HHS Public Access

Author manuscript

*Int J Cancer*. Author manuscript; available in PMC 2021 January 15.

Published in final edited form as:

*Int J Cancer*. 2020 January 15; 146(2): 363–372. doi:10.1002/ijc.32516.

# DNA repair and cancer in colon and rectum: novel players in genetic susceptibility

**Barbara Pardini**[1,2,*], **Alda Corrado**[3,*], **Elisa Paolicchi**[3], **Giovanni Cugliari**[1,2], **Sonja I. Berndt**[4], **Stephane Bezieau**[5], **Stephanie A. Bien**[6,7], **Hermann Brenner**[8,9,10], **Bette J. Caan**[11], **Peter T. Campbell**[12], **Graham Casey**[13], **Andrew T. Chan**[14], **Jenny Chang-Claude**[15], **Michelle Cotterchio**[16], **Manish Gala**[14], **Steven J. Gallinger**[17], **Robert W. Haile**[18], **Tabitha A. Harrison**[6], **Richard B. Hayes**[19], **Michael Hoffmeister**[8], **John L. Hopper**[20], **Li Hsu**[6], **Jeroen Huyghe**[6], **Mark A. Jenkins**[20], **Loic Le Marchand**[21], **Yi Lin**[6], **Noralane M. Lindor**[22], **Hongmei Nan**[23], **Polly A. Newcomb**[6], **Shuji Ogino**[24,25,26,27], **John D. Potter**[6], **Robert E. Schoen**[28], **Martha L. Slattery**[29], **Emily White**[6], **Ludmila Vodickova**[30,31,32], **Veronika Vymetalkova**[30,31,32], **Pavel Vodicka**[30,31,32], **Federica Gemignani**[3], **Ulrike Peters**[6,*], **Alessio Naccarati**[1,32,*], **Stefano Landi**[3,*]

[1]Italian Institute for Genomic Medicine (IIGM), Turin, Italy

[2]Department of Medical Sciences, University of Turin, Turin, Italy

[3]Department of Biology, University of Pisa, Pisa, Italy

[4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD. USA

[5]Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, France

[6]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[7]School of Public Health, University of Washington, Seattle, WA, USA

[8]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ) Heidelberg, Germany

Corresponding authors Alda Corrado, Department of Biology, University of Pisa, Via Derna, Pisa, Italy, corradoalda@gmail.com, Barbara Pardini, Italian Institute for Genomic Medicine (IIGM), Via Nizza 52, 10126, Turin, Italy, barbara.pardini@iigm.it.

*These authors contributed equally to the manuscript

Authors' contributions

Barbara Pardini (BP), Alda Corrado (AC), Elisa Paolicchi (EP), Giovanni Cugliari (GCu), Sonja I. Berndt (SIB), Stephane Bezieau (SB), Stephanie A. Bien (SAB), Hermann Brenner (HB), Bette J. Caan (BJC), Peter T. Campbell (PTC), Graham Casey (GC), Andrew T. Chan (ATC), Jenny Chang-Claude (JCC), Michelle Cotterchio (MC), Manish Gala (MG), Steven J. Gallinger (SJG), Robert W. Haile (RWH), Tabitha A. Harrison (TAH) Richard B. Hayes (RBH), Michael Hoffmeister (MH), John L. Hopper (JLH), Li Hsu (LH), Jeroen Huyghe (JH), Mark A. Jenkins (MAJ), Loic Le Marchand (LLM),Yi Lin (YL), Noralane M. Lindor (NML), Hongmei Nan (HN), Polly A. Newcomb (PAN), Shuji Ogino (SO), John D. Potter (JDP), Robert E. Schoen (RES), Martha L. Slattery (MLS), Emily White (EW), Ludmila Vodickova (LV), Veronika Vymetalkova (VV), Pavel Vodicka (PV), Federica Gemignani (FG), Ulrike Peters (UP), Alessio Naccarati (AN), Stefano Landi (SL).

Conceptualization of the study: BP, AC, AN, UP and SL

Sample collection, experiments, data collection and harmonization, statistical analyses: BP, AC, EP, GCu, SIB, SB, SAB, HB, BJC, PTC, GC, ATC, JCC, MC, MG, SJG, RWH, TAH, RBH, MH, JLH, LH, JH, MAJ, LLM, YL, NML, HN, PAN, SO, JDP, RES, MLS, EW, LV, VV, PV, FG, UP, AN, SL

Writing and Original Draft Preparation: BP, AC, GCu, SL and AN

Review & Editing: BP, AC, EP, GCu, SIB, SB, SAB, HB, BJC, PTC, GC, ATC, JCC, MC, MG, SJG, RWH, TAH, RBH, MH, JLH, LH, JH, MAJ, LLM, YL, NML, HN, PAN, SO, JDP, RES, MLS, EW, LV, VV, PV, FG, UP, AN, SL

The authors declare no conflict of interests

[9]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

[10]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

[11]Kaiser Permanente Medical Care Program of Northern California, Oakland, CA, USA

[12]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA

[13]Public Health Sciences, University of Virginia, VA, USA

[14]Division of Gastroenterology, Massachusetts General Hospital, Boston, MA, USA

[15]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ) Heidelberg, Germany

[16]Prevention and Cancer Control, Cancer Care Ontario, Toronto, ON, Canada

[17]Department of Surgery, Mount Sinai Hospital, Toronto, ON, Canada

[18]Stanford University School of Medicine, Stanford, CA, USA

[19]Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, NY, USA

[20]Melborne School of Population Health, The University of Melborne, Melborne, Australia

[21]Epidemiology Program, Research Cancer Center of Hawaii, University of Hawaii, Honolulu, HI, USA

[22]Department of Health Sciences Research, Mayo Clinic Arizona, Scottsdale, AZ, USA

[23]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA

[24]Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women’s Hospital, and Harvard Medical School;

[25]Department of Oncologic Pathology, Dana-Farber Cancer Institute;

[26]Department of Epidemiology, Harvard T.H. Chan School of Public Health; all in, Boston, MA, USA;

[27]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[28]Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

[29]Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT, USA

[30]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic

[31]Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czech Republic

[32]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, The Czech Academy of Sciences, Prague, Czech Republic

## Abstract

Inter-individual differences in DNA repair systems may play a role in modulating the individual risk of developing colorectal cancer.

To better ascertain the role of DNA repair gene polymorphisms on colon and rectal cancer risk individually, we evaluated 15,419 single nucleotide polymorphisms (SNPs) within 185 DNA repair genes using GWAS data from the Colon Cancer Family Registry (CCFR) and the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), which included 8,178 colon cancer, 2,936 rectum cancer cases and 14,659 controls.

Rs1800734 (in *MLH1 gene*) was associated with colon cancer risk (p-value=$3.5\times10^{-6}$) and rs2189517 (in *RAD51B*) with rectal cancer risk (p-value=$5.7\times10^{-6}$). The results had statistical significance close to the Bonferroni corrected p-value of $5.8\times10^{-6}$. Ninety-four SNPs were significantly associated with colorectal cancer risk after Binomial Sequential Goodness of Fit (BSGoF) procedure and confirmed the relevance of DNA mismatch repair (MMR) and homologous recombination pathways for colon and rectum cancer, respectively.

