



## OPEN ACCESS

## EDITED BY

Hua Tan,  
National Human Genome Research  
Institute (NIH), United States

## REVIEWED BY

Jingjing Chen,  
Dana–Farber Cancer Institute,  
United States  
Alfonso De Stefano,  
G. Pascale National Cancer Institute  
Foundation (IRCCS), Italy

## \*CORRESPONDENCE

Sevtap Savas  
✉ savas@mun.ca

## SPECIALTY SECTION

This article was submitted to  
Cancer Genetics,  
a section of the journal  
Frontiers in Oncology

RECEIVED 12 December 2022

ACCEPTED 27 February 2023

PUBLISHED 14 March 2023

## CITATION

Curtis AA, Yu Y, Carey M, Parfrey P,  
Yilmaz YE and Savas S (2023) Multifactor  
dimensionality reduction method identifies  
novel SNP interactions in the WNT protein  
interaction networks that are associated  
with recurrence risk in colorectal cancer.  
*Front. Oncol.* 13:1122229.  
doi: 10.3389/fonc.2023.1122229

## COPYRIGHT

© 2023 Curtis, Yu, Carey, Parfrey, Yilmaz and  
Savas. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Multifactor dimensionality reduction method identifies novel SNP interactions in the WNT protein interaction networks that are associated with recurrence risk in colorectal cancer

Aaron A. Curtis<sup>1</sup>, Yajun Yu<sup>2</sup>, Megan Carey<sup>1</sup>, Patrick Parfrey<sup>3</sup>,  
Yildiz E. Yilmaz<sup>4</sup> and Sevtap Savas<sup>1,5\*</sup>

<sup>1</sup>Division of Biomedical Sciences, Faculty of Medicine, Memorial University, St. John's, NL, Canada,

<sup>2</sup>Institute of Cardiovascular Research, Southwest Medical University, Luzhou, Sichuan, China,

<sup>3</sup>Discipline of Medicine, Faculty of Medicine, Memorial University, St. John's, NL, Canada,

<sup>4</sup>Department of Mathematics and Statistics, Faculty of Science, Memorial University, St. John's,

NL, Canada, <sup>5</sup>Discipline of Oncology, Faculty of Medicine, Memorial University, St. John's, NL, Canada

**Background:** Interactions among genetic variants are rarely studied but may explain a part of the variability in patient outcomes.

**Objectives:** In this study, we aimed to identify 1 to 3 way interactions among SNPs from five Wnt protein interaction networks that predict the 5-year recurrence risk in a cohort of stage I-III colorectal cancer patients.

**Methods:** 423 patients recruited to the Newfoundland Familial Colorectal Cancer Registry were included. Five Wnt family member proteins (Wnt1, Wnt2, Wnt5a, Wnt5b, and Wnt11) were selected. The BioGRID database was used to identify the proteins interacting with each of these proteins. Genotypes of the SNPs located in the interaction network genes were retrieved from a genome-wide SNP genotype data previously obtained in the patient cohort. The GMDR 0.9 program was utilized to examine 1-, 2-, and 3-SNP interactions using a 5-fold cross validation step. Top GMDR 0.9 models were assessed by permutation testing and, if significant, prognostic associations were verified by multivariable logistic regression models.

**Results:** GMDR 0.9 has identified novel 1, 2, and 3-way SNP interactions associated with 5-year recurrence risk in colorectal cancer. Nine of these interactions were multi loci interactions (2-way or 3-way). Identified interaction models were able to distinguish patients based on their 5-year recurrence-free status in multivariable regression models. The significance of interactions was the highest in the 3-SNP models. Several of the identified SNPs were eQTLs, indicating potential biological roles of the genes they were associated with in colorectal cancer recurrence.

**Conclusions:** We identified novel interacting genetic variants that associate with 5-year recurrence risk in colorectal cancer. A significant portion of the genes identified were previously linked to colorectal cancer pathogenesis or progression. These variants and genes are of interest for future functional and prognostic studies. Our results provide further evidence for the utility of GMDR models in identifying novel prognostic biomarkers and the biological importance of the Wnt pathways in colorectal cancer.

#### KEYWORDS

Wnt pathway, recurrence, multifactor dimensionality reduction, SNP interactions, colorectal cancer

## Background

One of the most common cancers in the world is colorectal cancer (1). This disease includes the cancers of the colon and rectum, has a number of identified genetic and environmental/life-style risk factors, shows geographic difference in incidence and survival rates, and is overall characterized by moderate to low survival rates (1–5). While there are a number of disease and patient related factors that help prognosis (6, 7), identifying additional factors is needed to improve the precision of prognosis (8). Personalized/Precision Medicine approaches can improve prognosis by focusing on new biomarkers. Germline (i.e. not tumor) genetic variations, such as SNPs, are candidate biomarkers, as they are relatively stable, abundant, and show variability among individuals (9). They also have the potential to help identify the biological bases of human conditions and phenotypes. As such, there has been an emphasis on examining SNP–outcome associations in cancers, including in colorectal cancer.

An important aspect of cancer genetics that these analyses miss is that potential interactions among SNPs may be associated with patient outcomes. For example, most of the studies in colorectal cancer have so far focused on individual SNPs' associations with outcome risk/survival times (10–15). However, it is possible that a SNP's genotypes may not be associated with the outcome on its own, but they may when they exist together with another SNP's genotypes (16). Examining interactions can be challenging, however, as the number of variables examined increases, so does the need for computational resources. Methodologies (such as, Multifactor Dimensionality Reduction [MDR]) that address this challenge and tools that utilize these methods (such as GMDR 0.9) have been developed to examine interactions in a relatively feasible way (16–18). We and others have previously shown the utility of MDR and examining interactions among genetic variables in colorectal cancer (19–23). As these studies showed, examining interactions is a promising research area and can reveal new biomarkers that can predict patient outcomes.

Wnt genes have important roles in normal cellular functions as well as colorectal cancer development and its progression (24–26). Therefore, they are excellent biological candidates to examine in

colorectal cancer. In this study, our aim was to use a Generalized MDR tool (GMDR 0.9 (18)) to examine the interactions among the germline variables of five Wnt protein interaction networks in relation to 5-year recurrence-free survival status in a cohort of stage I-III colorectal cancer patients.

## Data and methods

### Ethics statement

This study was conducted with ethics approval by the Health Research Ethics Authority of Newfoundland and Labrador (HREB #2018.051; #2009.106). This study was a secondary use of data study, hence, HREB waived the requirement for patient consent.

### Clinical and genetic patient data

The baseline features of the patient cohort are shown in Table 1. The patients included in this study were recruited to the Newfoundland Familial Colorectal Cancer Registry (NFCCR) between 1999–2003, and were followed up until 2018 (11, 27, 28). Clinical and pathological data were collected by the NFCCR using medical records, tumor registry data, and other resources as described in other publications (11, 27, 28). Genetic data was a part of the previously obtained genome-wide SNP genotype data (29). PLINK (1.7) (30) was used to manage the SNP genotype data used in this study. The 5-year local or distant recurrence-free survival (RMFS) status was the response variable. All patients were unrelated to each other and of Caucasian background (29).

