







## Article

# Interplay between Genome, Metabolome and Microbiome in Colorectal Cancer

Koldo Garcia-Etxebarria <sup>1,2,\*</sup>, Marc Clos-Garcia <sup>3,4</sup>, Oiana Telleria <sup>3</sup>, Beatriz Nafria <sup>4</sup>, Cristina Alonso <sup>5</sup>, Marta Iruarizaga-Lejarreta <sup>5</sup>, Andre Franke <sup>6</sup>, Anais Crespo <sup>7</sup>, Agueda Iglesias <sup>7</sup>, Joaquín Cubiella <sup>2,7</sup>, Luis Bujanda <sup>2,4</sup> and Juan Manuel Falcón-Pérez <sup>2,3,8,9</sup>

- <sup>1</sup> Grupo de Genética Gastrointestinal, Biodonostia, 20014 San Sebastián, Spain
  - <sup>2</sup> Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), 08036 Barcelona, Spain; joaquin.cubiella.fernandez@sergas.es (J.C.); luis.bujandafernandezdepierola@osakidetza.eus (L.B.); jfalcon@cicbiogune.es (J.M.F.-P.)
  - <sup>3</sup> Exosomes Laboratory, Centro de Investigación Cooperativa en Biociencias (CIC bioGUNE), 48160 Derio, Spain; marc.clos.garcia@sund.ku.dk (M.C.-G.); oiana.telleria@biodonostia.org (O.T.)
  - <sup>4</sup> Grupo de Enfermedades Gastrointestinales, Biodonostia, Universidad del País Vasco (UPV/EHU), 20014 San Sebastián, Spain; beatriz.nafrijimenez@osakidetza.eus
  - <sup>5</sup> OWL Metabolomics, Bizkaia Technology Park, 48160 Derio, Spain; calonso@owlmetabolomics.com (C.A.); miruarrizaga@owlmetabolomics.com (M.I.-L.)
  - <sup>6</sup> Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, 24105 Kiel, Germany; a.franke@ikmb.uni-kiel.de
  - <sup>7</sup> Department of Gastroenterology, Instituto de Investigación Sanitaria Galicia Sur, Complejo Hospitalario Universitario de Ourense, 32005 Ourense, Spain; anais.crespo.lois@sergas.es (A.C.); agueda.iglesias.gomez@sergas.es (A.I.)
  - <sup>8</sup> Basque Foundation for Sciences, Ikerbasque, 48013 Bilbao, Spain
  - <sup>9</sup> Metabolomics Platform, Centro de Investigación Cooperativa en Biociencias (CIC bioGUNE), 48160 Derio, Spain
- \* Correspondence: koldo.garcia@biodonostia.org



**Citation:** Garcia-Etxebarria, K.; Clos-Garcia, M.; Telleria, O.; Nafria, B.; Alonso, C.; Iruarizaga-Lejarreta, M.; Franke, A.; Crespo, A.; Iglesias, A.; Cubiella, J.; et al. Interplay between Genome, Metabolome and Microbiome in Colorectal Cancer. *Cancers* **2021**, *13*, 6216. <https://doi.org/10.3390/cancers13246216>

Academic Editors: Lisardo Bosca, Luis Bujanda, Ajay Goel and Ane Etxart

Received: 10 November 2021  
Accepted: 8 December 2021  
Published: 10 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** The development of colorectal cancer (CRC) is influenced by the environment, genetics and microbiota. Microbiome and metabolome analyses allowed for the finding of factors and markers associated with adenoma and CRC risk, but the interaction of host genomics with those omic layers remains unclear. Thus, our aim is to add host genome information to find new factors and markers associated with adenoma and CRC risk or to propose biological mechanisms involved in the risk. We found interactions between different omic layers that could be biologically relevant, and the three layers gave complementary information to predict adenoma and CRC risk. Our findings will help to find new markers for adenoma and CRC risk and to analyze biological mechanisms involved in adenoma and CRC development.

**Abstract:** Background: Colorectal cancer (CRC), a major health concern, is developed depending on environmental, genetic and microbial factors. The microbiome and metabolome have been analyzed to study their role in CRC. However, the interplay of host genetics with those layers in CRC remains unclear. Methods: 120 individuals were sequenced and association analyses were carried out for adenoma and CRC risk, and for selected components of the microbiome and metabolome. The epistasis between genes located in cholesterol pathways was analyzed; modifiable risk factors were studied using Mendelian randomization; and the three omic layers were used to integrate their data and to build risk prediction models. Results: We detected genetic variants that were associated to components of metabolome or microbiome and adenoma or CRC risk (e.g., in *LINC01605*, *PROKR2* and *CCSER1* genes). In addition, we found interactions between genes of cholesterol metabolism, and HDL cholesterol levels affected adenoma ( $p = 0.0448$ ) and CRC ( $p = 0.0148$ ) risk. The combination of the three omic layers to build risk prediction models reached high AUC values ( $>0.91$ ). Conclusions: The use of the three omic layers allowed for the finding of biological mechanisms related to the development of adenoma and CRC, and each layer provided complementary information to build risk prediction models.

**Keywords:** colorectal cancer; adenoma; metabolome; microbiome; host genome

## 1. Introduction

Colorectal cancer (CRC) is a major health concern that represents around 10% of the cancer incidence in the world [1], and it is the second cause of cancer death in developed countries [2]. CRC can be developed sporadically or as consequence of inflammatory processes [3], and its development is influenced by environmental factors, such as physical activity or diet [4], genetic factors and the microbiome composition [5]. Due to the burden that CRC entails, different screening strategies and diagnostic biomarkers have been developed to do an early detection of CRC [6].

As mentioned, genetic factors are involved in CRC risk and, therefore, they can be appropriate biomarkers. The study of genetic markers and their role in CRC risk have been carried out through the analysis of specific genetic variants [7–9] and genome-wide association studies (GWAS) [10]. In addition, the effect of lipid level on colorectal polyps [11] and CRC [12], as well as the effect of other modifiable risk factors [13], have been detected through genetic Mendelian randomization analyses. Finally, based on the information of those analyses, polygenic risk scores (PRS) have been developed to predict the risk of CRC based on the carriership of risk genetic variants, among other factors [14,15].

In the search of non-invasive biomarkers, in our previous works, we have analyzed the state, contribution and interaction of metabolome [16] and microbiome [17] to CRC risk. We found that six metabolites, among others, metabolites of cholesterol metabolism, and 16 bacteria genera that can be identified from fecal samples were able to discriminate cases from healthy controls [16,17]. In addition, we found alterations in the presence of the phyla Bacteroidete, Firmicutes and Fusobacteria in CRC or adenoma patients [17]. It has been suggested that the state of one omic layer could affect the state of other layers and their interplay could determine the final phenotype [18], e.g., the host genomic layer can influence the microbiome layer [19]. In addition, it has been proposed that the enterocytes are exposed to the metabolites produced by the diet, intestinal secretions and the digestion done by microbes; and the enterocytes can interact directly with microbes [20,21], but the effect of the host genetics in those interactions remains unclear, since, as far as we know, analyses of genome, metabolome and microbiome and their interplay are scarce in CRC, and limited to animal models [22].

Thus, our aim in this work is to add the information from host genomic layer to our previous results in metabolomic and microbiome layers to further characterize their interplay and effect in adenoma an CRC risk and development.

## 2. Materials and Methods

### 2.1. Samples

The patients included in this study were recruited from the COLONPREDICT study [23] and from the Biobank of the Instituto de Investigación Sanitaria Galicia Sur (<http://www.iisgaliciasur.es/home/biobank-iisgs/?lang=en> accessed on 3 December 2021). Those patients were referred to a colonoscopy due to gastrointestinal symptomatology; and patients with the following characteristics were excluded: age under 18, pregnancy, patients with previous history of colonic disease, patients requiring hospital admission, patients whose symptoms had ceased within three months of evaluation, and patients who declined to participate after reading the informed consent form. This study was approved by the Clinical Research Ethics Committee of Galicia (Code 2011/038).