Defects in MMR genes are known to be crucial for familial form of colorectal cancer but our findings suggest that specific genetic variations in *MLH1* are important also in the individual predisposition to sporadic colon cancer. Other SNPs associated with the risk of colon cancer (e.g. rs16906252 in *MGMT*) were found to affect mRNA expression levels in colon transverse and therefore working as possible cis-eQTL suggesting possible mechanisms of carcinogenesis.

### Keywords

Colon cancer; rectal cancer; DNA repair; single nucleotide polymorphisms; cancer susceptibility; genome-wide association studies

## Introduction

Cancer is the consequence of the complex interactions between genetic susceptibility and environmental factors. Among the genes playing a role in cancer susceptibility, DNA repair genes are important candidates since cancer is associated with inherited deficiencies of DNA repair [1]. Defects in DNA repair cause genetic instability leading to increased rates of somatic mutations, providing the biological bases of this phenomenon [2]. Concerning the gastro-intestinal tract, the Lynch syndrome, which is most commonly clinically manifested as hereditary nonpolyposis colorectal cancer (HNPCC), is one of the most characterized inherited forms bound to defects in the DNA mismatch repair (MMR) pathway and it accounts for about 1–5% of all colorectal cancer cases [3]. According to a multistep model of carcinogenesis [4], unrepaired mismatched bases (e.g. arising during DNA replication) cause a progressive accumulation of somatic mutations, predisposing replicating tissues with high turnover (such as the colon epithelium) to the malignant transformation [5]. The role of surveillance operated by MMR seems pivotal for colonocytes, as deficiencies within this

pathway are observed also at somatic level in 7–10% of the sporadic forms conferring the so-called “microsatellite instability” (MSI) phenotype [3, 6].

On the basis of the observations in HNPCC families, it has been hypothesized that moderate inter-individual differences in the activity of DNA repair systems could also play a role in modulating the individual risk to develop sporadic form of colorectal cancer in the general population [7–9]. Thus, various hypothesis-driven case-control studies have been carried out to evaluate the association between the risk of sporadic colorectal cancer and polymorphisms within candidate genes such as *OGG1*, *APEX*, *POLB*, *XRCC1*, and *MUTYH* (base excision repair, BER), *ERCC1*, *ERCC2*, *XPC*, and *ERCC5* (nucleotide excision repair, NER), *XRCC2* and *XRCC3* (double-strand breaks repair, DSB), and Poly(ADP-ribose) polymerase (*PARP*) [9, 10]. Positive associations were described for single-nucleotide polymorphisms (SNPs) within *APEX*, *ERCC1*, *MUTYH*, *OGG1*, *XPC*, *XPG*, *XRCC1, and XRCC3* genes [11–16], but some results were either discordant or not replicated [9, 11, 16–19] likely as the consequence of a limited statistical power. Genome-wide association studies (GWASs) could not confirm most of the positive associations within the DNA repair genes previously described [20–23]. Similarly, GWASs carried out on other types of cancer detected only few DNA repair SNPs (see the GWAS catalog https://www.ebi.ac.uk/gwas/home) such as in breast cancer the rs999737 near *RAD51L1*, likely affecting the DSB DNA repair [24]. Most probably, the low number of disease-associated DNA repair SNPs in GWAS could be due to the very small effect of each SNP or to moderately penetrant, rare, and population-specific alleles having various extents of linkage disequilibrium (LD) with the polymorphisms typically analyzed using commercial microarrays. Moreover, the effect of each SNP could be diluted in typical GWAS of overall colorectal cancer cases including tumors with different tumor molecular pathologies, as each risk allele is conceived to differentially influence specific carcinogenic mechanisms [25]. However, a previous study with adequate statistical power showed that the set of DNA repair SNPs, as a whole, could be associated with colorectal cancer risk [20]. Meta-analyses suggested also positive associations for rs1052133 and rs861539, respectively within *hOGG1* [26] and *XRCC3* [27] genes. These associations were observed only in specific ethnic groups, indirectly confirming the hypothesis of the moderately penetrant population-specific alleles [28]. In summary, further investigations are needed, in particular in large populations. In order to overcome the limitations imposed by the statistical power and in the attempt to draw more robust conclusions, we evaluated available SNPs within the full set of DNA repair genes in a large number of cases and controls combining data from two consortia: the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), and the Colon Cancer Family Registry (CCFR) [29]. We hypothesize that specific DNA repair pathways could be relevant in better describe risk association for colorectal cancer with particular care for cancer site subtypes. The large sample size allowed in fact to better investigating the role of DNA repair genes by stratifying for colon and rectal cancer separately.

## Material and methods

### Study population and genotyping

We included 14 studies from the CCFR and GECCO consortia as described previously and in the Supplementary Material (Text S1) **and** Table 1 [29–31]. All colorectal cancer cases were defined as colon or rectal adenocarcinoma and confirmed by medical records, pathologic reports, cancer registries, or death certificates. All analyses were restricted to individuals of European ancestry.

Methods of array-based genotyping, quality assurance/quality control and imputation, average sample and SNP call rates, and concordance rates for blinded duplicates have been previously published [32]. In brief, for quality insurance SNPs were excluded based on call rate (<98%), lack of Hardy-Weinberg Equilibrium (HWE) among controls (setting a threshold of $p<10^{-4}$), and low minor allele frequency (MAF) <0.05. We imputed the autosomal SNPs of all studies to the Northern Europeans from Utah (CEU population) in HapMap II. List of SNPs was restricted based on per-study minor allele count > 5 and imputation accuracy ($r^2$ >0.3). After imputation and quality-control exclusion, approximately 2.7M SNPs were available as complete genotype dataset. Imputations were done using the Haplotype Reference Consortium (HRC) r1.0 reference panel and Michigan Imputation Server, with phasing option set to ShapeIT v2.r790 [33–35].

### Selection of candidate genes and SNPs

To evaluate the association between polymorphic DNA repair genes and risk of colon and rectal cancer, we initially selected genes involved in many aspects of DNA repair processing as listed in: https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html. A total of 185 genes (Supplementary Table 1; Figure 1) were retrieved and for each of them all known SNPs reported for the gene region (including 5' and 3' near regions, as classified and reported in dbSNP) were evaluated. As one example, see *MLH1* at URL: https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?showRare=on&chooseRs=all&go=Go&locusId=4292 The complete list of 15,419 SNPs is reported in Supplementary Table 1.

### In silico analyses

In order to evaluate possible biological effects of specific SNPs, computational predictions were performed with the use of bioinformatics tools (Figure 1). First, we analyzed the presence of blocks of LD ($r^2$>0.8) by using "LD TAG SNP selection" available at http://archive.broadinstitute.org/mpg/snap/ldsearch.php and Haploreg V4.1 (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php). The latter is based on the ENCODE database and provides information for the analysis of the non-coding genome. Candidate regulatory SNPs were displayed together with their associated chromatin status and with the annotation of their protein binding sites (from the Roadmap Epigenomics and ENCODE projects). The information was also completed with the estimates of sequence conservation between mammals and the effects on the regulatory motifs and gene expression (from expression quantitative trait loci, eQTL, studies). Finally, for each SNP we examined gene expression levels as quantitative trait loci (cis-eQTLs), as available in GTEx Portal

(https://www.gtexportal.org) for intestinal tissues (i.e. transverse, n=246 and sigmoid, n=203).