### Identification of Wnt interactome networks using the BioGRID database

We focused on five WNT genes in non-canonical/ $\beta$ -catenin independent pathway: *Wnt1*, *Wnt2*, *Wnt5a*, *Wnt5b*, and *Wnt11* (25). We utilized the BioGRID database (31) to retrieve the

**TABLE 1** Baseline characteristics of the patient cohort included in the GMDR and logistic regression analysis.

Variable	N (Total 423)	%
<b>Tumor Location</b>		
Colon	270	64
Rectum	153	36
<b>Stage</b>		
I	81	19
II	185	44
III	157	37
<b>Adjuvant Chemotherapy</b>		
Yes	250	59
No	173	41
<b>Adjuvant Radiotherapy</b>		
Yes	116	27
No	307	73
<b>5-Year RMFS Status</b>		
Local or distant recurrence (-)	315	74
Local or distant recurrence (+)	108	26

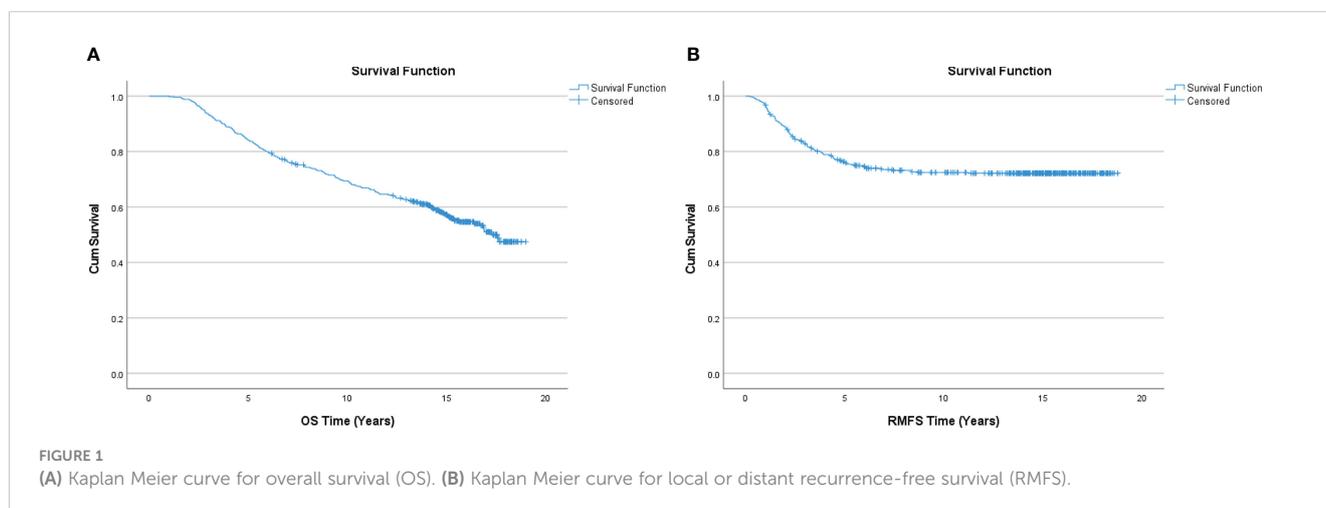
RMFS, local or distant recurrence free survival.

interaction networks for each of these genes. [Table S1](#) shows the genes in each protein interaction network after implementing quality control measures [described in detail in [Curtis et al. \(23\)](#)]. We then used the SNP genotype data and PLINK to retrieve the SNPs located in these genes using the following criteria: Minor Allele Frequency (MAF)  $\geq 0.05$ ; Hardy-Weinberg Equilibrium (HWE)  $> 0.0001$ ; and missing genotype data = 0%. Pruning was performed by PLINK to remove the SNPs in high-LD, as explained in [Curtis et al. \(23\)](#). As a result we ended up with 6 (in *Wnt1*), 18 (in *Wnt2*), 53 (in *Wnt5a*), 4 (in *Wnt5b*), and 20 (in *Wnt11*) genes

respectively ([Supplementary Table S1](#)), and around 1,000 SNPs in these Wnt interaction networks ([Supplementary Table S2](#)).

## GMDR 0.9 runs, permutation testing, and statistical analyses

GMDR 0.9 ([18, 32](#)) was previously downloaded from the UAB Department of Biostatistics Section on Statistical Genetics website. All analyses were done as described in [Curtis et al. \(23\)](#). In brief, we have conducted 1-way interactions (examining the interactions among the three potential genotypes of single SNPs); 2-way interactions (examining the interactions among the genotypes of two SNPs); and 3-way interactions (examining the interactions among the genotypes of three SNPs). Known prognostic markers, disease stage, tumor location, and adjuvant chemotherapy and radiotherapy statuses were used as variables. A 5-step cross-validation was implemented; where the dataset was partitioned into five parts, with four parts used as the training cohort and the remaining 5<sup>th</sup> part used as the testing (i.e. validation) cohort. Each of the five partitions was designated the testing cohort once and the results for each of these datasets were compared. GMDR analysis was repeated 20 times, using different random seeds, and the MDR model that was identified most frequently and with the highest Testing Balance Accuracy (TBA) score was selected as the top model for the examined interaction. Rarely, we used the higher Cross Validation Consistency (CVC) and specificity data to break a tie. We performed 1-way analysis iteratively for each dataset, and if a model was found to be significant (i.e. a SNP with a main effect was identified in the dataset), such SNPs were removed from the dataset. This was repeated until no significant 1-way model was found ([16, 33](#)). The significance of the top model was then examined using permutation testing. Models that were significant ( $p < 0.001$ ) were examined in multivariable logistic regression models, adjusting for disease stage, tumor location, and adjuvant chemotherapy and radiotherapy status. A  $p$ -value  $< 0.05$  was considered significant in multivariable logistic regression models. Kaplan Meier curves were created for overall survival (OS) and RMFS using the long-term follow up data ([28](#)) and



end point status (in OS, the end point was death from any cause). In addition, 18 patients who were censored before or at the 5 year time point and as such were excluded from the GMDR and logistic regression analyses, were included in the Kaplan Meier analyses to limit bias. Figure 1 shows the OS and RMFS curves for these patients. The 5-year OS probability was around 80% and a little bit less than that for the RMFS. R (version 3.5.3) (34) and SPSS (version 28.0.0.0 (190); 35) were used for data processing and statistical analyses.

## Functional annotation of genes and SNPs using databases

Identified SNPs and genes were examined for their functional and biological features using literature and databases. Specifically, SNPs that were identified in 1-SNP, 2-SNP, and 3-SNP interaction analyses were checked for their functional consequences in RegulomeDB (v2.0.3) (36), which uses a ranking system to describe the regulatory potential of SNPs (ranks 1a-1f specify expression quantitative trait loci [eQTLs]). In addition, SNPs were searched in the GTEx (data release v8) (37) database to see whether they were eQTLs. The latter database provided eQTL data for colon tissues only. The dbCPCO database (38) was utilized to search whether the identified SNPs were previously reported to be associated with clinical outcomes in colorectal cancer. Information about the genes/proteins was collected from literature and the Gene Entrez database (39). The dbSNP database (40) was utilized to retrieve information on SNP annotations (e.g. intronic, missense).

## Results

The clinical features of the patient cohort are summarized in Table 1.

By examining around 26,298,702 interactions (2-SNP interactions=173,425; 3-SNP interactions=26,125,277) in five Wnt interactome networks, our investigation identified 32 novel interactions associated with 5-year RMFS status in colorectal cancer.

Interactions were identified in each of the networks examined. Twenty-three 1-SNP interactions, where the genotypes of individual SNPs were identified as associated with the 5-year RMFS status in multivariable logistic regression models are shown in Supplementary Table S3. Additionally, we identified nine multi-SNP interactions (four 2-way and five 3-way interactions) that were associated with the 5-year RMFS status in the patient cohort, when adjusted for other prognostic markers in the logistic regression models (Table 2). Kaplan Meier curves for these interactions are shown in Figure 2 (for the Wnt5a network) and Supplementary Figure S1 (for the Wnt1, Wnt2, Wnt5b, and Wnt11 networks).