The samples and the clinical information were obtained from COLONPREDICT study and from the biobank. Patients were classified in three categories: normal, advanced adenoma ( $\geq 10$  mm, villous histology or high-grade dysplasia) and colorectal cancer (invasive adenocarcinoma). Unfortunately, information regarding the CRC stage or the characteris-

tics of the advanced adenomas was not available in the samples obtained from the biobank. Thus, this information could not be incorporated into the analyses.

## 2.2. Data Generation

The procedure to obtain metabolome and microbiome was described elsewhere [16,17]. Briefly, metabolomics analysis was carried out analyzing chloroform/methanol extracts using a UHPLC–time-of-flight (TOF)-MS-based platform. In the case of microbiome, variable regions V1 and V2 of the 16S rRNA gene were sequenced using Illumina MiSeq v3 (Illumina Inc., San Diego, CA, USA). Both omic layers were retrieved from matched fecal samples. Fecal hemoglobin (f-Hb) concentration was determined within the COLONPREDICT study in all patients using the automated OC-SENSOR MICRO analyser (Eiken Chemical Co., Ltd., Tokyo, Japan).

DNA of the individuals included in this study was extracted from blood. The Illumina Global Screening Array on Illumina iScan high-throughput screening system was used to genotype the individuals in the Institute of Clinical Molecular Biology (Kiel, Germany). The raw intensities were then analyzed using Illumina GenomeStudio software and GenCall algorithm was applied to call the alleles.

Once genotyped, samples and markers were analyzed using the quality control (QC) criteria usually applied in our group: exclusion of samples with  $\geq 15\%$  missing rates; exclusion of markers with non-called alleles; exclusion of markers with missing call rates  $> 0.05$ ; exclusion of samples with  $\geq 5\%$  missing rates; exclusion of related samples (PI-HAT  $> 0.1875$ ); exclusion of samples whose genotyped sex could not be determined; exclusion of samples with high heterozygosity rate (more than three times SD from the mean); only autosomal SNPs were kept; removal of markers with Hardy-Weinberg equilibrium  $p < 1 \times 10^{-5}$ ; removal of markers whose P of difference in missingness between cases and control was  $< 1 \times 10^{-5}$ ; and removal of samples which were outliers, identified using principal component analysis (deviation of more than 6 times interquartile range).

The QCed samples were imputed using the Sanger Imputation service. Haplotype Reference Consortium release 1.1 was used as reference panel and the pipeline EAGLE2 + PBWT was used to do the imputation. From the imputation we retrieved markers with INFO score  $\geq 0.80$  and biallelic markers.

After genotyping, quality control and imputation, 5,411,755 SNPs from 120 individuals (39 patients with adenomas, 41 colorectal cancer cases, and 40 controls) were kept. From those individuals, 104 (36 adenomas, 34 colorectal cancer and 34 controls) had available information for metabolome and microbiome.

## 2.3. Analyses

In genetic association analyses of adenoma and CRC, all genotyped individuals were used, regardless of the availability of metabolome and microbiome information. The analyses of adenoma vs. controls (N = 79), CRC vs. controls (N = 81), CRC vs. adenoma (N = 80) and CRC + adenoma vs. controls (N = 120) were carried out using logistic regression implemented in Plink [24], adjusting by age, sex and first 4 principal components.

In addition, linear regression implemented in Plink [24], adjusting by age, sex and the first four principal components was used to analyze the proportional abundance of Bacteroidetes (N = 104) and Firmicutes (N = 104); values of the Shannon index (N = 104); and the levels of ChoE(18:1) (N = 116), ChoE(18:2) (N = 116), ChoE(20:4) (N = 116), PE(16:0/18:1) (N = 116), SM(42:3) (N = 116) and TG(54:1) (N = 116); and logistic regression, adjusting by age, sex and the first four principal components, to analyze the presence of Fusobacteria (N = 104).

Analyses of epistasis of SNPs located in genes of interest were carried out using Plink [24]. In the case of “cholesterol esters metabolism” pathway, 1860 available SNPs were interrogated for interactions between them; and in the case of “Glycerophospho-

lipid, Sphingolipid metabolism, Glycosylphosphatidylinositol(GPI)-anchor biosynthesis" pathway, 1925 SNPs.

For carrying out Mendelian randomization (MR) analyses, TwoSampleMR [25] and gsmr [26] packages from R language were used [27]. First, the instruments (SNPs) for the traits of interest that are known to affect CRC (BMI, cholesterol, triglycerides, selenium, iron, vitamin B12, metabolism, body fat percentage, waist circumference, IL6 receptor and height [13]) available in MRC-IEU (<https://gwas.mrcieu.ac.uk> accessed on 15 April 2021) were retrieved through TwoSampleMR [25], and the analysis was carried out if 10 or more instruments were available. Then HEIDI outlier analysis was used to discard heterogenous instruments. With those instruments we carried out the MR analysis using MR Egger, as done in CRC previously [13]. The analysis was done for each comparison (adenoma vs. controls, CRC vs. controls, CRC vs. adenoma, and CRC + adenoma vs. controls).

To integrate the data from the three omic layers studied in this work (genome, metabolome and microbiome) MOFA2 package [28,29] of R language [27] was used. In genome, SNPs that were genotyped and have a MAF > 0.05 were included, in total 98,966 SNPs; in metabolome, 90 metabolites were included; and in microbiome 120 microbial genera were included, after removing the less variable genera using nearZeroVar function included in the package "caret" (<https://CRAN.R-project.org/package=caret> accessed on 28 June 2021) of R language. The analysis in MOFA2 was carried out using default settings.

Polygenic risk scores (PRS) were calculated using PRSice software [30]. As base summary statistics, 397 SNPs associated to CRC retrieved from GWAS Catalog [31] were used; the additive model was tested; and the analysis was adjusted by age, sex and the first four principal components. The analyses were performed in adenoma vs. controls, CRC vs. controls, CRC vs. adenoma, and CRC + adenoma vs. controls.

To build and test predictive models using metabolome (ChoE(18:1), ChoE(18:2), ChoE(20:4), PE(16:0/18:1), SM(42:3) and TG(54:1) levels), microbiome (*Fusobacterium*, *Gemella*, *Blautia*, *Butyrivibrio*, *Clostridium*, *Coprococcus*, *Dorea*, *Peptococcus*, *Peptostreptococcus*, *Staphylococcus*, *Streptococcus*, *Parvimonas* and *Selenomonas* genera abundance) and genome (PRS and PC1-PC4) data, a logistic regression was used in R language [27] for each comparison (adenoma vs. controls, CRC vs. controls, CRC vs. adenoma, and CRC + adenoma vs. controls). The models tested were sex, age, metabolome and microbiome (Model1); age, sex, metabolome, microbiome and fecal hemoglobin (Model 1 + f-Hb); age, sex and genetics (Model 2); age, sex, genetics and fecal hemoglobin (Model 2 + f-Hb); sex, age, metabolome, microbiome and genetics (Model 1 + Model 2); and age, sex, metabolome, microbiome, genetics and fecal hemoglobin (Model 1 + Model 2 + f-Hb). To calculate the performance of the model, the sensitivity, the specificity and the area under de curve pROC package of R language [27] was used.

### 3. Results

#### 3.1. Genetic Associations of Adenoma and Colorectal Cancer

In the genetic analyses of adenoma vs. healthy controls (AD vs. C), colorectal cancer vs. healthy controls (CRC vs. C), colorectal cancer vs. adenoma (CRC vs. AD) or adenomas and colorectal cancer vs. controls (CRC + AD vs. C) there was not any genome-wide significant SNP ( $p < 5 \times 10^{-8}$ ). The best SNP in each analysis were rs7178030 ( $p = 9.76 \times 10^{-5}$ ), rs11125152 ( $p = 0.0001$ ), rs33603 ( $p = 8.59 \times 10^{-5}$ ) and rs6510888 ( $p = 1.58 \times 10^{-5}$ ), respectively.

Among the SNPs previously associated with CRC (Supplementary Table S1), 2 of them had nominal p-value ( $p < 0.05$ ), in AD vs. C, 3 in CRC vs. C, 1 in CRC vs. AD and 1 in CRC + AD vs. C. In the case of SNPs associated with CRC in Spanish cohorts, none of them were significant.