## Statistical analysis

The association between SNPs and colon or rectal cancer risk was estimated using multiple logistic regression model with log-additive genetic effect. The model was adjusted for sex, age, genotype phase, batch effect, and principal components (PCs) for ancestry. The adjustment for multiple testing was initially approached by employing the Bonferroni's correction considering that, because of the presence of LD, about 4,300 independent haplotype-tagging SNPs (using a LD threshold with an $r^2 \geq 0.8$) could recapitulate the whole genetic variability contained in the full set of SNPs. The novel threshold of statistical significance was, then, $5.8 \times 10^{-6}$ (considering 2 sets of statistical tests, one for colon and one for rectum)

Moreover, as an alternative hypothesis-generating approach, we also tested the Binomial Sequential Goodness of Fit (BSGoF) method for multiple test adjustment. BSGoF (described in [36]) provides a good balance between false discovery rate (FDR) and power, particularly when the number of tests is large and the effect level is weak to moderate. We applied the BSGoF function to the total number of SNPs included in the study (n=15,419) to the p-value for SNP effect data (alpha=0.05, gamma=0.05).

## Ethics statement

All participants provided informed consent and studies were approved by their respective Institutional Review Boards. The overall project was reviewed and approved by the Fred Hutchinson Cancer Research Center Institutional Review Board (approval number: 1177). Each study was approved by the local IRB [University of Hawaii Human Studies Program (Colo23, Hawaii CCFR, and MEC); University of Utah Institutional Review Board (DALS); Partners Human Research Committee (NHS and PHS); Harvard School of Public Health Institutional Review Board (HPFS); Fred Hutchinson Cancer Research Center Institutional Review Board (PMH-CCFR, Seattle CCFR, VITAL, overall study); CSMC Institutional Review Boards (Cedars-Sinai CCFR); Cleveland Clinic Institutional Review Board (Cleveland Clinic CCFR); Mayo Clinic Institutional Review Board (Mayo Clinic CCFR); Mount Sinai Hospital Research Ethics Board (Ontario CCFR (OFFCR)); University of Melbourne Health Sciences Human Ethics Sub-Committee (Australasia CCFR); Ethics Committee of the Medical Faculty of the University of Heidelberg (DKFZ); NCI Special Studies Institutional Review Board (PLCO)]. For each participating study, participants or the next of kin in the case of deceased volunteers, provided either written informed consent to participate (the following CCFR sites: Australasia, Cedars-Sinai, Cleveland Clinic, Hawaii, Mayo Clinic and Ontario CCFRs), Colo23, DACHS, DALS, MEC, PHS, PLCO, VITAL, WHI) or they provided implied written consent by the return of the mailed questionnaires (NHS, HPFS) or the completion of telephone questionnaires (Seattle CCFR, PMH-CCFR). Additional consent to review medical records was obtained through signed written consent.

### Data availability

All custom Infinium OncoArray-500K array and Illumina HumanOmniExpressExome-8v1–2 array data used in the study have been deposited at dbGaP under accession number phs001415.v1.p1 and phs001315.v1.p1, respectively. Genotype data for the studies have been deposited at dbGaP under accession number phs001078.v1.p1.

## Results

In this work, we included 14 studies from CCFR and GECCO consortia as described in the Supplementary Material (Text S1) **and** Table 1 and elsewhere [29–31]. Overall, 15,419 SNPs within the 185 DNA repair genes were tested for different genotype distributions between 14,659 controls, 8,178 colon and 2,936 rectum cancer cases. The complete set of results is reported in Supplementary Table 2, whereas extracts concerning the htSNPs with the lowest p-values of association are showed grouped by gene in Tables 2 **and** 3 for colon and rectum, respectively.

The SNP rs1800734 in *MLH1* was significantly associated with the risk of colon cancer after Bonferroni's adjustment (OR=1.13, 95%CI= 1.07–1.18, p=3.5X$10^{-6}$; Table 2). Other two htSNPs within *MLH1*, i.e. rs6784088 (OR=0.94, 95%CI= 0.90–0.98, p= 3.3×$10^{-3}$) and rs9855475 (OR=0.94, 95%CI= 0.90–0.98, p=3.4X$10^{-3}$), were associated with the risk of colon cancer when BSGoF was applied (Table 2). These two latter were mildly in LD each other ($r^2$=0.71), whereas the strongest signal rs1800734 had a weak LD with them ($r^2$ = 0.33 and = 0.30, respectively). This SNP showed also an association with the risk of colorectal cancer, although at a lesser extent (OR=1.09; 95%CI= 1.04–1.14; p=5.6×$10^{-5}$).

Concerning rectal cancer, the strongest signal was found for rs2189517 within *RAD51B* (OR=1.15, 95% CI=1.08–1.22, p-value=5.7×$10^{-6}$), a gene involved in Homologous recombination repair (HR), statistically significant also following the Bonferroni's correction (Table 3). Interestingly, other 14 htSNPs were found associated at a lesser extent with the risk of rectal cancer, being statistically significant only when BSGoF was applied. The list of these SNPs includes rs12587232, rs187645011, rs7350713, rs6573841, rs111611396, rs1989974, rs117544253, rs77726787, rs8016488, rs11628293, rs113020754, rs80085210, rs113300322, and rs74933543. The significance levels ranged from 4.37×$10^{-5}$ to 1.6X$10^{-3}$ with the highest risk (OR=0.71 corresponding to 1.41 for the common versus the rare allele) for rs187645011. Rs2189517 was not in LD with the others ($r^2$<0.3) with the exception of rs12587232, having $r^2$ of 0.77. The 14 htSNPs were not in LD each other as well (max $r^2$<0.6). Rs2189517 was associated also with colorectal cancer risk (OR= 1.05, 95%CI= 1.02–1.09, p=1.7×$10^{-3}$) although statistically significant only following BSGoF correction.

When more exploratory and hypothesis-generating analyses were performed by considering statistically significant SNPs following BSGoF adjustment, several genes had multiple htSNPs associated with the risk of colon carcinoma, such as *ATM* (rs11212592, rs61915066), *FANCA* (rs2238526, rs3743860), *FANCE* (rs6907678, rs10947550), and *LIG1* (rs1971775, rs73054038). Because htSNPs are mostly independent each other, the presence of multiple signals provides a more robust indication for the role of the gene in the

susceptibility to the disease. Other genes, such as *EXO1*, *FEN1*, *PMS2*, *RBBP8*, and *TP53BP1*, had only one positive htSNPs (Table 2). For rectal carcinoma, multiple hits were found within *BLM* (rs2518967, rs35787687), *PMS1* (rs1233258, rs1233262) and *RAD51B* (14 hits). Single hits were found for *ATM* and *DCLRE1C* (Table 3).