In all logistic regression models, the direction of the Odds Ratios (ORs) were consistent with the high-risk and low-risk genotype combinations identified in the top MDR models.

As observed in other studies, as the order of interactions increased (from 1-SNP to 3-SNP), the significance of the association detected increased as well. In other words, the 3-SNP interaction models were the ones that best separated patients based

on their 5-year RMFS status (p-values  $10^{-7}$  to  $10^{-12}$ ). ORs were also higher in 3-way models compared 1-SNP and 2-SNP models (Table 2).

All SNPs identified in interaction models were common in the patient cohort (MAF  $\geq$  5%). Only one of the SNPs (FUCA2.rs11155297, NP\_114409.2:p.Ala233Glu) was a missense variant, whereas the other variants were intronic/non-coding or 3'-UTR variants (Supplementary Table S4). Interestingly, some of the SNPs were predicted to be functional (i.e. eQTLs) based on the RegulomeDB scores or GTEx data (Table 3). Another interesting finding was that both of the SNPs in a 2-way interaction were eQTLs (in sigmoid and transverse colon tissues): FUCA2.rs11155297 (eQTL for *ADAT2*) and TMED7.rs10075869 (eQTL for *AC010226.4*). Some of the genes in the interaction networks as well as those that are associated with eQTLs had literature findings indicative of their involvement in pathogenesis of colorectal cancer or its progression (Supplementary Table S5).

## Discussion

As a globally common disease with moderate to low survival rates (1–5), it is critical to identify new biomarkers that can help prognosis in colorectal cancer. One under-studied but promising research area is that of interactions, where multiple variables together relate to a phenotype, such as recurrence status in cancer patients. Here, we report our study that used GMDR 0.9 (18), a data reduction tool, to examine interactions among different SNPs from the Wnt interactome genes in a cohort of colorectal cancer patients from Newfoundland and Labrador (27). As a result, we were able to identify novel SNP interactions in the five WNT interactome sets that were predictive of 5-year RMFS status in colorectal cancer.

Our results underline the utility of MDR in identifying previously unknown interactions and potential prognostic biomarkers. As shown in Table 2 and Supplementary Table S3, GMDR 0.9, permutation testing, and multivariable logistic regression analyses identified interactions that can distinguish patients based on their 5-year RMFS risk when adjusted for prognostic covariates. Note that high and low risk patient group classifications made by GMDR 0.9 were supported by the direction of effect (i.e. ORs) in the regression models. This increases the confidence in the GMDR 0.9 risk classification system. Among the interactions identified, 2-SNP and 3-SNP interactions (n=9) are the most interesting ones, as they significantly contribute knowledge to the largely unknown multi-loci interactions in colorectal cancer. According to the dbCPCO database (38), only one of the SNPs identified in this study – HSPA5.rs12009 – was investigated in relation to colorectal cancer outcomes in a stage II-III patient cohort. It was not associated with either stage or time-to-recurrence in the study patient cohort which consisted of mixed-ethnicities (41). Of note, in our study HSPA5.rs12009 was identified in both WNT2 and WNT5A pathway analyses and is an eQTL. Prognostic associations of this and other SNPs identified in this study can be verified in additional colorectal cancer cohorts.

It is hard to predict the biological reasons behind such interactions without detailed experimental analyses. However,

TABLE 2 Results of the two-way and three-way interaction analyses.

a) Wnt1					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression 95% Confidence Interval (CI)
2-way					
ROR2.rs7037255_A, SFRP1.rs7843510_G	rs7037255_A = GG and rs7843510_G = GA or rs7037255_A = AG and rs7843510_G = GG or rs7037255_A = AA and rs7843510_G = AA or rs7037255_A = AA and rs7843510_G = GA or rs7037255_A = AA and rs7843510_G = GG → Low Risk <b>rs7037255_A = GG and rs7843510_G = AA or rs7037255_A = GG and rs7843510_G = GG or rs7037255_A = AG and rs7843510_G = AA or rs7037255_A = AG and rs7843510_G = GA → High Risk</b>	<0.001	4.587E-06	3.362	2.002 - 5.647
3-way					
LRP6.rs11609634_T, ROR2.rs7037255_A, UBR3.rs11691281_G	rs11609634_T = CC and rs7037255_A = GG and rs11691281_G = TT or rs11609634_T = CC and rs7037255_A = GG and rs11691281_G = GT or rs11609634_T = CC and rs7037255_A = AG and rs11691281_G = GG or rs11609634_T = CC and rs7037255_A = AA and rs11691281_G = TT or rs11609634_T = CC and rs7037255_A = AA and rs11691281_G = GT or rs11609634_T = CC and rs7037255_A = AA and rs11691281_G = GG or rs11609634_T = TC and rs7037255_A = GG and rs11691281_G = TT or rs11609634_T = TC and rs7037255_A = AA and rs11691281_G = TT or rs11609634_T = TC and rs7037255_A = AA and rs11691281_G = GT or rs11609634_T = TC and rs7037255_A = AA and rs11691281_G = GG or rs11609634_T = TT and rs7037255_A = GG and rs11691281_G = TT or rs11609634_T = TT and rs7037255_A = GG and rs11691281_G = GT or rs11609634_T = TT and rs7037255_A = GG and rs11691281_G = GG or rs11609634_T = TT and rs7037255_A = AG and	<0.001	9.407E-09	4.573	2.722 - 7.685

(Continued)

TABLE 2 Continued

a) Wnt1					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression 95% Confidence Interval (CI)
	rs11691281_G = TT or rs11609634_T = TT and rs7037255_A = AG and rs11691281_G = GG or rs11609634_T = TT and rs7037255_A = AA and rs11691281_G = TT or rs11609634_T = TT and rs7037255_A = AA and rs11691281_G = GG → Low Risk rs11609634_T = CC and rs7037255_A = GG and rs11691281_G = GG or rs11609634_T = CC and rs7037255_A = AG and rs11691281_G = TT or rs11609634_T = CC and rs7037255_A = AG and rs11691281_G = GT or rs11609634_T = TC and rs7037255_A = GG and rs11691281_G = GT or rs11609634_T = TC and rs7037255_A = GG and rs11691281_G = GG or rs11609634_T = TC and rs7037255_A = AG and rs11691281_G = TT or rs11609634_T = TC and rs7037255_A = AG and rs11691281_G = GT or rs11609634_T = TC and rs7037255_A = AG and rs11691281_G = GG or rs11609634_T = TT and rs7037255_A = AG and rs11691281_G = GT or rs11609634_T = TT and rs7037255_A = AA and rs11691281_G = GT → High Risk				
b) Wnt2					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
2-way					
GPC1.rs12695020_G, WLS.rs2116046_C	rs12695020_G = AA and rs2116046_C = TT or rs12695020_G = AA and rs2116046_C = CC or rs12695020_G = GA and rs2116046_C = CT or rs12695020_G = GA and rs2116046_C = CC or rs12695020_G = GG and rs2116046_C = TT → Low Risk rs12695020_G = AA and rs2116046_C = CT or rs12695020_G = GA and rs2116046_C = TT or rs12695020_G = GG and rs2116046_C = CT	0.004	0.00004949	2.640	1.652 - 4.220

(Continued)

