OR = 3.15; 95% CI, 1.06–9.41, P = 0.04).

Development and validation of a predictive model for colorectal cancer screening

Predictive modeling to discriminate between cases and controls was performed considering PRS along with age and sex, since although age and sex are not associated with PRS, they modify the risk of developing colorectal cancer. FIT value for each individual was also included in the model.

The results achieved in the PRS-based predictive model (AUROC: 0.614; 95% CI, 0.593–0.635) were superior to those obtained using only age and sex (AUROC: 0.597; 95% CI, 0.576–0.618) but lower than those obtained using age, sex, and FIT value (AUROC: 0.626; 95% CI, 0.605–0.647). Moreover, the results were statistically significant when compared with the baseline model (sex-age–based model). However, the model resulting from the combination of sex, age, FIT value, and PRS reached the highest accuracy for identifying patients with a HRL (AUROC: 0.639; 95% CI, 0.619–0.660) when compared with participants with nonrelevant findings for colorectal cancer risk at colonoscopy, illustrating the potential usefulness of incorporating PRS into the current colorectal cancer screening scenario. Results are shown in **Table 3** and **Fig. 2**.

A 10-fold cross-validation was also performed. All 10-fold cross-validated AUROC were very similar than the direct estimate of the model not supporting overfitting (Supplementary Table S3).

We then calculated the sensitivity, specificity, predictive positive and predictive negative values for four cut-offs (P10, P20, P80, P90) for the sex-age-FIT– and PRS-based model. As shown in **Table 4**, our positive predictive value (PPV) ranged from 44.25% to 61.75% and our negative predictive value (NPV) ranged from 74.11% to 59.75%. Therefore, when the cut-off point is low, the capacity to predict true positives is lower whereas the ability to detect true negatives is higher. On the other hand, as the cut-off point increases, PPV becomes higher and the NPV is reduced. For instance, comparing with the current scenario where every FIT-positive individual undergoes a colonoscopy, the P10 cut-off shows the impact in results by reducing 10% the number of colonoscopies when applying the sex-age-FIT and PRS-
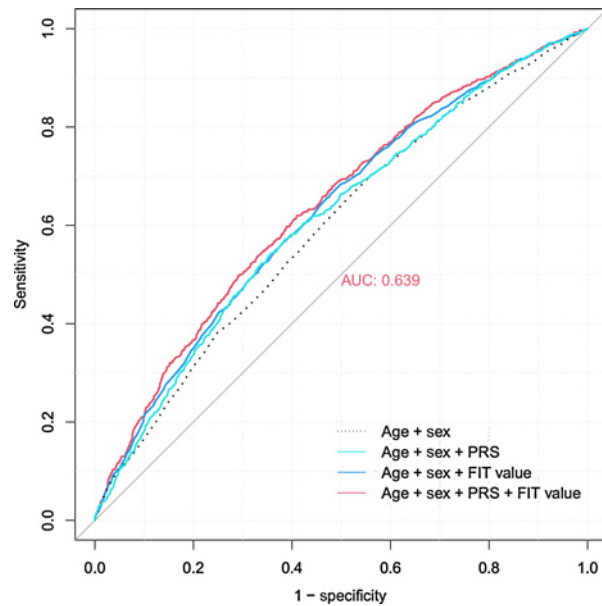

Figure 2.
AUROC of the different predictive models.

based model. By doing so, there will be still a sensitivity of approximately 94% (94% of individuals who have any relevant colorectal cancer lesion will test positive) and an improvement in specificity of approximately 13% (13% of additional individuals without any relevant colorectal cancer lesion will test negative and will not undergo a colonoscopy). Below this cut-off, the negative likelihood ratio (false negative/true negative) would be 0.48 and 209 unnecessary colonoscopies will not be performed (true negatives) but 73 HRL would not be detected (false negatives) only two of them being colorectal cancer. These results from our model show that number of colonoscopies could be significantly reduced by applying our model incorporating PRS and still maintain acceptable screening results.