Bonferroni's-positive htSNPs were also evaluated as potential cis-eQTL by investigating *in silico* their association with the gene expression using GTex portal. Rs1800734 was associated with *MLH1* expression in colon transverse but the statistical significance did not reach the genome-wide level ($p=9.9\times10^{-4}$, normalized effect size, NES, of 0.12). On the other hand, rs2189517 lacked completely any association with the expression of *RAD51B* in colonic tissues as well as in all other tissues available in GTex portal. To further investigate the role of these SNPs as eQTL or any other functional annotations, we have searched other databases (http://www.exsnp.org/; http://www.scandb.org/newinterface/about.html; and http://bioinfo.life.hust.edu.cn/PancanQTL/). However, no additional information were retrieved since data on these SNPs are largely missing.

## Discussion

In the present study, we comprehensively analyzed variations in 185 DNA repair genes in over 27,000 individuals [29] to ascertain their implication for colon and rectal cancer risk. Two SNPs in MMR and HR pathways (i.e. rs1800734 in *MLH1* and rs2189517 in *RAD51B)* were associated in a statistically significant way with increased risk for cancer in colon and rectum, respectively.

Differences in the activity of DNA repair systems could play a role in modulating individual cancer risk according to tumor location in the gut [7–9]. Mutations within *MLH1* (MMR) predispose to HNPCC type-2 [37]. Somatic mutations as well as hyper-methylation of the gene promoter were frequently observed also in sporadic colorectal cancer tissues associated with a MSI phenotype [38]. Rs1800734 encodes for a G to A transition at −93 from the transcription start site within the promoter region and it falls within NF-IL6 and GT-IIB transcription factor binding sites. The polymorphism has been associated with promoter methylation and gene silencing [39, 40] and a meta-analysis by Wang and colleagues [41] reported that carriers of the A-allele are at increased risk of colorectal cancer, in agreement with the present results. The association was even stronger among cases positive for MSI. However, according to another recent meta-analysis results were not conclusive [42]. Our results, carried out on a very large series of patients, suggest that rs1800734 plays a role particularly in colon, perhaps causing a decreased activity of the MMR in this tissue [39, 43]. This hypothesis is corroborated by the data from GTEx reporting that rs1800734 could act as a cis-eQTL by affecting mRNA expression levels in colon transverse ($p=9.9\times10^{-4}$ with NES of 0.12). Discrepancies among past studies could be ascribed to statistical limitations or to differences in the composition of colon/rectal cancer patients and to variable proportions of patients with MSI phenotypes.

Concerning the second positive association (i.e. rs2189517), it is important to observe that *RAD51B* is an important gene within the HR pathway. Interestingly, previous studies reported that various *RAD51B* SNPs in LD with those reported in our study were associated

with the susceptibility to prostate and breast cancer [44]. Finally, rs2189517 has been recently related to the risk of prostate cancer in a GWAS [45]. Furthermore, it should be stressed that other SNPs within the last intron of *RAD51B*, and not in LD with those presented here, were involved in the susceptibility to breast cancer in males [46], and females [47]. Germline mutations within *RAD51B* were also found to confer predisposition to familial breast and ovarian cancer [48], and cutaneous melanoma [49].

Subsequently, we also investigated the potential, although minor, involvement of other SNPs in DNA repair by a hypothesis-generating approach, which means a less conservative adjustment for multiple testing as applies for Bonferroni adjustment. Various htSNPs resulted significantly associated after BSGoF correction for multiple testing and confirmed and provided further evidence for our hypothesis of the relevance of DNA MMR and HR pathways for colon and rectal subtypes, respectively. In fact, MMR showed more signals such as rs12112229 in *PMS2*, rs4658549 in *EXO1*, and rs72812338 in *MSH2* associated with increased colon cancer risk. It should be also noted that PMS2 forms heterodimers with MLH1 to generate MutL-alpha complex. This last, together with MutS heterodimers, is pivotal for MMR to correct small insertion-deletion mispairing formed during DNA replication or recombination. Interestingly, *MSH2* and *EXO1* are also physically interacting each other for MMR activity [50]. Additionally, together with the RAD51B SNPs found in the present study, other htSNPs within the same gene resulted associated with risk of rectal cancer being the association of rs2189517 with the lowest P-value while that of rs187645011 with the highest OR. In summary, according to all these observations, *RAD51B* and HR appear as pivotal in the individual susceptibility to various types of tumors, including rectal carcinoma.

## Conclusions

All findings hereby presented suggest the importance of genetic variations within MMR genes (in particular those that could physically interact with each other for intact MMR activity) in the predisposition to non-inherited forms of colon cancer. In contrast, for rectal carcinoma, the strongest associations were observed for a SNP within *RAD51B,* a gene involved in HR. Thus, our results show that genetic variations within DNA repair genes, in particular MMR and HR, significantly affect the risk of colon and rectal carcinoma independently with a significant impact not only, as known, for the familial forms but also for the sporadic ones.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements:

ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

COLON CFR: CCFR: We graciously thank the generous contributions of our study participants, the dedication of our study staff, and the financial support from the U.S. National Cancer Institute, without which our important registry would not exist.

DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance.

Harvard cohorts: We would like to thank the participants and staff of the HPFS, NHS and PHS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

PLCO: The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible.

PMH: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Fundings

The study was funded by the Italian Institute for Genomic Medicine (IIGM) and Compagnia di San Paolo Torino, Italy (to A. Naccarati, B. Pardini and G. Cugliari); Fondazione Umberto Veronesi 'Post-doctoral fellowship Year 2014, 2015, 2016, and 2017' (to B. Pardini); Lega Italiana per La Lotta contro i Tumori (to B. Pardini and A. Naccarati), by the Grant Agency of the Czech Republic (17–16857S to A. Naccarati); by the Istituto Toscano Tumori (grant n. I56D15000010002 to S. Landi); by AZV Ministry of Health, Czech Republic (AZV 15–27580 and AZV 17–30920 to P. Vodicka and V. Vymetalkova). B. Pardini was supported by a Fulbright Research Scholarships (year 2018).

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

COLO2&3: National Institutes of Health (R01 CA60987).

The Colon Cancer Family Registry (CFR) Illumina GWAS was supported by funding from the National Cancer Institute, National Institutes of Health (grant numbers U01 CA122839, R01 CA143247). The Colon CFR/CORECT Affymetrix Axiom GWAS and OncoArray GWAS were supported by funding from National Cancer Institute, National Institutes of Health (grant number U19 CA148107 to S Gruber). The Colon CFR participant recruitment and collection of data and biospecimens used in this study were supported by the National Cancer Institute, National Institutes of Health (grant number U01 CA167551) and through cooperative agreements with the following Colon CFR centers: Australasian Colorectal Cancer Family Registry (NCI/NIH grant numbers U01 CA074778 and U01/U24 CA097735), USC Consortium Colorectal Cancer Family Registry (NCI/NIH grant numbers U01/U24 CA074799), Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (NCI/NIH grant number U01/U24 CA074800), Ontario Familial Colorectal Cancer Registry (NCI/NIH grant number U01/U24 CA074783), Seattle Colorectal Cancer Family Registry (NCI/NIH grant number U01/U24 CA074794), and University of Hawaii Colorectal Cancer Family Registry (NCI/NIH grant number U01/U24 CA074806), Additional support for case ascertainment was provided from the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute to Fred Hutchinson Cancer Research Center (Control Nos. N01-CN-67009 and N01-PC-35142, and Contract No. HHSN2612013000121), the Hawai'i Department of Health (Control Nos. N01-PC-67001 and N01-PC-35137, and Contract No. HHSN26120100037C, and the California Department of Public Health (contracts HHSN261201000035C awarded to the University of Southern California, and the following state cancer registries: AZ, CO, MN, NC, NH, and by the Victoria Cancer Registry and Ontario Cancer Registry.