TABLE 2 Continued

b) Wnt2					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
	or rs12695020_G = GG and rs2116046_C = CC → High Risk				
3-way					
HCK.rs980368_G, PPP6R3.rs2840367_C, SORL1.rs3862606_G	rs980368_G = AA and rs2840367_C = CT and rs3862606_G = AA or rs980368_G = AA and rs2840367_C = CT and rs3862606_G = GA or rs980368_G = AA and rs2840367_C = CC and rs3862606_G = AA or rs980368_G = AA and rs2840367_C = CC and rs3862606_G = GA or rs980368_G = GA and rs2840367_C = TT and rs3862606_G = AA or rs980368_G = GA and rs2840367_C = CT and rs3862606_G = GA or rs980368_G = GA and rs2840367_C = CT and rs3862606_G = GG or rs980368_G = GA and rs2840367_C = CC and rs3862606_G = GG or rs980368_G = GG and rs2840367_C = TT and rs3862606_G = GA or rs980368_G = GG and rs2840367_C = TT and rs3862606_G = GG or rs980368_G = GG and rs2840367_C = CT and rs3862606_G = AA or rs980368_G = GG and rs2840367_C = CT and rs3862606_G = GA or rs980368_G = GG and rs2840367_C = CC and rs3862606_G = AA or rs980368_G = GG and rs2840367_C = CC and rs3862606_G = GA → Low Risk rs980368_G = AA and rs2840367_C = TT and rs3862606_G = AA or rs980368_G = AA and rs2840367_C = TT and rs3862606_G = GA or rs980368_G = AA and rs2840367_C = TT and rs3862606_G = GG or rs980368_G = AA and rs2840367_C = CT and rs3862606_G = GG or rs980368_G = AA and rs2840367_C = CC and rs3862606_G = GG or rs980368_G = GA and rs2840367_C = TT and rs3862606_G = GA or rs980368_G = GA and rs2840367_C = TT and rs3862606_G = GG or rs980368_G = GA and rs2840367_C = CT and rs3862606_G = AA or rs980368_G = GA and rs2840367_C = CC and rs3862606_G = AA or rs980368_G = GA and rs2840367_C = CC and rs3862606_G = GA or rs980368_G = GG and rs2840367_C = TT and rs3862606_G = AA or rs980368_G = GG and rs2840367_C = CT and rs3862606_G = GG → High Risk	<0.001	9.244E-10	4.933	2.959 - 8.222

(Continued)

TABLE 2 Continued

c) Wnt5a					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
2-way					
FSTL1.rs1402372_T, ST14.rs704625_G	rs1402372_T = GG and rs704625_G = GG or rs1402372_T = TG and rs704625_G = GC or rs1402372_T = TG and rs704625_G = GG or rs1402372_T = TT and rs704625_G = CC or rs1402372_T = TT and rs704625_G = GG → Low Risk rs1402372_T = GG and rs704625_G = CC or rs1402372_T = GG and rs704625_G = GC or rs1402372_T = TG and rs704625_G = CC or rs1402372_T = TT and rs704625_G = GC → High Risk	<0.001	1.270E-06	3.281	2.029 - 5.306
3-way					
LRP6.rs10743980_T, WLS.rs2915124_C, WNT5B.rs10848523_A	rs10743980_T = CC and rs2915124_C = TT and rs10848523_A = AG or rs10743980_T = CC and rs2915124_C = TT and rs10848523_A = AA or rs10743980_T = CC and rs2915124_C = CT and rs10848523_A = GG or rs10743980_T = CC and rs2915124_C = CT and rs10848523_A = AG or rs10743980_T = CC and rs2915124_C = CC and rs10848523_A = AA or rs10743980_T = TC and rs2915124_C = TT and rs10848523_A = AG or rs10743980_T = TC and rs2915124_C = TT and rs10848523_A = AA or rs10743980_T = TC and rs2915124_C = CT and rs10848523_A = GG or rs10743980_T = TC and rs2915124_C = CT and rs10848523_A = AA or rs10743980_T = TC and rs2915124_C = CC and rs10848523_A = GG or rs10743980_T = TC and rs2915124_C = CC and rs10848523_A = AG or rs10743980_T = TC and rs2915124_C = CC and rs10848523_A = AA or rs10743980_T = TT and rs2915124_C = TT and rs10848523_A = GG or rs10743980_T = TT and rs2915124_C = CT and rs10848523_A = GG or rs10743980_T = TT and rs2915124_C = CT and rs10848523_A = AG or rs10743980_T = TT and rs2915124_C = CC and rs10848523_A = AG → Low Risk rs10743980_T = CC and rs2915124_C = TT and rs10848523_A = GG or rs10743980_T = CC and rs2915124_C = CT and rs10848523_A = AA or rs10743980_T = CC and rs2915124_C = CC and rs10848523_A = GG or rs10743980_T = CC and rs2915124_C = CC and rs10848523_A = AG or rs10743980_T = TC and rs2915124_C = TT and rs10848523_A = GG or rs10743980_T = TC and rs2915124_C = CT and rs10848523_A = AG or rs10743980_T = TT and rs2915124_C = TT and rs10848523_A = GG or rs10743980_T = TT and rs2915124_C = CT and rs10848523_A = AG or rs10743980_T = TT and rs2915124_C = CC and rs10848523_A = AG or	<0.001	1.755E-12	6.193	3.731 - 10.28

(Continued)

TABLE 2 Continued

c) Wnt5a					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
	rs10743980_T = TT and rs2915124_C = TT and rs10848523_A = AA or rs10743980_T = TT and rs2915124_C = CT and rs10848523_A = AA or rs10743980_T = TT and rs2915124_C = CC and rs10848523_A = GG → High Risk				
d) Wnt5b					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
2-way					
WNT5B.rs10773958_A, WNT5B.rs10491958_T	rs10773958_A = GG and rs10491958_T = CC or rs10773958_A = AG and rs10491958_T = TC or rs10773958_A = AG and rs10491958_T = TT or rs10773958_A = AA and rs10491958_T = CC → Low Risk rs10773958_A = GG and rs10491958_T = TC or rs10773958_A = GG and rs10491958_T = TT or rs10773958_A = AG and rs10491958_T = CC or rs10773958_A = AA and rs10491958_T = TC or rs10773958_A = AA and rs10491958_T = TT → High Risk	0.084	-	-	-
3-way					
KLRG2.rs9632774_A, WNT5B.rs11061856_T, WNT5B.rs4766399_G	rs9632774_A = GG and rs11061856_T = CC and rs4766399_G = GT or rs9632774_A = GG and rs11061856_T = TC and rs4766399_G = TT or rs9632774_A = GG and rs11061856_T = TC and rs4766399_G = GG or rs9632774_A = GG and rs11061856_T = TT and rs4766399_G = GT or rs9632774_A = AG and rs11061856_T = CC and rs4766399_G = TT or rs9632774_A = AG and rs11061856_T = TC and rs4766399_G = GT or rs9632774_A = AG and rs11061856_T = TC and rs4766399_G = GG or rs9632774_A = AA and rs11061856_T = CC and rs4766399_G = TT or rs9632774_A = AA and rs11061856_T = CC and rs4766399_G = GG or rs9632774_A = AA and rs11061856_T = TC and rs4766399_G = GT or rs9632774_A = AA and rs11061856_T = TC and rs4766399_G = GG or rs9632774_A = AA and rs11061856_T = TT and rs4766399_G = GT → Low Risk rs9632774_A = GG and rs11061856_T = CC and rs4766399_G = TT or rs9632774_A = GG and rs11061856_T = CC and rs4766399_G = GG or	0.011	2.894E-07	3.561	2.192 - 5.786