As previously mentioned, we also explored a comparison of extreme phenotypes (colorectal cancer and HRA as cases vs. negative colonoscopy as controls). We calculated AUROC values (Supplementary Fig. S3) and discriminative capacity including also a 10-fold cross-validation (Supplementary Tables S4 and S5). Again, results did not change relevantly but showed some improvement, an outcome likely related to considering phenotypically extreme individuals. This outcome could be expected since SNPs used in PRS were detected by comparing colorectal cancer cases and controls in previous GWAS and not HRA, IRA, or other HRLs. However, since a smaller set of samples was used, results were not as significant.

Table 3. Discriminative capacity of predictive models.

Predictive model	AUROC (95% CI)	AUROC improvement	P
Sex–age based	0.597 (0.576–0.618)		
PRS based	0.614 (0.593–0.635)	0.0174[a]	0.0041
FIT based	0.626 (0.605–0.647)	0.0291[a]	0.00009
FIT and PRS based	0.639 (0.619–0.660)	0.0133[b]	0.0049

[a]Compared with sex-age–based.
[b]Compared with FIT-based.

Table 4. Discriminative capacity of FIT- and PRS-based predictive model.

| | FIT- and PRS-based predictive model | | | |
	Sn	Sp	PPV	NPV
P10	0.94	0.13	0.44	0.74
P20	0.88	0.26	0.47	0.74
P80	0.29	0.87	0.61	0.62
P90	0.15	0.15	0.62	0.60

Abbreviations: Sn, sensitivity; Sp, specificity.

Discussion

Estimation of colorectal cancer risk can determine who should be screened. Being screened can result in prevention, by detecting and removing precancerous lesions and colorectal cancer, providing an earlier effective treatment for this disease. Screening includes FIT and colonoscopy, and ideally they should be tailored based on personal risk. FIT is inexpensive, safe, but less sensitive, whereas colonoscopy is expensive, invasive and more risky. Colorectal cancer has a heritable fraction being in most cases polygenic with thousands of genetic variants contributing to its development. Therefore, utilizing colorectal cancer variants to predict risk holds promise for risk stratification for primary and secondary prevention. Our study has explored the potential utility of PRS to improve colorectal cancer screening through genotyping 62 SNPs previously associated to colorectal cancer in 2,829 individuals from a colorectal cancer screening program.

We first evaluated the association between each SNP and colorectal cancer. Our results showed that most of the SNPs showed ORs in the same directions as previously described in the literature for colorectal cancer susceptibility, indicating that most of the SNPs used in our study behaved as expected in our population, even those SNPs described in Asian populations.

We then compared the median number of risk alleles between cases and controls to evaluate if there was a high number of risk alleles in the case group. The obtained results suggest that there is a significant increase in the number of risk alleles in the case group compared with controls. Moreover, by comparing the individuals in the highest decile of risk alleles with those in the first decile, we observed that there was almost a 2-fold increase in colorectal cancer risk (OR = 2.22; 95% CI, 1.59–3.12; $P < 0.0001$). Thus, the higher the number of risk alleles, the higher the risk of having a HRL. Based on this, it seems feasible to identify a subgroup of participants from the Barcelona colorectal cancer screening program whose risk of having precancerous lesions or colorectal cancer is explained by a high PRS, as to warrant the application of specific surveillance measures (regular colonoscopies) to those individuals exceeding a defined value or PRS cut-off.

Regarding our predictive model for colorectal cancer screening, the directly estimated results of our models evidenced an improvement of the discrimination compared to the current predictive model used in the colorectal cancer screening program and based on FIT alone. A 10-fold cross-validation was also performed. All 10-fold cross-validated AUROC were very similar than the direct estimate of the model not supporting overfitting. However, we believe that PRS should not be considered the best biomarker to discriminate between cases and controls in a two-step colorectal cancer screening program although it shows real potential to be used along with sex, age, FIT value, and other additional biomarkers. It could be also hypothesized that a PRS-only model could be used in a different scenario to identify those individuals with a high number of risk alleles who would benefit from starting the screening program before the age of 50.