When the association of SNPs located in genes of interest were inspected significant results ( $p < 0.05$ ) were found. In the case of genes involved in cholesterol esters metabolism pathway, in AD vs. C 7 genes had SNPs with  $p < 0.05$  value, 6 in CRC vs. C, 7 in CRC vs. AD and 7 CRC + AD vs. C (Table 1). In the case of genes involved in Glycerophospholipid,

Sphingolipid metabolism, Glycosylphosphatidylinositol(GPI)-anchor biosynthesis, in AD vs. C 8 genes had SNPs with  $p < 0.05$  value, 10 in CRC vs. C, 6 in CRC vs. AD and 9 CRC + AD vs. C (Table 2).

**Table 1.** Association study results of significant SNPs in genes of “Cholesterol esters metabolism pathway”. AD vs. C, adenoma vs. healthy controls; CRC vs. C, colorectal cancer vs. healthy controls; CRC vs. AD, colorectal cancer vs. adenoma; CRC + AD vs. C, adenomas and colorectal cancer vs. controls. OR, odds-ratio; CI 95%, 95% of confidence interval of odds-ratio.

Gene	Analysis	SNP	<i>p</i>	OR (CI 95%)
<i>LIPA</i>	AD vs. C	rs2246833	0.019	2.9 (1.2–7.1)
<i>LIPA</i>	CRC vs. AD	rs885561	0.028	0.4 (0.2–0.9)
<i>NCEH1</i>	AD vs. C	rs17756312	0.019	12.3 (1.5–100.7)
<i>NCEH1</i>	CRC vs. C	rs522028	0.017	3.8 (1.3–11.2)
<i>NCEH1</i>	CRC vs. AD	rs630736	0.049	0.3 (0.1–1.0)
<i>NCEH1</i>	CRC + AD vs. C	rs522028	0.014	2.8 (1.2–6.2)
<i>CES5A</i>	AD vs. C	rs34266217	0.013	0.1 (0.0–0.6)
<i>CES5A</i>	CRC vs. C	rs34266217	0.016	0.2 (0.0–0.7)
<i>CES5A</i>	CRC vs. AD	rs11864688	0.003	0.1 (0.0–0.4)
<i>CES5A</i>	CRC + AD vs. C	rs34266217	0.004	0.2 (0.1–0.6)
<i>SOAT1</i>	CRC vs. C	rs61824389	0.012	5.4 (1.5–20.2)
<i>SOAT1</i>	CRC vs. AD	rs61824371	0.028	2.6 (1.1–6.0)
<i>ABCA1</i>	AD vs. C	rs4149297	0.011	0.2 (0.0–0.7)
<i>ABCA1</i>	CRC vs. C	rs2487060	0.029	9.4 (1.2–70.3)
<i>ABCA1</i>	CRC vs. AD	rs2515617	0.007	0.3 (0.1–0.7)
<i>ABCA1</i>	CRC + AD vs. C	rs4149297	0.011	0.3 (0.1–0.7)
<i>SCARB1</i>	AD vs. C	rs1031605	0.012	15.8 (1.8–136.2)
<i>SCARB1</i>	CRC vs. C	rs7485656	0.021	0.3 (0.1–0.8)
<i>SCARB1</i>	CRC vs. AD	rs1031605	0.006	0.2 (0.0–0.6)
<i>SCARB1</i>	CRC + AD vs. C	rs7485656	0.012	0.4 (0.2–0.8)
<i>STARD3</i>	CRC + AD vs. C	rs11556624	0.04	0.1 (0.0–0.9)
<i>NPC2</i>	AD vs. C	rs8008540	0.025	2.5 (1.1–5.6)
<i>NPC2</i>	CRC + AD vs. C	rs10139122	0.026	0.4 (0.1–0.9)
<i>OSBPL5</i>	AD vs. C	rs11025414	0.012	0.2 (0.1–0.7)
<i>OSBPL5</i>	CRC vs. C	rs7122180	0.03	0.1 (0.0–0.8)
<i>OSBPL5</i>	CRC vs. AD	rs11025414	0.003	5.5 (1.8–16.6)
<i>OSBPL5</i>	CRC + AD vs. C	rs111415444	0.035	0.3 (0.1–0.9)

**Table 2.** Association study results of significant SNPs in genes of “Glycerophospholipid, Sphingolipid metabolism, Glycosylphosphatidylinositol(GPI)-anchor biosynthesis pathway”. AD vs. C, adenoma vs. healthy controls; CRC vs. C, colorectal cancer vs. healthy controls; CRC vs. AD, colorectal cancer vs. adenoma; CRC + AD vs. C, adenomas and colorectal cancer vs. controls. OR, odds-ratio; CI 95%, 95% of confidence interval of odds-ratio.

Gene	Analysis	SNP	<i>p</i>	OR (CI 95%)
<i>LPCAT1</i>	AD vs. C	rs2962043	0.022	0.3 (0.1–0.8)
<i>LPCAT1</i>	CRC vs. C	rs13170616	0.047	0.3 (0.1–1.0)
<i>LPCAT1</i>	CRC vs. AD	rs55707452	0.02	14.4 (1.5–136.8)
<i>LPCAT1</i>	CRC + AD vs. C	rs2962043	0.018	0.4 (0.2–0.8)
<i>PLA2G12B</i>	CRC vs. AD	rs77284811	0.02	0.1 (0.0–0.6)
<i>PLA2G4C</i>	AD vs. C	rs1985341	0.016	3.2 (1.2–8.2)
<i>PLA2G4C</i>	CRC vs. C	rs8110925	0.006	0.0 (0.0–0.4)
<i>PLA2G4C</i>	CRC + AD vs. C	rs8110925	0.005	0.1 (0.0–0.5)
<i>PLB1</i>	AD vs. C	rs1528178	0.009	16.6 (2.0–138.4)
<i>PLB1</i>	CRC vs. C	rs10176136	0.006	6.7 (1.7–25.9)
<i>PLB1</i>	CRC vs. AD	rs7607457	0.001	7.6 (2.3–25.7)
<i>PLB1</i>	CRC + AD vs. C	rs80005535	0.014	0.1 (0.0–0.6)
<i>PLD1</i>	AD vs. C	rs9823312	0.022	4.2 (1.2–14.2)

Table 2. Cont.

Gene	Analysis	SNP	p	OR (CI 95%)
<i>PLD1</i>	CRC vs. C	rs416158	0.007	0.2 (0.1–0.6)
<i>PLD1</i>	CRC vs. AD	rs41273601	0.008	4.1 (1.4–11.4)
<i>PLD1</i>	CRC + AD vs. C	rs416158	0.005	0.3 (0.1–0.7)
<i>PNPLA7</i>	CRC vs. C	rs12685481	0.033	0.2 (0.0–0.9)
<i>PNPLA7</i>	CRC + AD vs. C	rs2488592	0.04	0.4 (0.2–1.0)
<i>PLPP1</i>	CRC vs. C	rs111922583	0.019	13.9 (1.6–125.3)
<i>PLPP1</i>	CRC + AD vs. C	rs4865950	0.026	0.0 (0.0–0.7)
<i>PLPP3</i>	CRC vs. C	rs149181632	0.03	0.1 (0.0–0.8)
<i>PLPP3</i>	CRC vs. AD	rs17429642	0.029	0.2 (0.0–0.8)
<i>CERS4</i>	AD vs. C	rs7250035	0.017	4.2 (1.3–13.4)
<i>CERS4</i>	CRC vs. C	rs11666971	0.01	4.5 (1.4–14.1)
<i>CERS4</i>	CRC + AD vs. C	rs7250035	0.022	3.5 (1.2–10.1)
<i>PIGK</i>	AD vs. C	rs72683932	0.024	6.4 (1.3–31.7)
<i>PIGK</i>	CRC vs. C	rs12562725	0.014	0.2 (0.1–0.7)
<i>PIGK</i>	CRC + AD vs. C	rs72683932	0.008	6.5 (1.6–26.5)
<i>PIGZ</i>	AD vs. C	rs746037	0.006	3.5 (1.4–8.6)
<i>PIGZ</i>	CRC vs. C	rs76347663	0.032	7.9 (1.2–52.3)
<i>PIGZ</i>	CRC + AD vs. C	rs746037	0.008	2.7 (1.3–5.8)
<i>SMPD3</i>	AD vs. C	rs9940621	0.048	0.3 (0.1–1.0)
<i>SMPD3</i>	CRC vs. AD	rs7200932	0.006	0.3 (0.1–0.7)

Moreover, epistatic interactions between the genes of those pathways were analyzed (Figure 1). The landscape of the interaction between genes was slightly different in each analysis, for example, the connections between *SOAT1* and *SCARB1* was different in CRC vs. C and in CRC vs. AD. On the whole, the number of interactions was higher in AD + CRC vs. C than in the rest of comparisons and all genes carried SNPs that interact with SNPs of other genes, with the exception of *CEL* and *LIPE* genes in cholesterol esters metabolism pathway; and *SMPD1* gene in Glycerophospholipid, Sphingolipid metabolism, Glycosylphosphatidylinositol(GPI)-anchor biosynthesis, as well as *PLA2G3* and *PLA2G12B* genes but only in CRC vs. C (Figure 1).