(Continued)

TABLE 2 Continued

d) Wnt5b					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
	rs9632774_A = GG and rs11061856_T = TC and rs4766399_G = GT or rs9632774_A = AG and rs11061856_T = CC and rs4766399_G = GT or rs9632774_A = AG and rs11061856_T = CC and rs4766399_G = GG or rs9632774_A = AG and rs11061856_T = TC and rs4766399_G = TT or rs9632774_A = AA and rs11061856_T = CC and rs4766399_G = GT or rs9632774_A = AA and rs11061856_T = TC and rs4766399_G = TT → High Risk				
e) Wnt11					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
2-way					
FUCA2.rs11155297_T, TMED7.rs10075869_G	rs11155297_T = GG and rs10075869_G = GA or rs11155297_T = TG and rs10075869_G = AA → Low Risk rs11155297_T = GG and rs10075869_G = AA or rs11155297_T = GG and rs10075869_G = GG or rs11155297_T = TG and rs10075869_G = GA or rs11155297_T = TG and rs10075869_G = GG or rs11155297_T = TT and rs10075869_G = AA or rs11155297_T = TT and rs10075869_G = GA or rs11155297_T = TT and rs10075869_G = GG → High Risk	0.004	5.879E-06	3.328	1.978 - 5.598
3-way					
C1orf54.rs10157197_A, TMED7.rs698366_A, WNT11.rs17749202_C	rs10157197_A = GG and rs698366_A = CC and rs17749202_C = TT or rs10157197_A = GG and rs698366_A = CC and rs17749202_C = CT or rs10157197_A = GG and rs698366_A = AC and rs17749202_C = CT or rs10157197_A = GG and rs698366_A = AC and rs17749202_C = CC or rs10157197_A = GG and rs698366_A = AA and rs17749202_C = TT or rs10157197_A = GG and rs698366_A = AA and rs17749202_C = CC or rs10157197_A = AG and rs698366_A = CC and rs17749202_C = CC or rs10157197_A = AG and rs698366_A = AC and rs17749202_C = TT or rs10157197_A = AG and rs698366_A = AC and rs17749202_C = CT or rs10157197_A = AG and rs698366_A = AC and rs17749202_C = CC or rs10157197_A = AG and rs698366_A = AA and rs17749202_C = TT or	0.001	5.655E-09	4.149	2.571 - 6.695

(Continued)

TABLE 2 Continued

e) Wnt11					
μTop Model SNPs	Top Model Risk Categorization	Permutation Testing P-value	Logistic Regression P-value	Odds Ratio (OR)	Logistic Regression Confidence Interval (CI)
	rs10157197_A = AA and rs698366_A = CC and rs17749202_C = CT or rs10157197_A = AA and rs698366_A = CC and rs17749202_C = CC or rs10157197_A = AA and rs698366_A = AC and rs17749202_C = TT or rs10157197_A = AA and rs698366_A = AC and rs17749202_C = CC or rs10157197_A = AA and rs698366_A = AA and rs17749202_C = CT → Low Risk rs10157197_A = GG and rs698366_A = CC and rs17749202_C = CC or rs10157197_A = GG and rs698366_A = AC and rs17749202_C = TT or rs10157197_A = GG and rs698366_A = AA and rs17749202_C = CT or rs10157197_A = AG and rs698366_A = CC and rs17749202_C = TT or rs10157197_A = AG and rs698366_A = CC and rs17749202_C = CT or rs10157197_A = AG and rs698366_A = AA and rs17749202_C = CT or rs10157197_A = AG and rs698366_A = AA and rs17749202_C = CC or rs10157197_A = AA and rs698366_A = CC and rs17749202_C = TT or rs10157197_A = AA and rs698366_A = AC and rs17749202_C = CT or rs10157197_A = AA and rs698366_A = AA and rs17749202_C = TT or rs10157197_A = AA and rs698366_A = AA and rs17749202_C = CC → High Risk				

High risk genotypes are shown in red font. μThe letter at the end is the minor allele.  
 - says that there is no results to report.

whether or not these SNPs are likely to have biological roles in the phenotype can be first assessed using existing information, for example, based on their genomic/genic locations, predictive tools/databases, and previously conducted large-scale experiments (such as eQTL analyses). In this regard, RegulomeDB (36) and GTEx (37)

database information suggested that a number of the identified variants were eQTLs that were associated with the expression levels of genes. In four cases, these genes included the genes that the SNPs were located in. For example, WLS.rs2116046 was an eQTL for *GNG12-AS1* as well as *WLS* in transverse colon (note that sequences

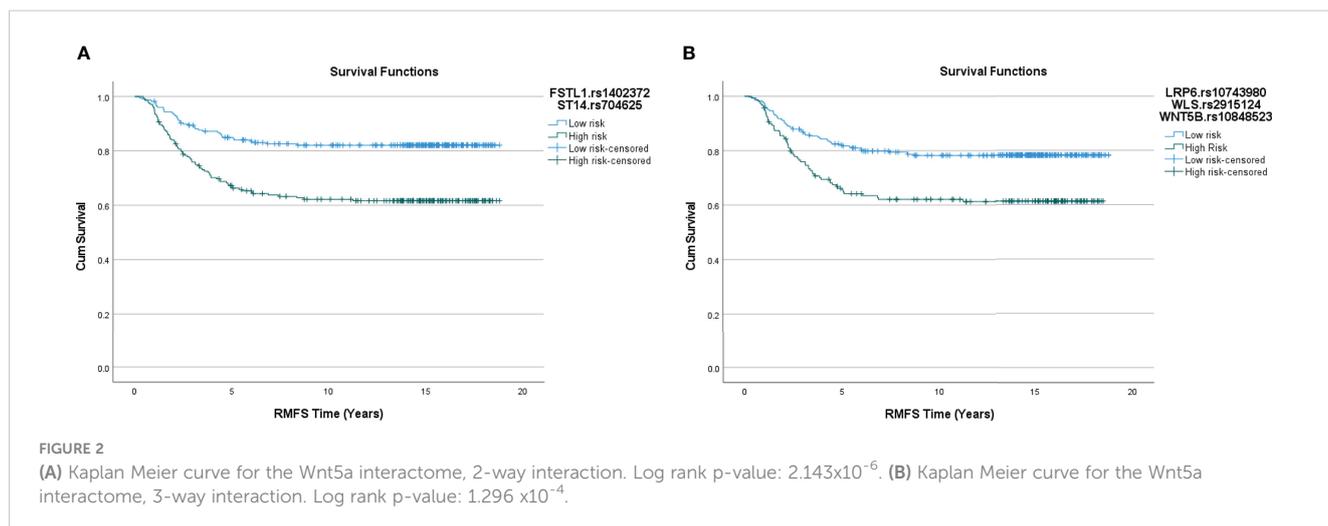


TABLE 3 eQTL information for the SNPs identified.