DACHS: This work was supported by the German Research Council (BR 1704/6–1, BR 1704/6–3, BR 1704/6–4, CH 117/1–1, HO 5117/2–1, HE 5998/2–1, KL 2354/3–1, RO 2270/8–1 and BR 1704/17–1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: National Institutes of Health (R01 CA48998 to M. L. Slattery).

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA164930, U01 CA137088, R01 CA059045). This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704.

arvard cohorts (HPFS, NHS, PHS): HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003) and PHS by the National Institutes of Health (R01 CA042182).

MEC: National Institutes of Health (R37 CA54281, P01 CA033619, and R01 CA063464).

OFCCR: National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see CCFR section above. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

PMH: National Institutes of Health (R01 CA076366 to P.A. Newcomb).

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

## Abbreviation list

| | |
|---|---|
| **BSGoF** | binomial sequential goodness of fit |
| **CCFR** | Colon Cancer Family Registry |
| **CI** | confidence interval |
| **eQTL** | expression quantitative trait loci |
| **FDR** | false discovery rate |
| **GECCO** | Genetics and Epidemiology of Colorectal Cancer |
| **GWAS** | genome-wide association studies |
| **HNPCC** | hereditary nonpolyposis colorectal cancer |
| **HR** | homologous recombination |
| **HRC** | Haplotype Reference Consortium |
| **htSNPs** | haplotype tagging SNPs |
| **HWE** | Hardy-Weinberg Equilibrium |
| **LD** | linkage disequilibrium |
| **MAF** | minor allele frequency |
| **MMR** | mismatch repair pathway |
| **MSI** | microsatellite instability |

| | |
|---|---|
| **NES** | normalized effect size |
| **OR** | odds ratio |
| **PARP** | poly(ADP-ribose) polymerase |
| **PCs** | principal components |
| **SNPs** | single-nucleotide polymorphisms |

## References


1. Lord CJ, Ashworth A. The DNA damage response and cancer therapy. Nature 2012;481: 287–94. [PubMed: 22258607]
2. Naccarati A, Rosa F, Vymetalkova V, Barone E, Jiraskova K, Di Gaetano C, Novotny J, Levy M, Vodickova L, Gemignani F, Buchler T, Landi S, et al. Double-strand break repair and colorectal cancer: gene variants within 3' UTRs and microRNAs binding as modulators of cancer risk and clinical outcome. Oncotarget 2016;7: 23156–69. [PubMed: 26735576]
3. Poulogiannis G, Frayling IM, Arends MJ. DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. Histopathology 2010;56: 167–79. [PubMed: 20102395]
4. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61: 759–67. [PubMed: 2188735]
5. Li SK, Martin A. Mismatch Repair and Colon Cancer: Mechanisms and Therapies Explored. Trends Mol Med 2016;22: 274–89. [PubMed: 26970951]
6. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology 2010;138: 2073–87 e3. [PubMed: 20420947]
7. Pardini B, Naccarati A, Novotny J, Smerhovsky Z, Vodickova L, Polakova V, Hanova M, Slyskova J, Tulupova E, Kumar R, Bortlik M, Barale R, et al. DNA repair genetic polymorphisms and risk of colorectal cancer in the Czech Republic. Mutation research 2008;638: 146–53. [PubMed: 17991492]
8. Slyskova J, Cordero F, Pardini B, Korenkova V, Vymetalkova V, Bielik L, Vodickova L, Pitule P, Liska V, Matejka VM, Levy M, Buchler T, et al. Post-treatment recovery of suboptimal DNA repair capacity and gene expression levels in colorectal cancer patients. Molecular carcinogenesis 2015;54: 769–78. [PubMed: 24585457]
9. Naccarati A, Pardini B, Hemminki K, Vodicka P. Sporadic colorectal cancer and individual susceptibility: a review of the association studies investigating the role of DNA repair genetic polymorphisms. Mutation research 2007;635: 118–45. [PubMed: 17419091]
10. Laporte GA, Leguisamo NM, Kalil AN, Saffi J. Clinical importance of DNA repair in sporadic colorectal cancer. Critical reviews in oncology/hematology 2018;126: 168–85. [PubMed: 29759559]
11. Li Y, Li S, Wu Z, Hu F, Zhu L, Zhao X, Cui B, Dong X, Tian S, Wang F, Zhao Y. Polymorphisms in genes of APE1, PARP1, and XRCC1: risk and prognosis of colorectal cancer in a northeast Chinese population. Medical oncology 2013;30: 505. [PubMed: 23430444]
12. Gil J, Gaj P, Misiak B, Ostrowski J, Karpinski P, Jarczynska A, Kielan W, Sasiadek MM. CYP1A1 Ile462Val polymorphism and colorectal cancer risk in Polish patients. Medical oncology 2014;31: 72. [PubMed: 24939416]
13. Hua RX, Zhu J, Jiang DH, Zhang SD, Zhang JB, Xue WQ, Li XZ, Zhang PF, He J, Jia WH. Association of XPC Gene Polymorphisms with Colorectal Cancer Risk in a Southern Chinese Population: A Case-Control Study and Meta-Analysis. Genes (Basel) 2016;7.
14. Tao H, Shinmura K, Suzuki M, Kono S, Mibu R, Tanaka M, Kakeji Y, Maehara Y, Okamura T, Ikejiri K, Futami K, Yasunami Y, et al. Association between genetic polymorphisms of the base excision repair gene MUTYH and increased colorectal cancer risk in a Japanese population. Cancer science 2008;99: 355–60. [PubMed: 18271935]

15. Forat-Yazdi M, Gholi-Nataj M, Neamatzadeh H, Nourbakhsh P, Shaker-Ardakani H. Association of XRCC1 Arg399Gln Polymorphism with Colorectal Cancer Risk: A HuGE Meta Analysis of 35 Studies. Asian Pacific journal of cancer prevention : APJCP 2015;16: 3285–91. [PubMed: 25921133]

16. Eskandari E, Rezaifar A, Hashemi M. XPG Asp1104His, XRCC2 Rs3218536 A/G and RAD51 135G/C Gene Polymorphisms and Colorectal Cancer Risk: A Meta-Analysis. Asian Pacific journal of cancer prevention : APJCP 2017;18: 1805–13. [PubMed: 28749109]

17. Zhang T, Zhang DM, Zhao D, Hou XM, Ma SC, Liu XJ. Lack of association between the XPD Lys751Gln polymorphism and colorectal cancer risk: a meta-analysis. OncoTargets and therapy 2014;7: 1255–60. [PubMed: 25050067]

18. Vineis P, Manuguerra M, Kavvoura FK, Guarrera S, Allione A, Rosa F, Di Gregorio A, Polidoro S, Saletta F, Ioannidis JP, Matullo G. A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility. Journal of the National Cancer Institute 2009;101: 24–36. [PubMed: 19116388]