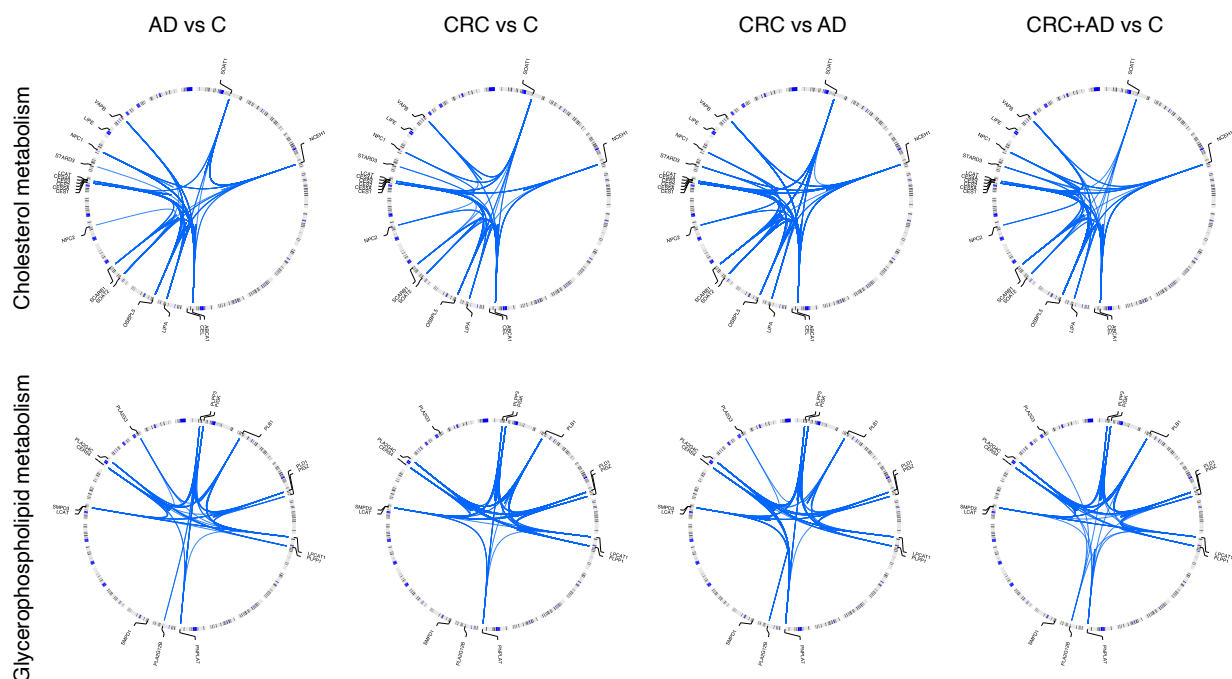


Figure 1. Epistatic interaction between SNPs of genes of cholesterol esters metabolism and Glycerophospholipid, Sphingolipid metabolism, Glycosylphosphatidylinositol(GPI)-anchor biosynthesis pathways.

### 3.2. Genetic Associations of Selected Microbiome and Metabolome Traits

Among the 34 suggestive signals found in the 4 traits related to the microbiome (abundance of Bacteroidetes 9 suggestive *loci*, abundance of Firmicutes 8, and values of Shannon index 17; while there was not any suggestive signal in the presence of Fusobacteria) there were six signals that showed nominal significance in adenoma or CRC (Table 3): rs2470641, which decreases the abundance of Bacteroidetes, decreases the risk of CRC ( $p = 0.049$ , OR = 0.3); four SNPs which were involved in the abundance of Firmicutes, were associated with higher AD risk (rs9538330 and rs7166734), or they had different frequencies in CRC and AD (rs2662642 and rs13280938); and rs73523611, which increases the Shannon diversity index, was significant in AD vs. C ( $p = 0.044$ , OR = 3.5).

**Table 3.** Suggestive SNPs associated to microbiome and metabolome traits, which are significant ( $p < 0.05$ ) in adenoma or colorectal cancer. AD vs. C, adenoma vs. healthy controls; CRC vs. C, colorectal cancer vs. healthy controls; CRC vs. AD, colorectal cancer vs. adenoma; CRC + AD vs. C, adenomas and colorectal cancer vs. controls. OR, odds-ratio; CI 95%, 95% of confidence interval of effect or odds-ratio.