$\mu$ Variant	*Chr	*Location (hg19)	**RegulomeDB rank	***GTEX eQTL in colon sigmoid - target gene (chr)	***GTEX eQTL in colon transverse - target gene (chr)
<b>1-WAY</b>					
HSPA5.rs12009_C	9	127997302-127997303	<b>#1b</b>	PRPS1P2 (chr9)	PRPS1P2 (chr9)
MKRN2.rs5746255_C	3	12624320-12624321	<b>#1f</b>	MKRN2 (chr3)	MKRN2 (chr3)
DDX58.rs944582_G	9	32465769-32465770	7	GVQW1 (chr9); ACO1 (chr9)	DDX58 (chr9); ACO1 (chr9)
DDX58.rs4384073_G	9	32474689-32474690	5	ACO1 (chr9)	No
NME7.rs1080266_A	1	169176589-169176590	5	NME7 (chr1)	NME7 (chr1)
<b>2-WAY</b>					
ROR2.rs7037255_A	9	94696953-94696954	5	No	SPTLC1 (chr9)
SFRP1.rs7843510_G	8	41135768-41135769	5	No	No
GPC1.rs12695020_G	2	241403956-241403957	3a	n/a	n/a
WLS.rs2116046_C	1	68662251-68662252	7	No	****GNG12-AS1 (chr1)
FUCA2.rs11155297_T	6	143825103-143825104	5	ADAT2 (chr6)	ADAT2 (chr6)
TMED7.rs10075869_G	5	114956958-114956959	5	AC010226.4 (chr5)	AC010226.4 (chr5)
<b>3-WAY</b>					
LRP6.rs11609634_T	12	12309686-12309687	4	No	No
ROR2.rs7037255_A	9	94696953-94696954	5	No	SPTLC1 (chr9)
UBR3.rs11691281_G	2	170861133-170861134	7	No	No
HCK.rs980368_G	20	30681542-30681543	4	No	RP11-358N2.2 (chr20)
PPP6R3.rs2840367_C	11	68302100-68302101	7	No	No
SORL1.rs3862606_G	11	121330086-121330087	3a	No	No
C1orf54.rs10157197_A	1	150250635-150250636	3a	MRPS21 (chr1)	MRPS21 (chr1)
TMED7.rs698366_A	5	114951139-114951140	7	AC010226.4 (chr5); TICAM2 (chr5)	AC010226.4 (chr5); TICAM2 (chr5)
WNT11.rs17749202_C	11	75897373-75897374	2b	No	No

\*Based on the RegulomeDB database (36) \*\*The RegulomeDB (36) rank scores signifying eQTLs are shown in bold. \*\*\*Based on the GTEX database (data release v8) (37). \*\*\*\*According to the dbSNP database (40), this SNP is also located in GNG12-AS1. #According to RegulomeDB, two of these SNPs (HSPA5.rs12009\_C and MKRN2.rs5746255\_C) are eQTLs for RABEPK (Rab9 effector protein with kelch motifs), and PPARG (peroxisome proliferator activated receptor gamma) and MKRN2 (makorin ring finger protein 2), respectively, in monocytes. Therefore, in this table and manuscript, we prioritize and focus on the data retrieved from the GTEx database where the tissues are colon. n/a: variant was not found in the GTEx database.  $\mu$ The letter at the end is the minor allele.

of *WLS* and *GNG12-AS1* overlap, and rs2116046 is also located in *GNG12-AS1* (Table 3). An interesting case was identified in the 2-way interaction results where both of the SNPs identified turned out to be eQTLs: *FUCA2*.rs11155297 (eQTL for *ADAT2* in both sigmoid and transverse colon) and *TMED7*.rs10075869 (eQTL for *AC010226.4* in both sigmoid and transverse colon). Little is known about the latter gene but *ADAT2* is involved in tRNA modification (42), upregulated through translational control by *BRCA1* and a marker of *BRCA1* depletion in cell lines and human breast tumors (43), identified as over-expressed in different solid tumors (44), and linked to malignant transformation through its roles in affecting chromatin, transcriptional processes, and apoptosis (44). Also, while neither *TMED7* (transmembrane p24 trafficking protein 7) nor *FUCA2* (alpha-L-fucosidase 2) genes are known to be linked to colorectal cancer, *FUCA2* is associated with various cancers, tumor microenvironment and prognostic features (45). Although this is the first time *ADAT2* has been linked to colorectal cancer, our data and literature information make *ADAT2* (as well as *FUCA2*.rs11155297) interesting candidates for future studies in colorectal cancer progression and prognosis.

Additional genes are worth discussion. Supplementary Table S5 shows the information collected about the genes in the five Wnt interactome networks as well as the genes associated with the eQTL SNPs. A significant portion of the genes was already linked to colorectal cancer pathogenesis or progression/prognosis by previous studies. For example, expression levels, deletion, or biological functions of *ROR2*, *SFRP1*, *LRP6*, *PITX2*, *WLS*, *GPC1*, *HCK*, *HPN*, *MKRN2*, and *DDX58* were associated with disease features, patient outcomes/prognosis, or invasive and other malignant features of colorectal tumors (46–57). This literature information supports our findings and can be partly attributed to the fact that the Wnt pathway is one of the most studied pathways in colorectal cancer, increasing the chances of finding literature information on genes functioning in Wnt-related biological processes. Overall, future biological studies and/or interventions can be planned for the genes and eQTLs identified by our analyses.

This study has a number of strengths and limitations. This is one of the few large-scale studies examining such a large number of interactions in colorectal cancer. Interactions identified are novel. The patient cohort is a well annotated cohort and the genes selected have been previously shown to have abnormalities/functional roles in colorectal cancer development and or progression (24–26). The 5-step cross validation and repeating the analyses 20 times helped reduce the false-positive findings, in addition to the permutation testing. Additionally, we examined SNP interactions in protein interaction networks, increasing the biological plausibility of the identified interactions. Our cohort, however, consists of only Caucasian patients, therefore, our results may not be applicable to other ethnicities/populations. The identified variants associated with the 5-year recurrence-free survival need to be verified in other patient cohorts for generalizability of our findings. The study focused on common (MAFs  $\geq 5\%$ ) SNPs from the autosomal chromosomes, and hence, missed examining the associations of rare variables and variables from sex-chromosomes. As we reported earlier, GMDR 0.9 has certain limitations, so it may have missed interactions (23). However, use

of permutation testing, repeating the MDR procedure and choosing the most frequently identified MDR model for each examined interaction, and multivariable regression modeling also have limited the false-positive findings.

In conclusion, we present novel 1 to 3 way SNP interactions that predict the 5-year RMFS status in colorectal cancer. These interactions are excellent candidates for further verification in other patient cohorts. We also identified a number of genes that are biologically linked to colorectal cancer: they form an exciting set for future studies or interventions in colorectal cancer. Our results also indicate that MDR and other data reduction methods should be utilized more widely for comprehensive investigations of statistical interactions in patient prognosis. Finally, our findings also re-emphasize and strengthen the importance of Wnt protein pathways in colorectal cancer.

## Data availability statement

Data that support the findings of this study are available from the Newfoundland Colorectal Cancer Registry/Memorial University of Newfoundland. However, restrictions apply to the availability of this data, and so data are not publicly available. The data used in this study cannot be made publicly available as patients were not consented to make their data publicly available or accessible. Clinical and genetic data are available from the Newfoundland Colorectal Cancer Registry (NFCCR) upon reasonable request for researchers who meet the criteria for access to confidential data. Permission to obtain the data can be requested from Newfoundland Colorectal Cancer Registry (PP; [pparfrey@mun.ca](mailto:pparfrey@mun.ca)) and Research, Grant, and Contract Services ([rgcs@mun.ca](mailto:rgcs@mun.ca)) at Memorial University of Newfoundland, St. John's, NL, Canada, and the ethics approval shall be obtained from the Health Research Ethics Board (HREB), Ethics Office, Health Research Ethics Authority, Suite 200, 95 Bonaventure Avenue, St. John's, NL, A1B 2X5, Canada. The GMDR 0.9 code can be requested from the developers, Drs. Xiang-Yang Lou, Jun Zhu, or Ming D. Li. Requests to access these datasets should be directed to PP; [pparfrey@mun.ca](mailto:pparfrey@mun.ca)) and Research, Grant, and Contract Services ([rgcs@mun.ca](mailto:rgcs@mun.ca)).