19. Pardini B, Naccarati A, Polakova V, Smerhovsky Z, Hlavata I, Soucek P, Novotny J, Vodickova L, Tomanova V, Landi S, Vodicka P. NBN 657del5 heterozygous mutations and colorectal cancer risk in the Czech Republic. Mutation research 2009;666: 64–7. [PubMed: 19393249]

20. Tomlinson IP, Houlston RS, Montgomery GW, Sieber OM, Dunlop MG. Investigation of the effects of DNA repair gene polymorphisms on the risk of colorectal cancer. Mutagenesis 2012;27: 219–23. [PubMed: 22294770]

21. Houlston RS. COGENT (COlorectal cancer GENeTics) revisited. Mutagenesis 2012;27: 143–51. [PubMed: 22294761]

22. Shen L, Kondo Y, Rosner GL, Xiao L, Hernandez NS, Vilaythong J, Houlihan PS, Krouse RS, Prasad AR, Einspahr JG, Buckmeier J, Alberts DS, et al. MGMT promoter methylation and field defect in sporadic colorectal cancer. Journal of the National Cancer Institute 2005;97: 1330–8. [PubMed: 16174854]

23. Broderick P, Dobbins SE, Chubb D, Kinnersley B, Dunlop MG, Tomlinson I, Houlston RS. Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients-a Systematic Review. Gastroenterology 2017;152: 75–7 e4. [PubMed: 27713038]

24. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nature genetics 2009;41: 579–84. [PubMed: 19330030]

25. Ogino S, Chan AT, Fuchs CS, Giovannucci E. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. Gut 2011;60: 397–411. [PubMed: 21036793]

26. Lu M, Sun L, Zhou J, Zhang J. Assessment of the association between hOGG1 C8069G polymorphism and colorectal cancer. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 2014;35: 2373–7. [PubMed: 24297334]

27. Wang Z, Zhang W. Association between XRCC3 Thr241Met polymorphism and colorectal cancer risk. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 2013;34: 1421–9.

28. Liu L, Miao L, Ji G, Qiang F, Liu Z, Fan Z. Association between XRCC1 and XRCC3 polymorphisms and colorectal cancer risk: a meta-analysis of 23 case-control studies. Molecular biology reports 2013;40: 3943–52. [PubMed: 23271134]

29. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, Berndt SI, Bezieau S, Brenner H, Butterbach K, Caan BJ, Campbell PT, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology 2013;144: 799–807 e24.

30. Hutter CM, Chang-Claude J, Slattery ML, Pflugeisen BM, Lin Y, Duggan D, Nan H, Lemire M, Rangrej J, Figueiredo JC, Jiao S, Harrison TA, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. Cancer research 2012;72: 2036–44. [PubMed: 22367214]

31. Newcomb PA, Baron J, Cotterchio M, Gallinger S, Grove J, Haile R, Hall D, Hopper JL, Jass J, Le Marchand L, Limburg P, Lindor N, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2007;16: 2331–43.
32. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nature genetics 2007;39: 989–94. [PubMed: 17618283]
33. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, et al. Next-generation genotype imputation service and methods. Nature genetics 2016;48: 1284–7. [PubMed: 27571263]
34. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nature methods 2013;10: 5–6. [PubMed: 23269371]
35. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics 2016;48: 1279–83. [PubMed: 27548312]
36. Castro-Conde I, de Una-Alvarez J. Adjusted p-values for SGoF multiple test procedure. Biom J 2015;57: 108–22. [PubMed: 25323102]
37. Knudson AG. Hereditary predisposition to cancer. Annals of the New York Academy of Sciences 1997;833: 58–67. [PubMed: 9616740]
38. Oliveira C, Westra JL, Arango D, Ollikainen M, Domingo E, Ferreira A, Velho S, Niessen R, Lagerstedt K, Alhopuro P, Laiho P, Veiga I, et al. Distinct patterns of KRAS mutations in colorectal carcinomas according to germline mismatch repair defects and hMLH1 methylation status. Human molecular genetics 2004;13: 2303–11. [PubMed: 15294875]
39. Ito E, Yanagisawa Y, Iwahashi Y, Suzuki Y, Nagasaki H, Akiyama Y, Sugano S, Yuasa Y, Maruyama K. A core promoter and a frequent single-nucleotide polymorphism of the mismatch repair gene hMLH1. Biochem Bioph Res Co 1999;256: 488–94.
40. Goldsborough AS, Kornberg TB. Reduction of transcription by homologue asynapsis in Drosophila imaginal discs. Nature 1996;381: 807–10. [PubMed: 8657287]
41. Wang T, Liu Y, Sima L, Shi L, Wang Z, Ni C, Zhang Z, Wang M. Association between MLH1 −93G>a polymorphism and risk of colorectal cancer. PLoS One 2012;7: e50449.
42. Chen H, Shen Z, Hu Y, Xiao Q, Bei D, Shen X, Ding K. Association between MutL homolog 1 polymorphisms and the risk of colorectal cancer: a meta-analysis. Journal of cancer research and clinical oncology 2015;141: 2147–58. [PubMed: 25986311]
43. Perera S, Mrkonjic M, Rawson JB, Bapat B. Functional effects of the MLH1–93G>A polymorphism on MLH1/EPM2AIP1 promoter activity. Oncology reports 2011;25: 809–15. [PubMed: 21206982]
44. Nowacka-Zawisza M, Wisnik E, Wasilewski A, Skowronska M, Forma E, Brys M, Rozanski W, Krajewska WM. Polymorphisms of homologous recombination RAD51, RAD51B, XRCC2, and XRCC3 genes and the risk of prostate cancer. Anal Cell Pathol (Amst) 2015;2015: 828646.
45. Amin Al Olama A, Dadaev T, Hazelett DJ, Li Q, Leongamornlert D, Saunders EJ, Stephens S, Cieza-Borrella C, Whitmore I, Benlloch Garcia S, Giles GG, Southey MC, et al. Multiple novel prostate cancer susceptibility signals identified by fine-mapping of known risk loci among Europeans. Human molecular genetics 2015;24: 5589–602. [PubMed: 26025378]
46. Orr N, Lemnrau A, Cooke R, Fletcher O, Tomczyk K, Jones M, Johnson N, Lord CJ, Mitsopoulos C, Zvelebil M, McDade SS, Buck G, et al. Genome-wide association study identifies a common variant in RAD51B associated with male breast cancer risk. Nature genetics 2012;44: 1182–4. [PubMed: 23001122]
47. Pelttari LM, Khan S, Vuorela M, Kiiski JI, Vilske S, Nevanlinna V, Ranta S, Schleutker J, Winqvist R, Kallioniemi A, Dork T, Bogdanova NV, et al. RAD51B in Familial Breast Cancer. PLoS One 2016;11: e0153788.