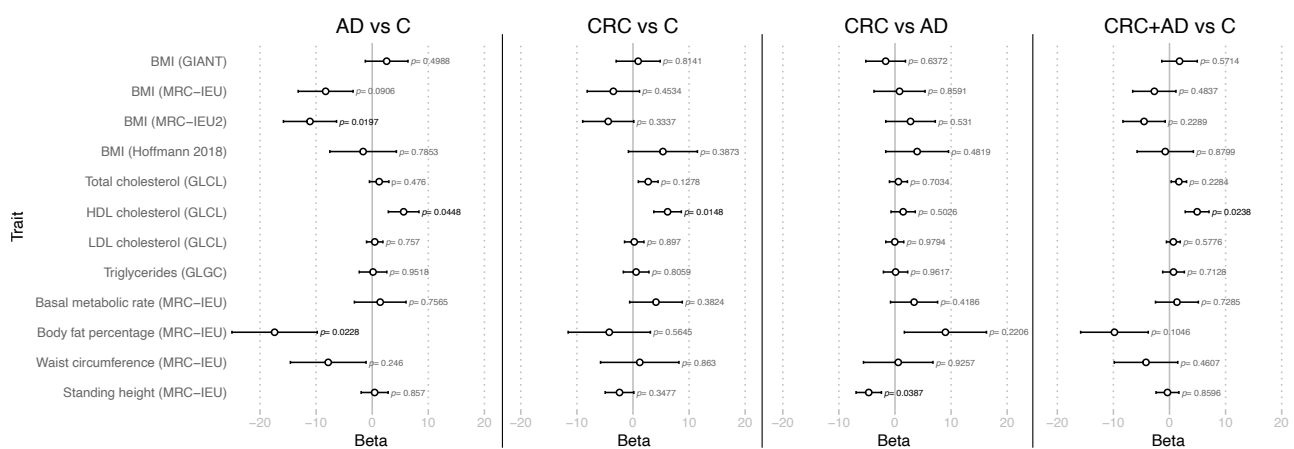
Traits	SNP	Gene	Association with Trait		Analysis	$p$	OR (CI 95%)
			$p$	Effect (CI 95%)			
Bacteroidetes	rs2470641	<i>LOC105376940</i>	$1.7 \times 10^{-6}$	−0.086 (−0.119–0.053)	CRC vs. C	0.049	0.3 (0.1–0.9)
Firmicutes	rs2662642	<i>SPAG16</i>	$1.1 \times 10^{-7}$	0.072 (0.048–0.097)	CRC vs. AD	0.034	2.2 (1.1–4.5)
Firmicutes	rs13280938	<i>LINC01605</i>	$1.6 \times 10^{-6}$	0.14 (0.087–0.194)	CRC vs. AD	0.034	6.8 (1.2–39.8)
Firmicutes	rs9538330	-	$4.6 \times 10^{-6}$	−0.073 (−0.103–0.044)	AD vs. C	0.025	3.5 (1.2–10.6)
Firmicutes	rs7166734	<i>LOC107983974</i>	$2.5 \times 10^{-6}$	−0.077 (−0.107–0.047)	AD vs. C	0.0389	2.7 (1.1–6.9)
Shannon	rs73523611	<i>CRAT37</i>	$7.66 \times 10^{-7}$	0.42 (0.26–0.57)	AD vs. C	0.044	3.5 (1.0–11.9)
ChoE(18:1)	rs142370545	-	$1.97 \times 10^{-6}$	−3.1 (−4.4–1.9)	CRC + AD vs. C	0.049	0.1 (0.01–0.9)
ChoE(18:1)	rs691563	<i>SLC39A12</i>	$1.14 \times 10^{-6}$	1.0 (0.6–1.3)	CRC vs. C	0.010	5.5 (1.5–20.4)
ChoE(18:1)	rs691563	<i>SLC39A12</i>	$1.14 \times 10^{-6}$	1.0 (0.6–1.3)	CRC + AD vs. C	0.022	3.4 (1.2–10.2)
ChoE(18:1)	rs6085078	<i>PROKR2</i>	$1.73 \times 10^{-6}$	0.69 (0.42–0.96)	CRC vs. C	0.021	2.9 (1.2–7.2)
ChoE(18:1)	rs6085078	<i>PROKR2</i>	$1.73 \times 10^{-6}$	0.69 (0.42–0.96)	CRC + AD vs. C	0.037	2.0 (1.0–3.8)
ChoE(18:2)	rs62249239	<i>LOC100130207</i>	$1.17 \times 10^{-6}$	1.7 (1.0–2.3)	CRC vs. AD	0.036	9.8 (1.1–83.6)
ChoE(20:4)	rs2146594	<i>LOC102724520</i>	$4.36 \times 10^{-6}$	−1.1 (−1.6–0.8)	CRC vs. C	0.007	0.2 (0.1–0.7)
ChoE(20:4)	rs56191847	<i>PCDH15</i>	$3.24 \times 10^{-6}$	−1.4 (−2.0–0.9)	CRC vs. C	0.020	0.3 (0.1–0.8)
ChoE(20:4)	rs56191847	<i>PCDH15</i>	$3.24 \times 10^{-6}$	−1.4 (−2.0–0.9)	CRC + AD vs. C	0.032	0.4 (0.2–0.9)
PE(16:0/18:1)	rs145994977	-	$6.89 \times 10^{-7}$	−3.5 (−4.8–2.2)	CRC + AD vs. C	0.035	0.1 (0.005–0.8)
PE(16:0/18:1)	rs17450393	<i>SNX2</i>	$1.33 \times 10^{-6}$	−3.1 (−4.3–1.9)	CRC + AD vs. C	0.036	0.1 (0.01–0.8)
PE(16:0/18:1)	rs2048236	<i>EPM2A</i>	$3.21 \times 10^{-6}$	−1.7 (−2.4–1.1)	AD vs. C	0.027	0.09 (0.01–0.8)
PE(16:0/18:1)	rs2048236	<i>EPM2A</i>	$3.21 \times 10^{-6}$	−1.7 (−2.4–1.1)	CRC vs. C	0.030	0.1 (0.004–0.8)
PE(16:0/18:1)	rs2048236	<i>EPM2A</i>	$3.21 \times 10^{-6}$	−1.7 (−2.4–1.1)	CRC + AD vs. C	0.005	0.1 (0.01–0.4)
PE(16:0/18:1)	rs73153821	<i>CREB3L2</i>	$3.31 \times 10^{-6}$	−2.4 (−3.4–1.5)	CRC + AD vs. C	0.022	0.1 (0.01–0.7)
SM(42:3)	rs17018424	<i>CCSER1</i>	$2.28 \times 10^{-6}$	−1.7 (−2.4–1.1)	CRC vs. AD	0.012	0.3 (0.1–0.8)
SM(42:3)	rs186239	-	$1.13 \times 10^{-6}$	1.7 (1.1–2.4)	CRC vs. C	0.028	5.1 (1.2–21.6)
SM(42:3)	rs186239	-	$1.13 \times 10^{-6}$	1.7 (1.1–2.4)	CRC vs. AD	0.041	2.7 (1.0–7.2)
SM(42:3)	rs4745374	<i>CARNMT1</i>	$2.77 \times 10^{-6}$	1.5 (0.9–2.2)	CRC vs. C	0.014	3.4 (1.3–9.3)
SM(42:3)	rs4745374	<i>CARNMT1</i>	$2.77 \times 10^{-6}$	1.5 (0.9–2.2)	CRC vs. AD	0.015	3.1 (1.2–7.6)
SM(42:3)	rs11058736	-	$1.53 \times 10^{-6}$	2.4 (1.4–3.3)	CRC vs. C	0.030	5.1 (1.2–21.9)
SM(42:3)	rs11058736	-	$1.53 \times 10^{-6}$	2.4 (1.4–3.3)	CRC vs. AD	0.014	7 (1.5–33.1)
SM(42:3)	rs6085078	<i>PROKR2</i>	$2.32 \times 10^{-6}$	1.3 (0.8–1.7)	CRC vs. C	0.021	2.9 (1.2–7.2)
SM(42:3)	rs6085078	<i>PROKR2</i>	$2.32 \times 10^{-6}$	1.3 (0.8–1.7)	CRC + AD vs. C	0.037	2.0 (1.0–3.8)
SM(42:3)	rs118183318	-	$1.75 \times 10^{-6}$	−2.5 (−3.5–1.5)	CRC vs. C	0.012	0.03 (0.002–0.5)
SM(42:3)	rs118183318	-	$1.75 \times 10^{-6}$	−2.5 (−3.5–1.5)	CRC vs. AD	0.027	0.1 (0.01–0.7)
SM(42:3)	rs118183318	-	$1.75 \times 10^{-6}$	−2.5 (−3.5–1.5)	CRC + AD vs. C	0.032	0.3 (0.1–0.9)
TG(54:1)	rs117956865	<i>EPS15L1</i>	$2.73 \times 10^{-8}$	−3.1 (−4.1–2.1)	CRC vs. AD	0.039	12.1 (1.1–127.7)

In the case of the metabolites analyzed, among the 168 suggestive signals (25 *loci* in ChoE(18:1), 3 in ChoE(18:2), 12 in ChoE(20:4), 41 in PE(16:0/18:1), 12 in SM(42:3), 58 in TG(54:1)), 17 signals showed an association with adenoma or CRC (Table 3). Three SNPs involved in the levels of ChoE(18:1) had  $p < 0.05$  in CRC vs. C or CRC + AD vs. C analyses. In the case of ChoE(18:2) rs62249239 had  $p < 0.05$  in CRC vs. AD ( $p = 0.036$ ; OR = 9.8). Four SNPs that affect PE(16:0/18:1) levels had  $p < 0.05$  in AD vs. C, CRC vs. C or CRC + AD vs.

C analyses. Six SNPs that affect SM(42:3) levels had  $p < 0.05$  in CRC vs. C, CRC vs. AD or CRC + AD vs. C analyses. In the case of TG(54:1), rs117956865 had an association in CRC vs. AD ( $p = 0.039$ , OR = 12.1). Among those SNPs, rs6085078 was associated with the increase of ChoE(18:1) and SM(42:3) levels and it was associated in CRC vs. C analysis ( $p = 0.021$ , OR = 2.9) and CRC + AD vs. C ( $p = 0.037$ , OR = 2.0).

### 3.3. Mendelian Randomization

Traits that have been previously described to affect CRC were used in a Mendelian randomization analysis to measure their effect in the adenoma and CRC traits of our cohort (Figure 2). In the case of AD vs. C, BMI ( $P = 0.0197$ ) and body fat percentage ( $p = 0.0228$ ) affected negatively, while HDL cholesterol affected positively ( $p = 0.0448$ ); in CRC vs. C HDL cholesterol affected positively ( $p = 0.0148$ ); in CRC vs. AD standing height affected negatively ( $p = 0.0387$ ); and in CRC + AD vs. C HDL cholesterol affected positively ( $p = 0.0238$ ).