## Ethics statement

The studies involving human participants were reviewed and approved by Health Research Ethics Authority of Newfoundland and Labrador. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

AC: performed all MDR and statistical analyses; managed the data; helped interpret the results and draft the manuscript: YY: performed the bioinformatics analyses; helped interpret the results

and draft the manuscript; MC: helped collect the outcome data; PP: led the NFCCR and helped collect the clinical and genetic patient data; YEY: helped with statistical approach and permutation testing. SS: conceived the idea; supervised the assistants; helped interpret the results; drafted; finalized and submitted the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This study was funded by the Medical Research Foundation (MRF) of the Faculty of Medicine, Memorial University (PI: SS). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Acknowledgments

We are indebted to the patients recruited to the NFCCR. We thank the NFCCR staff, investigators, and NL Tumor Registry staff for helping collect the data used for this study. We also thank Dr. Guobo Chen for valuable correspondence related to their GMDR 0.9 software. SS is a senior scientist of the Beatrice Hunter Cancer Research Institute (BHCRI).

## Conflict of interest

SS is an associate editor with *Frontiers in Oncology/Genetics*.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- 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. doi: 10.3322/caac.21660
- Boyle P, Langman JS. ABC Of colorectal cancer: Epidemiology. *BMJ* (2000) 321:805–8. doi: 10.1136/bmj.321.7264.805
- Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* (2010) 138:2044–58. doi: 10.1053/j.gastro.2010.01.054
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* (2015) 136:E359–86. doi: 10.1002/ijc.29210
- Hermelink R, Leitzmann MF, Markozannes G, Tsilidis K, Pukrop T, Berger F, et al. Sedentary behavior and cancer—an umbrella review and meta-analysis. *Eur J Epidemiol* (2022) 37:447–60. doi: 10.1007/s10654-022-00873-6
- Compton CC. Colorectal carcinoma: diagnostic, prognostic, and molecular features. *Mod Pathol* (2003) 16:376–88. doi: 10.1097/01.MP.0000062859.46942.93
- Zlobec I, Lugli A. Prognostic and predictive factors in colorectal cancer. *J Clin Pathol* (2008) 61:561–9. doi: 10.1136/jcp.2007.054858
- Savas S, Liu G. Genetic variations as cancer prognostic markers: review and update. *Hum Mutat* (2009) 30:1369–77. doi: 10.1002/humu.21078
- 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature* (2010) 467:1061–73. doi: 10.1038/nature09534
- Theodoratou E, Din FV, Farrington SM, Cetnarskyj R, Barnettson RA, Porteous ME, et al. Association between common mtDNA variants and all-cause or colorectal cancer mortality. *Carcinogenesis* (2010) 31:296–301. doi: 10.1093/carcin/bgp237
- Negandhi A, Hyde A, Dicks E, Younghusband B, Parfrey P, Green R, et al. MTHFR Glu429Ala and ERCC5 His46His polymorphisms are associated with

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1122229/full#supplementary-material>

### SUPPLEMENTARY TABLE 1

Proteins in the WNT interaction networks.

### SUPPLEMENTARY TABLE 2

SNPs included in this study.

### SUPPLEMENTARY TABLE 3

Single SNP interaction models identified and results of the logistic regression analyses.

### SUPPLEMENTARY TABLE 4

SNP functional annotations (based on the dbSNP database)

### SUPPLEMENTARY TABLE 5

Information on the genes.

### SUPPLEMENTARY FIGURE 1

Kaplan Meier curves for select 2-way and 3-way interactions.

prognosis in colorectal cancer patients: Analysis of two independent cohorts from Newfoundland. *PLoS One* (2013) 8:e61469. doi: 10.1371/journal.pone.0061469

12. Phipps AI, Passarelli MN, Chan AT, Harrison TA, Jeon J, Hutter CM, et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *Carcinogenesis* (2016) 37:87–95. doi: 10.1093/carcin/bgv161

13. Yu Y, Werdyani S, Carey M, Parfrey P, Yilmaz YE, Savas S. A comprehensive analysis of SNPs and CNVs identifies novel markers associated with disease outcomes in colorectal cancer. *Mol Oncol* (2021) 15:3329–47. doi: 10.1002/1878-0261.13067

14. Labadie JD, Savas S, Harrison TA, Banbury B, Huang Y, Buchanan DD, et al. Genome-wide association study identifies tumor anatomical site-specific risk variants for colorectal cancer survival. *Sci Rep* (2022) 12:127. doi: 10.1038/s41598-021-03945-x

15. Quintanilha JCF, Wang J, Sibley AB, Xu W, Espin-Garcia O, Jiang C, et al. Genome-wide association studies of survival in 1520 cancer patients treated with bevacizumab-containing regimens. *Int J Cancer* (2022) 150:279–89. doi: 10.1002/ijc.33810

16. Lee S, Kwon M-S, Oh JM, Park T. Gene-gene interaction analysis for the survival phenotype based on the cox model. *Bioinformatics* (2012) 28:i582–8. doi: 10.1093/bioinformatics/bts415

17. Moutsier AA, Ritchie MD. Multifactor dimensionality reduction: An analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Hum Genomics* (2006) 2:318. doi: 10.1186/1479-7364-2-5-318

18. Lou X-Y, Chen G-B, Yan L, Ma JZ, Zhu J, Elston RC, et al. A generalized combinatorial approach for detecting gene-by-Gene and gene-by-Environment interactions with application to nicotine dependence. *Am J Hum Genet* (2007) 80:1125–37. doi: 10.1086/518312

19. Pander J, Wessels JAM, Gelderblom H, van der Straaten T, Punt CJA, Guchelaar H-J. Pharmacogenetic interaction analysis for the efficacy of systemic treatment in