48. Golmard L, Caux-Moncoutier V, Davy G, Al Ageeli E, Poirot B, Tirapo C, Michaux D, Barbaroux C, d'Enghien CD, Nicolas A, Castera L, Sastre-Garau X, et al. Germline mutation in the RAD51B gene confers predisposition to breast cancer. BMC Cancer 2013;13: 484. [PubMed: 24139550]

49. Wadt KA, Aoude LG, Golmard L, Hansen TV, Sastre-Garau X, Hayward NK, Gerdes AM. Germline RAD51B truncating mutation in a family with cutaneous melanoma. Familial cancer 2015;14: 337–40. [PubMed: 25600502]

50. Schmutte C, Marinescu RC, Sadoff MM, Guerrette S, Overhauser J, Fishel R. Human exonuclease I interacts with the mismatch repair protein hMSH2. Cancer research 1998;58: 4537–42. [PubMed: 9788596]

## Novelty & Impact Statements

The results presented in this study provide new insights on candidate SNPs (rs1800734 in MLH1 gene and rs2189517 in RAD51B) involved in DNA repair that may spur downstream investigation into the biology of risk for colon and rectal cancers with a reflection in improving drug development and clinical guidelines, such as personalized screening decisions.

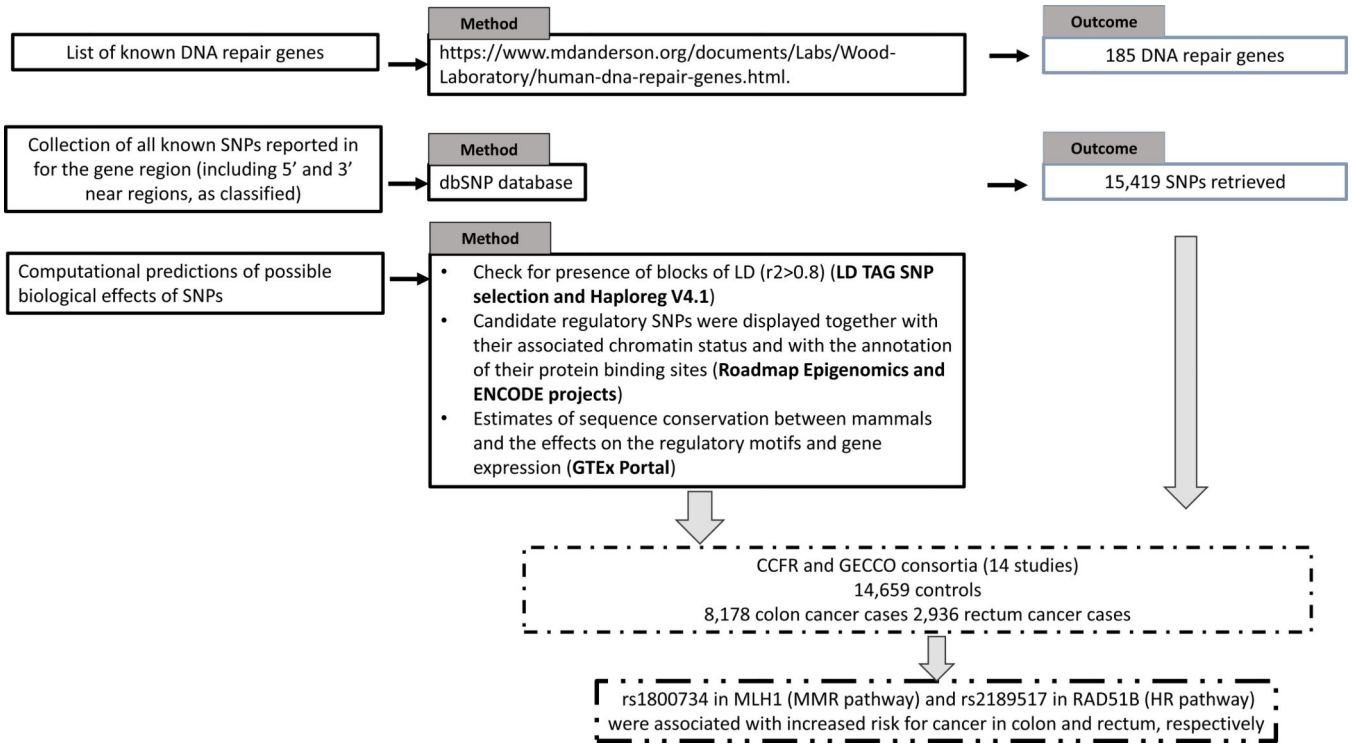


**Figure 1.**
Workflow of the study

## Table 1

Description of study populations included in the Colon Cancer Family Registry (CCFR) and the Genetics and Epidemiology of ColorectalCancer Consortium(GECCO).

| Study* | Total** | Sex | | Controls | Cases | Cancer site | |
|---|---|---|---|---|---|---|---|
| | | *Females* | *Males* | N=14662 | N=11898 | *Colon (proximal / distal)* | *Rectum* |
| ASTERISK | 1839 | 763 | 1076 | 947 | 892 | 622 (249 / 373) | 260 |
| CCFR Set 1 | 2016 | 1010 | 1006 | 978 | 1038 | 700 (317 / 375) | 448 |
| CCFR Set 2 | 717 | 389 | 328 | 386 | 331 | 237 (97 / 127) | 135 |
| Colo 2&3 | 211 | 94 | 117 | 124 | 87 | 59 (35 / 24) | 27 |
| DACHS Set 1 | 3409 | 1393 | 2016 | 1702 | 1707 | 1037 (548 / 487) | 668 |
| DACHS Set 2 | 1164 | 435 | 729 | 498 | 666 | 385 (210 / 175) | 281 |
| DALS Set 1 | 1411 | 612 | 799 | 709 | 702 | 702 (329 / 358) | 0 |
| DALS Set 2 | 863 | 410 | 453 | 461 | 402 | 410 (209 / 185) | 0 |
| HPFS Set 1 | 456 | 0 | 456 | 229 | 227 | 158 (82 / 76) | 48 |
| HPFS Set 2 | 348 | 0 | 348 | 172 | 176 | 111 (54 / 57) | 40 |
| HPFS_AD | 656 | 0 | 656 | 343 | 313 | n/a | n/a |
| MEC | 672 | 311 | 361 | 346 | 326 | 241 (155 / 86) | 81 |
| NHS Set 1 | 1165 | 1165 | 0 | 774 | 391 | 305 (175 / 123) | 82 |
| NHS Set 2 | 339 | 339 | 0 | 181 | 158 | 112 (67 / 44) | 35 |
| NHS_AD | 1090 | 1090 | 0 | 577 | 513 | n/a | n/a |
| OFCCR | 1116 | 579 | 537 | 522 | 594 | 396 (204 / 164) | 188 |
| PHS | 764 | 0 | 764 | 389 | 375 | 286 (122 / 121) | 84 |
| PLCO Set 1 | 2496 | 664 | 1832 | 1972 | 524 | 516 (323 / 193) | 5 |
| PLCO Set 2 | 889 | 379 | 510 | 414 | 475 | 320 (213 / 102) | 161 |
| PMH-CCFR | 398 | 398 | 0 | 122 | 276 | 206 (132 / 72) | 64 |
| VITAL | 566 | 267 | 299 | 287 | 279 | 215 (143 / 69) | 66 |
| WHI Set 1 | 1991 | 1991 | 0 | 1523 | 468 | 456 (308 / 147) | 14 |
| WHI Set 2 | 1984 | 1984 | 0 | 1006 | 978 | 704 (482 / 222) | 249 |

Numbers may not add up to 100% of available subjects because of missing information; n/a information not available

*For the complete list and description of the studies, see Supplementary materials. ASTERISK, Colo2&3, DALS Set 2, DACHS Set 1, PMH-CCFR, MEC, PLCO Set 2, WHI Set 2 and VITAL were genotyped on the Illumina CytoSNP BeadChip. WHI Set 1 was genotyped using Illumina 550K, 550K duo, and 610K platforms (only 550K and 550K duo if not utilizing hip fracture controls). PLCO Set 1 was genotyped using Illumina 550K and 610K platforms (also the 550K Duo platform if using the PLCO rematch set). DALS Set 1 was genotyped using Illumina 610K and 550K platforms. OFCCR was genotyped using Affymetrix GeneChip Human mapping 100K and 500K Array Set and a 10K non-synonymous SNP chip. CCFR was genotyped using Illumina Human1M and Human1M-Duo platforms. DACHS Set 2, HPFS, NHS, and PHS were genotyped on the OmniExpress platform.