**Figure 2.** Mendelian Randomization analyses in adenoma and colorectal cancer. AD vs. C, adenoma vs. healthy controls; CRC vs. C, colorectal cancer vs. healthy controls; CRC vs. AD, colorectal cancer vs. adenoma; CRC + AD vs. C, adenomas and colorectal cancer vs. controls. Between parenthesis, the source of the SNPs used as exposure in the analysis.

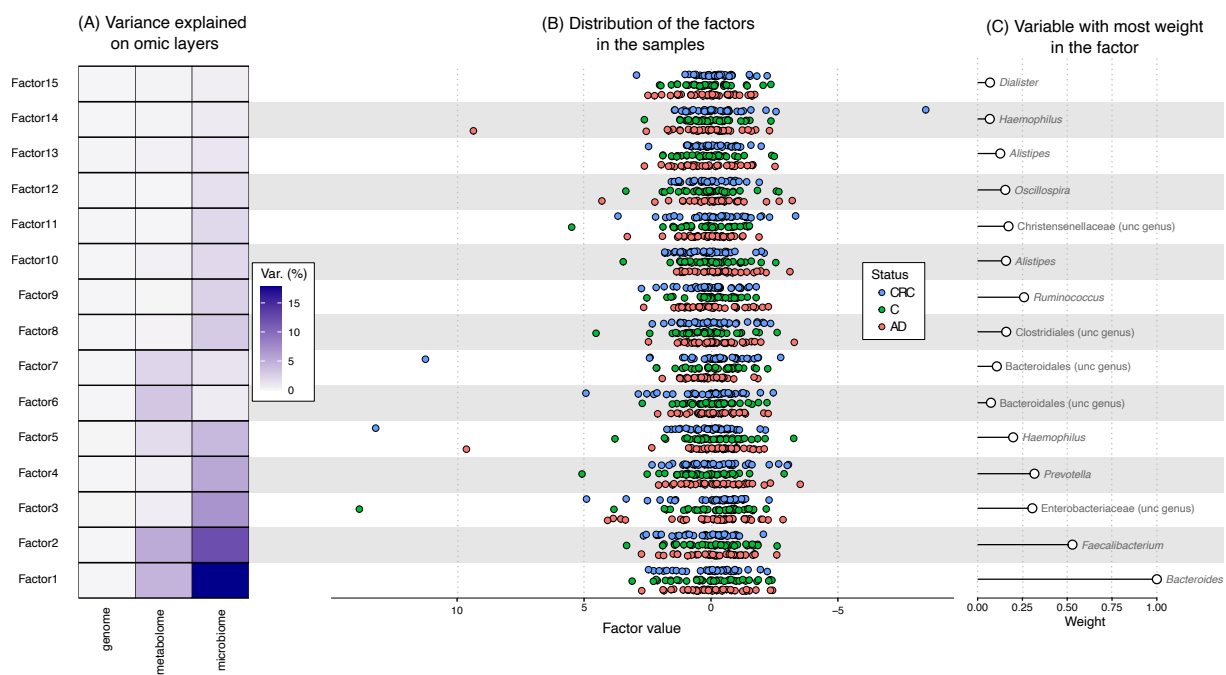
### 3.4. Multiomic Integration

The integration of the data from the three omic layers analyzed in this work showed that microbiome, represented by microbial genera, was the layer which more variance explained (Figure 3). In addition, the metabolome explained part of the variation, especially in Factors 1 and 2, while the host genome did not explain the variation. Except for Factor 1 and 2 for metabolome and microbiome, there was not any shared covariance among the different omic layers.

The distribution of the factors in the samples showed that some samples had extreme values, while most of the samples had a similar distribution of the factors regardless their status. Although some clustering can be identified (some AD samples had higher values of Factor 3 and CRC samples in Factor 6), there was not any clear difference between the different status (Figure 3).

In all the Factors the variable with most weight come from the microbiome layer (Figure 3): *Bacteroides* for Factor 1 and *Faecalibacterium* for Factor 2 were the factors with highest weight.





**Figure 3.** Factors derived from the integration of genome, metabolome and microbiome data. (A) The variance explained by each omic layer in each Factor. (B) the distribution of the values of the factors in each sample. CRC, colorectal cancer; C, healthy controls; AD, adenoma. (C) The variable with the most weight in the Factor. “Unc genus”, genus that was not classified, the identified lowest level is showed.

### 3.5. Polygenic Risk Scores and Predictive Model

Polygenic risk scores were computed using the information of SNPs previously associated to CRC and optimized for each analysis: in AD vs. CRC 147 SNPs were used, in CRC vs. C 191 SNPs, in CRC vs. AD 69 SNPs, and in CRC + AD vs. C 147 SNPs. Based on those SNPs the PRS for each individual and analysis was calculated.

Those PRS scores, as well as the genetic distance of the individuals (reflected in the first 4 PC components), were applied to build a predictive model. Compared with the previous model built by our group using metabolome and microbiome data, the performance of the use of genetic information was not better in any of the four comparisons (Table 4). The inclusion of fecal hemoglobin (f-Hb) test as predictive value improved the performance of genetic analyses, and the combination of genetics and f-Hb improved the combination of microbiome and metabolome in CRC + AD vs. control (Table 4). Finally, the combination of metabolome, microbiome and genetic data improved the predictive model in the four comparisons (Table 4).

**Table 4.** Results of the area under curve (AUC) of the receiver operating characteristic (ROC) curve in the different models tested. AD vs. C, adenoma vs. healthy controls; CRC vs. C, colorectal cancer vs. healthy controls; CRC vs. AD, colorectal cancer vs. adenoma; CRC + AD vs. C, adenomas and colorectal cancer vs. controls. Model 1, sex, age, metabolome and microbiome data as predictive values; Model 2, sex, age and genetic data (polygenic score and PC1–PC4 components) as predictive values; f-Hb, fecal hemoglobin test as predictive value.

Model	Measurement	AD vs. C	CRC vs. C	AD vs. CRC	CRC + AD vs. C
Model 1	AUC	0.91 (0.84–0.94)	0.87 (0.79–0.91)	0.76 (0.66–0.81)	0.91 (0.86–0.94)
	Specificity	0.81 (0.67–1)	0.92 (0.81–1)	0.8 (0.66–0.91)	0.94 (0.69–1)
	Sensitivity	0.94 (0.64–1)	0.83 (0.69–0.94)	0.72 (0.58–0.86)	0.79 (0.65–0.97)
Model 1 + f-Hb	AUC	1	1	0.77 (0.67–0.82)	0.81 (0.73–0.85)
	Specificity	1	1	0.66 (0.49–0.8)	0.72 (0.58–0.86)
	Sensitivity	1	1	0.89 (0.78–0.97)	0.9 (0.83–0.97)

Table 4. Cont.

Model	Measurement	AD vs. C	CRC vs. C	AD vs. CRC	CRC + AD vs. C
Model 2	AUC	0.85 (0.76–0.89)	0.85 (0.76–0.89)	0.67 (0.55–0.73)	0.84 (0.76–0.88)
	Specificity	0.82 (0.61–0.92)	0.82 (0.61–0.97)	0.7 (0.32–0.92)	0.79 (0.61–0.95)
	Sensitivity	0.86 (0.7–0.97)	0.82 (0.6–0.98)	0.7 (0.35–0.97)	0.84 (0.65–0.96)
Model 2 + f-Hb	AUC	0.92 (0.86–0.95)	0.95 (0.9–0.98)	0.75 (0.64–0.81)	0.94 (0.89–0.96)
	Specificity	0.82 (0.68–0.92)	0.95 (0.82–1)	0.8 (0.65–0.98)	0.84 (0.68–1)
	Sensitivity	0.97 (0.89–1)	0.9 (0.78–1)	0.7 (0.43–0.86)	0.94 (0.7–1)
Model 1 + Model 2	AUC	0.97 (0.94–0.99)	1	1	0.96 (0.93–0.98)
	Specificity	0.97 (0.83–1)	1	1	1 (0.89–1)
	Sensitivity	0.89 (0.78–1)	1	1	0.86 (0.77–0.97)
Model 1 + Model 2 + f-Hb	AUC	1	1	1	0.93 (0.87–0.96)
	Specificity	1	1	1	0.89 (0.78–0.97)
	Sensitivity	1	1	1	0.97 (0.93–1)

#### 4. Discussion

In the present study we have taken advantage of adding host genetic information to the omic layers that we studied previously, namely, the metabolome [16] and microbiome [17]. As far as we know, the interplay between those layers in CRC have been analyzed in animal models [22], but not in humans. On that study two mice strains were compared, since they showed differences in the number of intestinal adenomas, although one strain was the founder of the other; and the two strains have gut microbiota from two different sources [22]. The number of adenomas was influenced by the genetic variation and microbiome, and the metabolome was useful to identify the relevant biological functions affected by genetic variation [22]. In our case, it was expected that the heterogeneity of genome and microbiome would be higher and, therefore, the results could not be completely comparable. However, in both works the multi-omic approach has been useful to find biological functions relevant for the development of adenoma or colorectal cancer.