- metastatic colorectal cancer. *Ann Oncol* (2011) 22:1147–53. doi: 10.1093/annonc/mdq572
20. Afzal S, Gusella M, Jensen SA, Vainer B, Vogel U, Andersen JT, et al. The association of polymorphisms in 5-fluorouracil metabolism genes with outcome in adjuvant treatment of colorectal cancer. *Pharmacogenomics* (2011) 12:1257–67. doi: 10.2217/pgs.11.83
21. Afzal S, Gusella M, Vainer B, Vogel UB, Andersen JT, Broedbaek K, et al. Combinations of polymorphisms in genes involved in the 5-fluorouracil metabolism pathway are associated with gastrointestinal toxicity in chemotherapy-treated colorectal cancer patients. *Clin Cancer Res* (2011) 17:3822–9. doi: 10.1158/1078-0432.CCR-11-0304
22. Sarac SB, Rasmussen CH, Afzal S, Thirstrup S, Jensen SA, Colding-Jørgensen M, et al. Data-driven assessment of the association of polymorphisms in 5-fluorouracil metabolism genes with outcome in adjuvant treatment of colorectal cancer. *Bas. Clin Pharmacol Toxicol* (2012) 111:189–97. doi: 10.1111/j.1742-7843.2012.00885.x
23. Curtis A, Yu Y, Carey M, Parfrey P, Yilmaz YE, Savas S. Examining SNP-SNP interactions and risk of clinical outcomes in colorectal cancer using multifactor dimensionality reduction based methods. *Front Genet* (2022) 13:902217. doi: 10.3389/fgene.2022.902217
24. de Lau W, Barker N, Clevers H. WNT signaling in the normal intestine and colorectal cancer. *Front Biosci* (2007) 12:471–91. doi: 10.2741/2076
25. Nie X, Liu H, Liu L, Wang Y-D, Chen W-D. Emerging roles of wnt ligands in human colorectal cancer. *Front Oncol* (2020) 10:1341. doi: 10.3389/fonc.2020.01341
26. Disoma C, Zhou Y, Li S, Peng J, Xia Z. Wnt/ $\beta$ -catenin signaling in colorectal cancer: Is therapeutic targeting even possible? *Biochimie* (2022) 195:39–53. doi: 10.1016/j.biochi.2022.01.009
27. Woods MO, Younghusband HB, Parfrey PS, Gallinger S, McLaughlin J, Dicks E, et al. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. *Gut* (2010) 59:1369–77. doi: 10.1136/gut.2010.208462
28. Yu Y, Carey M, Pollett W, Green J, Dicks E, Parfrey P, et al. The long-term survival characteristics of a cohort of colorectal cancer patients and baseline variables associated with survival outcomes with or without time-varying effects. *BMC Med* (2019) 17:150. doi: 10.1186/s12916-019-1379-5
29. Xu W, Xu J, Shestopaloff K, Dicks E, Green J, Parfrey P, et al. A genome wide association study on Newfoundland colorectal cancer patients' survival outcomes. *biomark Res* (2015) 3:6. doi: 10.1186/s40364-015-0031-6
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* (2007) 81:559–75. doi: 10.1086/519795
31. Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, Willems A, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* (2021) 30:187–200. doi: 10.1002/pro.3978
32. Chen G-B, Xu Y, Xu H-M, Li MD, Zhu J, Lou X-Y. Practical and theoretical considerations in study design for detecting gene-gene interactions using MDR and GMDR approaches. *PLoS One* (2011) 6:e16981. doi: 10.1371/journal.pone.0016981
33. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* (2001) 69:138–47. doi: 10.1086/321276
34. Core Team, R. R. *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (2017).
35. IBM SPSS Statistics for Windows. (2017).
36. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* (2012) 22:1790–7. doi: 10.1101/gr.137323.112
37. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. *Nat Genet* (2013) 45:580–5. doi: 10.1038/ng.2653
38. Savas S, Younghusband HB. dbCPCO: a database of genetic markers tested for their predictive and prognostic value in colorectal cancer. *Hum Mutat* (2010) 31:901–7. doi: 10.1002/humu.21285
39. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res* (2011) 39:D52–7. doi: 10.1093/nar/gkq1237
40. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* (2001) 29:308–11. doi: 10.1093/nar/29.1.308
41. Winder T, Bohanes P, Zhang W, Yang D, Power DG, Ning Y, et al. GRP78 promoter polymorphism rs391957 as potential predictor for clinical outcome in gastric and colorectal cancer patients. *Ann Oncol* (2011) 22:2431–9. doi: 10.1093/annonc/mdq771
42. Schaub M, Keller W. RNA Editing by adenosine deaminases generates RNA and protein diversity. *Biochimie* (2002) 84:791–803. doi: 10.1016/s0300-9084(02)01446-3
43. Berthel E, Vincent A, Eberst L, Torres AG, Dacheux E, Rey C, et al. Uncovering the translational regulatory activity of the tumor suppressor BRCA1. *Cells* (2020) 9:941. doi: 10.3390/cells9040941ac
44. Caron C, Lestrat C, Marsal S, Escoffier E, Curtet S, Virolle V, et al. Functional characterization of ATAD2 as a new cancer/testis factor and a predictor of poor prognosis in breast and lung cancers. *Oncogene* (2010) 29:5171–81. doi: 10.1038/nc.2010.259
45. Liu Q, Dong H-T, Zhao T, Yao F, Xu Y, Chen B, et al. Cancer-associated adipocytes release FUCA2 to promote aggressiveness in TNBC. *Endocr. Relat Cancer* (2022) 29:139–49. doi: 10.1530/ERC-21-0243
46. Hirose H, Ishii H, Mimori K, Tanaka F, Takemasa I, Mizushima T, et al. The significance of PITX2 overexpression in human colorectal cancer. *Ann Surg Oncol* (2011) 18:3005–12. doi: 10.1245/s10434-011-1653-z
47. Mei H, Lian S, Zhang S, Wang W, Mao Q, Wang H. High expression of ROR2 in cancer cell correlates with unfavorable prognosis in colorectal cancer. *Biochem Biophys Res Commun* (2014) 453:703–9. doi: 10.1016/j.bbrc.2014.09.141
48. Xu H, Jiang W, Zhu F, Zhu C, Wei J, Wang J. Expression of wntless in colorectal carcinomas is associated with invasion, metastasis, and poor survival. *APMIS* (2016) 124:522–8. doi: 10.1111/apm.12534
49. Yao Q, An Y, Hou W, Cao Y-N, Yao M-F, Ma N-N, et al. LRP6 promotes invasion and metastasis of colorectal cancer through cytoskeleton dynamics. *Oncotarget* (2017) 8:109632–45. doi: 10.18632/oncotarget.22759
50. Gu C, Wang X, Long T, Wang X, Zhong Y, Ma Y, et al. FSTL1 interacts with VIM and promotes colorectal cancer metastasis via activating the focal adhesion signalling pathway. *Cell Death Dis* (2018) 9:1–14. doi: 10.1038/s41419-018-0695-6
51. Roseweir AK, Powell AGMT, Horstman SL, Inthagard J, Park JH, McMillan DC, et al. Src family kinases, HCK and FGR, associate with local inflammation and tumour progression in colorectal cancer. *Cell Signal* (2019) 56:15–22. doi: 10.1016/j.cellsig.2019.01.007
52. Chang Z, Liu X, Zhao W, Xu Y. Identification and characterization of the copy number dosage-sensitive genes in colorectal cancer. *Mol Ther Methods Clin Dev* (2020) 18:501–10. doi: 10.1016/j.omtm.2020.06.020
53. Busuioic C, Ciocan-Cartita CA, Braicu C, Zanoaga O, Raduly L, Trif M, et al. Epithelial-mesenchymal transition gene signature related to prognostic in colon adenocarcinoma. *J Pers. Med* (2021) 11:476. doi: 10.3390/jpm11060476
54. He Y, Gong P, Wang S, Xu Q, Chen J. The significance of homeodomain transcription factor 2 in colon cancer cells. *Biomed Eng. Online* (2021) 20:81. doi: 10.1186/s12938-021-00912-5
55. Deng Y, Fu H, Han X, Li Y, Zhao W, Zhao X, et al. Activation of DDX58/RIG-I suppresses the growth of tumor cells by inhibiting STAT3/CSE signaling in colon cancer. *Int J Oncol* (2022) 61:1–13. doi: 10.3892/ijo.2022.5410
56. Lu F, Chen S, Shi W, Su X, Wu H, Liu M. GPC1 promotes the growth and migration of colorectal cancer cells through regulating the TGF- $\beta$ 1/SMAD2 signaling pathway. *PLoS One* (2022) 17:e0269094. doi: 10.1371/journal.pone.0269094
57. Zaragoza-Huesca D, Nieto-Olivares A, García-Molina F, Ricote G, Montenegro S, Sánchez-Cánovas M, et al. Implication of hepsin from primary tumor in the prognosis of colorectal cancer patients. *Cancers* (2022) 14:3106. doi: 10.3390/cancers14133106