Most studies previously developed using PRS have used individuals from case-control or cohort studies not recruited specifically in a colorectal cancer screening scenario (16–18, 20, 28, 29). Likewise, there are few studies based on individuals from colorectal cancer screening programs where colonoscopy is used (21, 22, 30, 31). Accordingly, these previous studies focused mainly on improving the definition of risk-adapted screening ages. On the contrary, the individuals included in this study were part of the Barcelona colorectal cancer screening program where FIT was used as intermediate test to select who should undergo a colonoscopy when positive. Thus, we aimed at improving the discriminating power of the intermediate test used to better direct individuals to colonoscopy. Therefore, our study corresponds to the first investigation analyzing PRS in a two-step colorectal cancer screening program.

The number of genetic variants included to calculate PRS has increased as more colorectal cancer risk components were unfolded by GWAS and data meta-analyses over the years. Previous studies used the first 10 variants (16), 21 variants (29), 27 variants (28), 37 variants (20), 48 variants (21), 63 variants (17), and the final currently known 140 variants (18). Also, besides using the colorectal cancer risk linked to individual GWAS variants to estimate PRS, other recent approaches have escalated variant selection to use a selection of SNPs by linkage disequilibrium (10,000 variants), or even a genome-wide approach (18). Our study used 65 colorectal cancer risk variants since when genotyping started in 2017, it corresponded to the maximum number of variants known at that time. Even though our study did not used the maximum number of known colorectal cancer susceptibility genetic variants, a more modest sample size, or other additional approaches, it reached a similar AUROC values when compared with previous studies, exemplifying again the limited capacity of the colorectal cancer genetic susceptibility to be used in risk prediction. In fact, it should be expected from complex diseases such as colorectal cancer that risk prediction could not rely only on genetic susceptibility but also in other causes contributing to disease such as environmental factors or microbiome, and those should be taken into account in the risk prediction model.

Another matter of concern in PRS models is their transferability to all ancestral populations so their application in disease prediction in the clinical setting is robust. Most PRS have been developed and optimized using genotyping data from individuals of European ancestry and may be less accurate when risk prediction is applied to other populations (32). To circumvent this matter to some extent, we included three genetic variants (rs12080929, rs11987193, rs3987) associated with colorectal cancer risk and identified in previous GWAS performed in the Spanish population (33, 34).

This study has some limitations. It should be commented that our cohort sample size or the number of analyzed genetic variants to calculate PRS may probably be not large enough to reach stronger conclusions. Moving into affordable genome-wide genotyping in the near future could be possible and useful for the risk prediction of colorectal cancer and other diseases. It is also feasible that including additional information in the PRS model such as gut microbiota data could enhance risk prediction. It is a plausible hypothesis that certain colonic microbes or alterations of the typical resident colonic flora may create a microenvironment that is more favorable to tumor development (35). In that direction, microbiome analysis is being performed by using DNA extracted from FIT samples from this same cohort and hopefully it would add an additional layer of information to the conclusions reached in the present study. Also, additional information

if available regarding body mass index, metabolic syndrome, smoking, anti-inflammatory drugs, and antibiotic use could surely enrich our prediction model.

In summary, our study corresponds to the first investigation analyzing PRS in a two-step colorectal cancer screening program. PRS could help improve the current results in colorectal cancer screening programs. However, its capacity is limited as proven by this study and should be complemented by additional biomarkers such as microbiome or environmental factors.