Due to our sample size the statistical power of our analyses was limited. As a consequence, we were not able to find new significant genetic variants associated to adenoma or CRC, or to find significant results for genetic variants previously associated to CRC [7–9], but we were able to find connections with other layers, to measure the effect of modifiable risk factors, and to predict the risk of CRC. In addition, we could not include histopathological characteristics to our analyses due to the lack of information in some samples, and we only compare the three categories (controls, adenoma and CRC). Although some biases could be present in the statistical analyses due to the sample size, the most relevant findings were consistent with previous works.

Among the SNPs that were associated to microbiome or metabolome traits and had nominal association with adenoma or CRC, we found genetic variants in genes associated with CRC: rs13280938 is located in the long non-coding RNA *LINC01605*, which has been described as important regulator in CRC [32]; rs6085078 of *PROKR2* gene, a receptor which is associated with colorectal progression [33]; and rs17018424 of *CCSER1* gene, a gene involved in cell division defect [34], and the overexpression of a deletion in this gene results in overexpression of adenocarcinoma associated genes [35]. Thus, our approach allowed to propose biological mechanisms that could link components of different layers. Those biological mechanisms could explain the detection of the risk of developing adenoma and colorectal cancer by different omic layers and, therefore, their utility as risk biomarkers. The present results could be a basis to suggest follow-up analyses and analyses in cell- and animal-models to understand such mechanisms and the implications for adenoma and CRC risk.

In the case of modifiable risk factors, we were able to detect the effect of the levels of HDL cholesterol, body fat percentage and BMI in adenoma and CRC, results that are consistent with previous studies [11–13]. Thus, host genome and metabolome indicated

the importance of the role of cholesterol in adenoma and CRC risk, since some relevant metabolites from the metabolome analysis were cholesterol products [16]. When we scrutinized the SNPs of genes of pathways related to cholesterol metabolism, there were some significant ( $p < 0.05$ ) SNPs, but more interestingly, the epistatic interactions between SNPs of those genes were significant. Thus, the results observed in the metabolome layer could be an effect of interaction between genes, rather than the effect of individual genetic variants; and that cholesterol levels are a risk factor also in adenoma.

When genetic information was used to build the predictive model of risk, the genetic markers used were not as good as the model built using metabolome and microbiome [17] information. However, when the information of the three layers was combined, the performance was higher than when they were separated. That could suggest that the risk factors captured by each omic layer is slightly different and complementary. This complementarity of the omic layers is consistent with the lack of covariance we observed in the integration of the three layers.

## 5. Conclusions

In conclusion, using the information of three omic layers, we were able to detect some interactions between them to propose biological mechanisms implicated in the development of adenoma and CRC which could be interesting to analyze in follow-up studies and to build an informative model for predicting the risk to adenoma and CRC.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13246216/s1>, Supplementary Table S1: Association study results of SNP previously reported to be associated with colorectal cancer.

**Author Contributions:** Conceptualization, K.G.-E., J.C., L.B., J.M.F.-P.; methodology, K.G.-E., M.C.-G.; software, K.G.-E., M.C.-G.; formal analysis, K.G.-E., M.C.-G.; resources, O.T., B.N., C.A., M.I.-L., A.F., A.C., A.I., J.C., L.B., J.M.F.-P.; data curation, M.C.-G., O.T., B.N., J.C.; writing—original draft preparation, K.G.-E.; writing—review and editing, K.G.-E., M.C.-G., J.C., L.B., J.M.F.-P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by the European Regional Development Fund (ERDF), through the funding programme INTERREG V-A España-Portugal (POCTEP) 2014–2020, Contract Number: 0181\_NANOEATERS\_1\_EP, project Cdu10 and the Instituto de Salud Carlos III PI17/00837 and DTS19/00160 to JC; Centro de Investigación Biomédica en Red en el Área temática de Enfermedades Hepáticas y Digestivas (CIBERehd) is funded by the Institute of Health Carlos III. Also, Instituto de Salud Carlos III (PI12/01604), BG2016-INVESTIGACION COLABORATIVA EN MEDICINA DE PRECISION Y BIOMARCADORES (Ref. KK-2016/00026) funded by Basque Government to JMF-P. We thank Spanish Ministry for the Severo Ochoa Excellence Accreditation (SEV-2016-0644) of CIC bioGUNE. All of them co-financed by ERDF (FEDER) Funds from the European Commission, “A way of making Europe”.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Clinical Research Ethics Committee of Galicia (Code 2011/038).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the authors. The data are not publicly available due to ethical reasons (genotype data cannot be shared).

**Conflicts of Interest:** M.I.-L. and C.A.: OWL Metabolomics’ employees. The authors declare no conflict of interest.

## References

1. Stewart, B.; Wild, C. *World Cancer Report 2014*; IARC Publications: Lyon, France, 2014.
2. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer Incidence and Mortality Worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2015**, *136*, E359–E386. [[CrossRef](#)]