**Sample sizes based on GECCO GIGSv3/HRCv1 data.

## Table 2

htSNPs with the lowest p-values for the association with risk of colon cancer, grouped by gene

| Colon (proximal+distal) | | | |
|---|---|---|---|
| **ht SNP ID** | **OR (95% CI)** | **p-value for SNP effect** | **BSGoF-Adjusted p-value** |
| ***ATM*** ***(DSBR)*** | | | |
| rs11212592 | 0.92 (0.87–0.97) | $3.30X10^{-03}$ | 0.021 |
| rs61915066 | 1.13 (1.04–1.24) | $4.81X10^{-03}$ | 0.021 |
| ***EXO1*** ***(MMR, EPN)*** | | | |
| rs4658549 | 1.06 (1.02–1.11) | $5.33X10^{-03}$ | 0.021 |
| ***FANCA*** ***(DCLR)*** | | | |
| rs2238526 | 0.94 (0.91–0.98) | $3.98X10^{-03}$ | 0.021 |
| rs3743861 | 0.94 (0.91–0.98) | $6.13X10^{-03}$ | 0.021 |
| ***FANCE*** ***(DCLR)*** | | | |
| rs6907678 | 0.94 (0.90–0.98) | $1.79X10^{-03}$ | 0.020 |
| rs10947550 | 0.94 (0.90–0.98) | $2.32X10^{-03}$ | 0.020 |
| ***FEN1*** ***(BER, EPN)*** | | | |
| rs4246215 | 0.93 (0.89–0.97) | $1.59X10^{-03}$ | 0.020 |
| ***LIG1*** ***(NER,BER)*** | | | |
| rs1971775 | 1.06 (1.02–1.11) | $4.38X10^{-03}$ | 0.021 |
| rs73054038 | 0.92 (0.87–0.98) | $5.32X10^{-03}$ | 0.021 |
| ***MLH1*** ***(MMR)*** | | | |
| rs1800734 | 1.13 (1.07–1.18) | $^{\S}\mathbf{3.48X10^{-06}}$ | 0.019 |
| rs6784088 | 0.94 (0.90–0.98) | $3.28X10^{-03}$ | 0.020 |
| rs9855475 | 0.94 (0.90–0.98) | $3.42X10^{-03}$ | 0.021 |
| ***PMS2*** ***(MMR)*** | | | |
| rs12112229 | 1.07 (1.02–1.13) | $3.15X10^{-03}$ | 0.020 |
| ***RBBP8*** ***(HR)*** | | | |
| rs113047993 | 1.15 (1.07–1.25) | $2.36X10^{-04}$ | 0.019 |
| ***TP53BP1*** ***(NHEJ)*** | | | |
| rs17782975 | 0.88 (0.82–0.96) | $1.91X10^{-03}$ | 0.020 |

$^{\S}$statistically significant after Bonferroni's correction

## Table 3

htSNPs with the lowest p-values for the association with risk of rectal cancer, grouped by gene

| | Rectum | | |
|---|---|---|---|
| **ht SNP ID** | **OR(95% CI)** | **p-value for SNP effect** | **BSGoF-Adjusted p-value** |
| ***ATM*** **(DSBR)** | | | |
| rs11212592 | 0.87 (0.80–0.94) | $1.67X10^{-03}$ | $2.86X10^{-03}$ |
| ***BLM*** **(DSBR)** | | | |
| rs2518967 | 1.14 (1.06–1.24) | $5.97X10^{-04}$ | $1.35X10^{-03}$ |
| rs35787687 | 1.14 (1.05–1.23) | $1.07X10^{-04}$ | $2.32X10^{-03}$ |
| ***DCLRE1C*** **(NHEJ)** | | | |
| rs7920514 | 1.12 (1.04–1.21) | $1.37X10^{-03}$ | $2.50X10^{-03}$ |
| ***PMS1*** **(MMR)** | | | |
| rs1233258 | 0.87 (0.80–0.93) | $2.37X10^{-04}$ | $7.22X10^{-04}$ |
| rs1233262 | 0.89 (0.83–0.95) | $1.21X10^{-04}$ | $2.32X10^{-03}$ |
| ***RAD51B*** **(HR)** | | | |
| rs2189517 | 1.15 (1.08–1.22) | $^{\S}\mathbf{5.73X10^{-06}}$ | $1.24X10^{-05}$ |
| rs12587232 | 1.13 (1.06–1.20) | $4.37X10^{-05}$ | $6.56X10^{-04}$ |
| rs187645011 | 0.71 (0.61–0.84) | $7.53X10^{-05}$ | $6.90X10^{-04}$ |
| rs7350713 | 0.74 (0.63–0.87) | $2.47X10^{-04}$ | $7.44X10^{-04}$ |
| rs6573841 | 0.86 (0.80–0.93) | $3.43X10^{-04}$ | $7.59X10^{-04}$ |
| rs111611396 | 0.74 (0.63–0.87) | $3.53X10^{-04}$ | $7.66X10^{-04}$ |
| rs1989974 | 0.76 (0.65–0.88) | $4.20X10^{-04}$ | $7.81X10^{-04}$ |
| rs117544253 | 0.72 (0.60–0.86) | $4.70X10^{-04}$ | $8.69X10^{-04}$ |
| rs77726787 | 0.74 (0.63–0.87) | $4.76X10^{-04}$ | $9.30X10^{-04}$ |
| rs8016488 | 1.15 (1.06–1.25) | $5.27X10^{-04}$ | $1.31X10^{-03}$ |
| rs11628293 | 1.12 (1.05–1.20) | $6.28X10^{-04}$ | $1.37X10^{-03}$ |
| rs113020754 | 0.75 (0.64–0.88) | $6.50X10^{-04}$ | $1.54X10^{-03}$ |
| rs80085210 | 0.74 (0.62–0.88) | $9.30X10^{-04}$ | $2.30X10^{-03}$ |
| rs113300322 | 0.77 (0.66–0.90) | $9.49X10^{-04}$ | $2.30X10^{-03}$ |
| rs74933543 | 1.11 (1.04–1.19) | $1.59X10^{-03}$ | $2.85X10^{-03}$ |

$^{\S}$statistically significant after Bonferroni's correction