Authors' Disclosures

B. Bellosillo reports personal fees from Biocartis, Novartis, Qiagen, Merck-Serono, BMS, Pfizer, Amgen, Janssen; grants and personal fees from Thermo Fisher Scientific, AstraZeneca; and personal fees from Eli Lilly and Company outside the submitted work. J.M. Borràs reports grants from Department of Health during the conduct of the study. V. Moreno reports grants from Agency for Management of University and Research Grants (AGAUR) of the Catalan Government, Instituto de Salud Carlos III, cofunded by FEDER funds; and grants from Spanish Association Against Cancer (AECC) Scientific Foundation during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

C. Arnau-Collell: Conceptualization, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing–original draft. **A. Díez-Villanueva:** Data curation, software, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **B. Bellosillo:** Resources, data curation, validation, investigation, visualization, methodology, writing–review and editing. **J.M. Augé:** Resources, data curation, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **J. Muñoz:** Data curation, validation, investigation, visualization, methodology, writing–review and editing. **E. Guinó:** Resources, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **L. Moreira:** Resources, data curation, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **A. Serradesanferm:** Resources, data curation, validation, investigation, visualization, writing–review and editing. **À. Pozo:** Resources, data curation, validation, investigation, visualization, writing–review and editing. **I. Torà-Rocamora:** Resources, data curation, validation, investigation, writing–review and editing. **L. Bonjoch:** Resources, validation, investigation, visualization, methodology, writing–review and editing. **G. Ibañez-Sanz:** Resources, data curation, software, validation, investigation, visualization, methodology, writing–review and editing. **M. Obon-Santacana:** Resources, data curation, software, validation, investigation, visualization, methodology, writing–review and editing. **F. Moratalla-Navarro:** Resources, data curation, software, validation, investigation, visualization, methodology, writing–review and editing. **R. Sanz-Pamplona:** Resources, data curation, software, validation, investigation, visualization, methodology, writing–review and editing. **C. Márquez Márquez:** Resources, data curation, investigation, visualization, writing–review and editing. **R. Rueda Miret:** Resources, data curation, investigation, visualization, writing–review and editing. **R. Pérez Berbegal:** Resources, data curation, investigation, visualization, writing–review and editing. **G. Piquer Velasco:** Resources, data curation, investigation, visualization, writing–review and editing. **C. Hernández Rodríguez:** Resources, data curation, investigation, visualization, writing–review and editing. **J. Grau:** Resources, data curation, supervision, investigation, project administration, writing–review and editing. **A. Castells:** Resources, formal analysis, supervision, funding acquisition, investigation, project administration, writing–review and editing. **J.M. Borràs:** Resources, data curation, supervision, funding acquisition, investigation, project administration, writing–review and editing. **X. Bessa:** Resources, data curation, supervision, funding acquisition, investigation, project administration, writing–review and editing. **V. Moreno:** Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration. **S. Castellví-Bel:** Conceptualization, resources, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **CRIPEV Consortium:** Resources, data curation, funding acquisition, investigation, project administration, writing–review and editing.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.

2. Morson BC. The evolution of colorectal carcinoma. Clin Radiol 1984;35:425–31.

3. Winawer SJ, Fletcher RH, Miller L, Godlee F, Stolar MH, Mulrow CD, et al. Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology 1997;112:594–642.

4. Official Journal of the European Union. The Council of the European Union Recommendation of 2 December 2003 on cancer screening [cited 2021 Sep 20]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32003H0878#x0026;qid=1649843665041#x0026;from=EN.

5. European Colorectal Cancer Screening Guidelines Working Group, von Karsa L, Patnick J, Segnan N, Atkin W, Halloran S, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. Endoscopy 2013;45:51–9.

6. Wilschut JA, Habbema JD, van Leerdam ME, Hol L, Lansdorp-Vogelaar I, Kuipers EJ, et al. Fecal occult blood testing when colonoscopy capacity is limited. J Natl Cancer Inst 2011;103:1741–51.

7. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med 2012;366:697–706.

8. Young GP, Rabeneck L, Winawer SJ. The global paradigm shift in screening for colorectal cancer. Gastroenterology 2019;156:843–51.

9. Robertson DJ, Lee JK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, et al. Recommendations on fecal immunochemical testing to screen for colorectal neoplasia: a consensus statement by the US multi-society task force on colorectal cancer. Gastroenterology 2017;152:1217–37.