3. Lasry, A.; Zinger, A.; Ben-Neriah, Y. Inflammatory networks underlying colorectal cancer. *Nat. Immunol.* **2016**, *17*, 230–240. [[CrossRef](#)]
4. Cross, A.J.; Ferrucci, L.M.; Risch, A.; Graubard, B.I.; Ward, M.H.; Park, Y.; Hollenbeck, A.R.; Schatzkin, A.; Sinha, R. A Large Prospective Study of Meat Consumption and Colorectal Cancer Risk: An Investigation of Potential Mechanisms Underlying this Association. *Cancer Res.* **2010**, *70*, 2406–2414. [[CrossRef](#)] [[PubMed](#)]
5. Wong, S.H.; Yu, J. Gut microbiota in colorectal cancer: Mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 690–704. [[CrossRef](#)]
6. Levin, B.; Lieberman, D.A.; McFarland, B.; Andrews, K.S.; Brooks, D.; Bond, J.; Dash, C.; Giardiello, F.M.; Glick, S.; Johnson, D.; et al. Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J. Clin.* **2008**, *58*, 130–160. [[CrossRef](#)] [[PubMed](#)]
7. Abulí, A.; Fernández-Rozadilla, C.; Alonso-Espinaco, V.; Muñoz, J.; Gonzalo, V.; Bessa, X.; González, D.; Clofent, J.; Cubiella, J.; Morillas, J.D.; et al. Case-control study for colorectal cancer genetic susceptibility in EPICOLON: Previously identified variants and mucins. *BMC Cancer* **2011**, *11*, 339. [[CrossRef](#)] [[PubMed](#)]
8. Abulí, A.; Bessa, X.; González, J.R.; Ruiz-Ponte, C.; Cáceres, A.; Muñoz, J.; Gonzalo, V.; Balaguer, F.; Fernández-Rozadilla, C.; González, D.; et al. Susceptibility Genetic Variants Associated With Colorectal Cancer Risk Correlate With Cancer Phenotype. *Gastroenterol.* **2010**, *139*, 788–796. [[CrossRef](#)]
9. Burns, M.B.; Montassier, E.; Abrahante, J.; Priya, S.; Niccum, D.E.; Khoruts, A.; Starr, T.K.; Knights, D.; Blekman, R. Colorectal cancer mutational profiles correlate with defined microbial communities in the tumor microenvironment. *PLoS Genet.* **2018**, *14*, e1007376. [[CrossRef](#)] [[PubMed](#)]
10. Law, P.J.; The PRACTICAL Consortium; Timofeeva, M.; Fernandez-Rozadilla, C.; Broderick, P.; Studd, J.; Fernandez-Tajes, J.; Farrington, S.; Svinti, V.; Palles, C.; et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* **2019**, *10*, 1–15. [[CrossRef](#)] [[PubMed](#)]
11. Passarelli, M.N.; Newcomb, P.A.; Makar, K.W.; Burnett-Hartman, A.N.; Potter, J.D.; Upton, M.P.; Zhu, L.-C.; Rosenfeld, M.E.; Schwartz, S.M.; Rutter, C.M. Blood lipids and colorectal polyps: Testing an etiologic hypothesis using phenotypic measurements and Mendelian randomization. *Cancer Causes Control* **2015**, *26*, 467–473. [[CrossRef](#)]
12. Rodriguez-Broadbent, H.; Law, P.J.; Sud, A.; Palin, K.; Tuupainen, S.; Gylfe, A.; Hänninen, U.A.; Cajuso, T.; Tanskanen, T.; Kondelin, J.; et al. Mendelian randomisation implicates hyperlipidaemia as a risk factor for colorectal cancer. *Int. J. Cancer* **2017**, *140*, 2701–2708. [[CrossRef](#)]
13. Cornish, A.J.; Law, P.J.; Timofeeva, M.; Palin, K.; Farrington, S.M.; Palles, C.; A Jenkins, M.; Casey, G.; Brenner, H.; Chang-Claude, J.; et al. Modifiable pathways for colorectal cancer: A mendelian randomisation analysis. *Lancet Gastroenterol. Hepatol.* **2020**, *5*, 55–62. [[CrossRef](#)]
14. Ibáñez-Sanz, G.; Díez-Villanueva, A.; Alonso, M.H.; Moranta, F.R.; Pérez-Gómez, B.; Bustamante, M.; Martin, V.; Llorca, J.; Amiano, P.; Ardanaz, E.; 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. [[CrossRef](#)]
15. Thomas, M.; Sakoda, L.C.; Hoffmeister, M.; Rosenthal, E.A.; Lee, J.K.; van Duijnhoven, F.J.; Platz, E.A.; Wu, A.H.; Dampier, C.H.; de la Chapelle, A.; et al. Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am. J. Hum. Genet.* **2020**, *107*, 432–444. [[CrossRef](#)]
16. Cubiella, J.; Clos-Garcia, M.; Alonso, C.; Martinez-Arranz, I.; Perez-Cormenzana, M.; Barrenetxea, Z.; Berganza-Granda, J.; Rodríguez-Llopis, I.; D'Amato, M.; Bujanda, L.; et al. Targeted UPLC-MS Metabolic Analysis of Human Faeces Reveals Novel Low-Invasive Candidate Markers for Colorectal Cancer. *Cancers* **2018**, *10*, 300. [[CrossRef](#)] [[PubMed](#)]
17. Clos-Garcia, M.; Garcia, K.; Alonso, C.; Iruarrizaga-Lejarreta, M.; D'Amato, M.; Crespo, A.; Iglesias, A.; Cubiella, J.; Bujanda, L.; Falcón-Pérez, J.M. Integrative Analysis of Fecal Metagenomics and Metabolomics in Colorectal Cancer. *Cancers* **2020**, *12*, 1142. [[CrossRef](#)]
18. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.; Kim, D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97. [[CrossRef](#)] [[PubMed](#)]
19. Kurilshikov, A.; Medina-Gomez, C.; Bacigalupe, R.; Radjabzadeh, D.; Wang, J.; Demirkan, A.; Le Roy, C.I.; Garay, J.A.R.; Finnicum, C.T.; Liu, X.; et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* **2021**, *53*, 156–165. [[CrossRef](#)]
20. Janney, A.; Powrie, F.; Mann, E.H. Host–microbiota maladaptation in colorectal cancer. *Nat. Cell Biol.* **2020**, *585*, 509–517. [[CrossRef](#)] [[PubMed](#)]
21. Dalal, N.; Jalandra, R.; Sharma, M.; Prakash, H.; Makharia, G.K.; Solanki, P.R.; Singh, R.; Kumar, A. Omics technologies for improved diagnosis and treatment of colorectal cancer: Technical advancement and major perspectives. *Biomed. Pharmacother.* **2020**, *131*, 110648. [[CrossRef](#)]
22. Moskowitz, J.E.; Doran, A.G.; Lei, Z.; Busi, S.B.; Hart, M.L.; Franklin, C.L.; Sumner, L.W.; Keane, T.M.; Amos-Landgraf, J.M. Integration of genomics, metagenomics, and metabolomics to identify interplay between susceptibility alleles and microbiota in adenoma initiation. *BMC Cancer* **2020**, *20*, 1–16. [[CrossRef](#)] [[PubMed](#)]

23. Cubiella, J.; on Behalf of the COLONPREDICT Study Investigators; Vega, P.; Salve, M.; Díaz-Ondina, M.; Alves, M.T.; Quintero, E.; Álvarez-Sánchez, V.; Fernández-Bañares, F.; Boadas, J.; et al. Development and external validation of a faecal immunochemical test-based prediction model for colorectal cancer detection in symptomatic patients. *BMC Med.* **2016**, *14*, 128. [[CrossRef](#)] [[PubMed](#)]
24. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **2015**, *4*, 7. [[CrossRef](#)]
25. Hemani, G.; Zheng, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **2018**, *7*, 1–29. [[CrossRef](#)]
26. Zhu, Z.; Zheng, Z.; Zhang, F.; Wu, Y.; Trzaskowski, M.; Maier, R.; Robinson, M.R.; McGrath, J.J.; Visscher, P.M.; Wray, N.R.; et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **2018**, *9*, 1–12. [[CrossRef](#)]
27. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2008.
28. Argelaguet, R.; Velten, B.; Arnol, D.; Dietrich, S.; Zenz, T.; Marioni, J.C.; Buettner, F.; Huber, W.; Stegle, O. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **2018**, *14*, e8124. [[CrossRef](#)] [[PubMed](#)]
29. Argelaguet, R.; Arnol, D.; Bredikhin, D.; Deloro, Y.; Velten, B.; Marioni, J.C.; Stegle, O. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **2020**, *21*, 1–17. [[CrossRef](#)]
30. Choi, S.W.; O'Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **2019**, *8*, 082. [[CrossRef](#)] [[PubMed](#)]
31. Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **2019**, *47*, D1005–D1012. [[CrossRef](#)]
32. Forrest, M.E.; Saiakhova, A.; Beard, L.; Buchner, D.A.; Scacheri, P.C.; LaFramboise, T.; Markowitz, S.; Khalil, A.M. Colon Cancer-Upregulated Long Non-Coding RNA lincDUSP Regulates Cell Cycle Genes and Potentiates Resistance to Apoptosis. *Sci. Rep.* **2018**, *8*, 7324. [[CrossRef](#)] [[PubMed](#)]
33. Benlahfid, M.; Traboulsi, W.; Sergent, F.; Benharouga, M.; Elhattabi, K.; Erguibi, D.; Karkouri, M.; Elattar, H.; Fadil, A.; Fahmi, Y.; et al. Endocrine gland-derived vascular endothelial growth factor (EG-VEGF) and its receptor PROKR2 are associated to human colorectal cancer progression and peritoneal carcinomatosis. *Cancer Biomark.* **2018**, *21*, 345–354. [[CrossRef](#)] [[PubMed](#)]
34. Patel, K.; Scrimieri, F.; Ghosh, S.; Zhong, J.; Kim, M.-S.; Ren, Y.R.; Morgan, R.A.; Iacobuzio-Donahue, C.A.; Pandey, A.; Kern, S.E. FAM190A Deficiency Creates a Cell Division Defect. *Am. J. Pathol.* **2013**, *183*, 296–303. [[CrossRef](#)] [[PubMed](#)]
35. Kang, S.U.; Park, J.T. Functional evaluation of alternative splicing in the FAM190A gene. *Genes Genom.* **2018**, *41*, 193–199. [[CrossRef](#)] [[PubMed](#)]