10. Murphy N, Moreno V, Hughes DJ, Vodicka L, Vodicka P, Aglago EK, et al. Lifestyle and dietary environmental factors in colorectal cancer susceptibility. Mol Aspects Med 2019;69:2–9.

11. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.

12. Frank C, Sundquist J, Yu H, Hemminki A, Hemminki K. Concordant and discordant familial cancer: Familial risks, proportions and population impact. Int J Cancer 2017;140:1510–6.

13. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet 2019;51:76–87.

14. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun 2019;10:2154.

15. Kraft P, Wacholder S, Cornelis MC, Hu FB, Hayes RB, Thomas G, et al. Beyond odds ratios–communicating disease risk based on genetic profiles. Nat Rev Genet 2009;10:264–9.

16. 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;62:871–81.

17. Jeon J, Du M, Schoen RE, Hoffmeister M, Newcomb PA, Berndt SI, et al. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. Gastroenterology 2018;154:2152–64.

18. Thomas M, Sakoda LC, Hoffmeister M, Rosenthal EA, Lee JK, van Duijnhoven FJB, et al. Genome-wide modeling of polygenic risk score in colorectal cancer risk. Am J Hum Genet 2020;107:432–44.

19. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219–24.

20. Frampton MJ, Law P, Litchfield K, Morris EJ, Kerr D, Turnbull C, et al. Implications of polygenic risk for personalised colorectal cancer screening. Ann Oncol 2016;27:429–34.

21. Weigl K, Thomsen H, Balavarca Y, Hellwege JN, Shrubsole MJ, Brenner H. Genetic risk score is associated with prevalence of advanced neoplasms in a colorectal cancer screening population. Gastroenterology 2018;155:88–98.

22. Guo F, Weigl K, Carr PR, Heisser T, Jansen L, Knebel P, et al. Use of polygenic risk scores to select screening intervals after negative findings from colonoscopy. Clin Gastroenterol Hepatol 2020;18:2742–51.

23. Auge JM, Pellise M, Escudero JM, Hernandez C, Andreu M, Grau J, et al. Risk stratification for advanced colorectal neoplasia according to fecal hemoglobin concentration in a colorectal cancer screening program. Gastroenterology 2014;147:628–36.

24. Jass JR, Sobin LH. Histological typing of intestinal tumours. In: WHO International Histological Classification of Tumours. 2nd ed. Berlin-New York: Springer-Verlag; 1989.

25. Castells A, Andreu M, Binefa G, Fité A, Font R, Espinàs JA. Postpolypectomy surveillance in patients with adenomas and serrated lesions: a proposal for risk stratification in the context of organized colorectal cancer-screening programs. Endoscopy 2015;47:86–7.

26. von Karsa L, Patnick J, Segnan N. European guidelines for quality assurance in colorectal cancer screening and diagnosis. 1st ed. Luxembourg: European Commission, Publications Office of the European Union; 2010.

27. Click B, Pinsky PF, Hickey T, Doroudi M, Schoen RE. Association of colonoscopy adenoma findings with long-term colorectal cancer incidence. JAMA 2018;319:2021–31.

28. Hsu L, Jeon J, Brenner H, Gruber SB, Schoen RE, Berndt SI, et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. Gastroenterology 2015;148:1330–9.

29. 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.

30. Balavarca Y, Weigl K, Thomsen H, Brenner H. Performance of individual and joint risk stratification by an environmental risk score and a genetic risk score in a colorectal cancer screening setting. Int J Cancer 2020;146:627–34.

31. Northcutt MJ, Shi Z, Zijlstra M, Shah A, Zheng S, Yen EF, et al. Polygenic risk score is a predictor of adenomatous polyps at screening colonoscopy. BMC Gastroenterol 2021;21:65.

32. Cavazos TB, Witte JS. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. HGG Adv 2021;2:100017.

33. 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.

34. 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:e101178.

35. Saus E, Iraola-Guzmán S, Willis JR, Brunet-Vega A, Gabaldón T. Microbiome and colorectal cancer: roles in carcinogenesis and clinical potential. Mol Aspects Med 2019;69:93–106.